



Sveučilište u Zagrebu

Filozofski fakultet

Ivana Dević

**RAZVOJ ONTOLOGIJE IZ
ENCIKLOPEDIČKI ORGANIZIRANOGA
ZNANJA**

DOKTORSKI RAD

Zagreb, 2017.



University of Zagreb

Faculty of Humanities and Social Sciences

Ivana Dević

**ONTOLOGY DEVELOPMENT FROM THE
ENCYCLOPAEDIC ORGANIZATION OF
KNOWLEDGE**

DOCTORAL THESIS

Zagreb, 2017.



Sveučilište u Zagrebu

Filozofski fakultet

Ivana Dević

**RAZVOJ ONTOLOGIJE IZ
ENCIKLOPEDIČKI ORGANIZIRANOGA
ZNANJA**

DOKTORSKI RAD

Mentori:
prof. dr. sc. Damir Boras
dr. sc. Zdenko Jecić

Zagreb, 2017.



University of Zagreb

Faculty of Humanities and Social Sciences

Ivana Dević

**ONTOLOGY DEVELOPMENT FROM THE
ENCYCLOPAEDIC ORGANIZATION OF
KNOWLEDGE**

DOCTORAL THESIS

Supervisors:
prof. dr. sc. Damir Boras
dr. sc. Zdenko Jecić

Zagreb, 2017.

PODACI O MENTORIMA

Prof. dr. sc. Damir Boras diplomirao je na Elektrotehničkom fakultetu Sveučilišta u Zagrebu. Poslijediplomski studij informacijskih znanosti završio je na Filozofskom fakultetu Sveučilišta u Zagrebu, gdje je 1998. obranio doktorski rad pod naslovom „ Teorija i pravila segmentacije teksta na hrvatskom jeziku”.

Od siječnja 1975. zaposlio se u Referalnom centru Sveučilišta u Zagrebu, a od 1984. godine prelazi na Filozofski fakultet u Zagrebu. U Leksikografskom zavodu Miroslav Krleža je od 2000. do 2006. godine obavljao dužnost prvog pomoćnika ravnatelja za informatizaciju i znanstvenu suradnju. Od 2009. do 2014. je obnašao dužnost dekana na Filozofskog fakulteta u Zagrebu. Predstojnik je i utemeljitelj Katedre za leksikografiju i enciklopediku pri Odsjeku za informacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu od 2004. Predsjednik je Upravnog vijeća Nacionalnog centra za vanjsko vrednovanje obrazovanja NCVVO od listopada 2009. Od 2011. do 2012. je bio predsjednik Upravnoga vijeća Hrvatske izvještajne novinske agencije (Hine). Kao predložnik Vijeća društveno-humanističkog područja te Fakulteta političkih znanosti, Filozofskoga fakulteta i Učiteljskoga fakulteta, tajnim je glasovanjem izabran za 82. rektora Sveučilišta u Zagrebu. Svoje četverogodišnje mandatno razdoblje je započeo u listopadu 2014. godine.

Sudjelovao je i sudjeluje kao istraživač na više projekata pri Ministarstvu znanosti (od 1990 do danas). Tako je primjerice bio voditelj znanstvenog projekta Ministarstva znanosti i tehnologije RH "130464 Hrvatska rječnička baština i prikaz rječničkoga znanja" od 2002. do 2006., glavni istraživač na međunarodnom projektu uz potporu UNESCO-a "Croatian dictionary heritage" od 2003. do 2006., voditelj projekta za primjenu informacijske tehnologije Ministarstva znanosti, obrazovanja i sporta "130500 Portal hrvatskoga računalnoga nazivlja" od 2004. – 2005., voditelj znanstvenog projekta Ministarstva znanosti obrazovanja i sporta RH u području informacijskih znanosti "130-130I679-1380 Hrvatska rječnička baština i hrvatski europski identitet" od 1.1.2007., voditelj znanstvenog programa Ministarstva znanost, obrazovanja i sporta RH u području informacijskih znanosti "Izvori za hrvatsku baštinu i hrvatski europski identitet", koji obuhvaća šest znanstvenih projekata iz područja informacijskih znanosti, društvenih znanosti i humanističkih znanosti od 1.1.2007. Kodirektor je međunarodne znanstvene konferencije *Information Technology and Journalism*, Dubrovnik, IUC od 2004.

Osim istaknutog nastavnog rada, ostvario je veliki doprinos i u stručnom radu. Bio je predsjednik stručnoga savjeta projekta razvoja informacijske baze Informacijsko-dokumentacijskog centra u Hini od 2008. do 2010. Član je uređivačkog odbora međunarodnog časopisa za pitanja medija, novinarstva, masovnog komuniciranja i odnosa s javnostima *Medianali* (ISSN 1846-436X) od 2007. Glavni je urednik međunarodnog znanstvenog časopisa *Studia lexicographica* (ISSN 1846-6745), Leksikografski zavod Miroslav Krleža od 2007. Član je Upravnog vijeća Informacijskog sustava visokih učilišta Republike Hrvatske ISVU od veljače 2009.

Objavio više od devedeset znanstvenih i stručnih radova, knjiga i srednjoškolskih i visokoškolskih udžbenika te više od 40 pozvanih predavanja i priopćenja na stranim i domaćim znanstvenim skupovima iz područja informacijskih i komunikacijskih znanosti, programiranja, računalne obrade teksta i jezika, leksikografije i enciklopedistike.

Dr. sc. Zdenko Jecić je diplomirao arhitekturu na Arhitektonskom fakultetu u Zagrebu 1993. Magistrirao je 1999. na Građevinskom fakultetu u Zagrebu s temom "Primjena modela pri određivanju oblika vlačno napregnutih konstrukcija". Na Filozofski fakultet u Zagrebu je 2009. obranio doktorski rad pod naslovom "Virtualna enciklopedija, redefiniranje zadaće enciklopedijske djelatnosti".

Zaposlen je kao leksikograf u hrvatskoj središnjoj leksikografskoj ustanovi Miroslav Krleža. Docent je u naslovnom zvanju na Filozofskom fakultetu u Zagrebu od 2013. Izvodi kolegiji Uvod u enciklopediku (diplomski studij Informacijskih znanosti) te Enciklopedistika (doktorski studij Informacijskih znanosti).

Istaknuo se uredničkim radom na više leksikografskih i enciklopedičkih djela. Svakako je vrijedno istaknuti rad u uredništvu *Tehničkoga leksikona* od 2002-2007. Glavni je urednik *Hrvatske tehničke enciklopedije* od 2014. Radi u uredništvu *Hrvatske enciklopedije* od 2000., gdje je od 2011. do 2014. obnašao i dužnost pomoćnika glavnog urednika. Bio je suradnik *Hrvatskog općeg leksikona* od 2007. do 2009., kao i *Likovnog leksikona* od 2009. Predsjednik je Znanstvenog vijeća Zavoda te je član uredništva časopisa *Studia lexicographica*.

Objavio je više znanstvenih i stručnih radova, poglavlja u nekoliko knjiga te predavanja održana na znanstvenim skupovima.

Sažetak

Globalna mreža, internet, ubrzano se transformira u semantičku mrežu prelazeći od povezivanja dokumenata na povezivanje podataka, odnosno dosadašnji web portali s klasičnim bazama informacija i znanja postaju povezani podaci (engl. linked data) globalnog oblaka (engl. cloud computing).

Usporedbom organizacijske strukture tradicionalnih enciklopedija na papirnatom mediju s onima u mrežnom okruženju uočavaju se određene razlike proizašle iz različitih vrsta medija koje omogućavaju nove funkcionalnosti pretraživanja. Promjene kroz koje prolazi enciklopedičko djelo zahtijeva uspostavljanje novog načina modeliranja organizacije enciklopedičkog znanja u mrežnom okruženju koje će svoje utemeljenje pronaći na analizi specifičnosti strukture enciklopedičkog članka uvažavanjem temeljnih postavki semantičkog weba, principa ontologijskog oblikovanja i potreba korisnika kako bi se osigurala njihova dostupnost i bolja iskoristivost. U svrhu očuvanja korisnosti enciklopedije u današnjoj mrežnoj informacijskoj eksploziji potrebno je poboljšati sposobnost predstavljanja njihovog sadržaja na smisleni (semantički) način u mrežnom okruženju.

Cilj ove doktorske disertacije je istražiti koji elementi enciklopedičke organizacije znanja mogu pružiti podršku za razvoj ontologije te razviti metodu kojom će se generirati ontologija na osnovi enciklopedički organiziranog znanja.

Proučavanjem literature, analizom sličnih ontologijskih modela odabranih svjetskih projekata i znanja pohranjenog u biografskim člancima Hrvatske enciklopedije (HE) iz područja hrvatske književnosti predložio se ontologijski model koji na djelotvoran način opisuje enciklopedičko znanje navedenog područja. Primijenjena je metodologija analize sadržaja odabranih 1170 članaka HE te METHONTOLOGY metoda. Upotrijebio se Protege softver za razvoj ontologije. Klasna hijerarhija ontologije se razvija FCA pristupom te se provodi LSA metoda u svrhu određivanja pojma kao skupa srodnih termina te pripadnosti pojedinih dokumenta (članaka) tom pojmu, čime je omogućena automatska klasifikacija članaka pojedinim ontologijskim klasama. Razvijena ontologija poslužit će za organizaciju, pretraživanje i pregledavanje znanja mrežne HE iz odabranog područja hrvatske književnosti, kao i za dobivanje preciznih odgovora na složena pitanja. Obuhvaćen je veliki broj odnosa potreban za opis produkcije književnika pojedinih nacionalnih književnosti, njihovih životopisa, međusobnih odnosa, odnosa između pojedinih književnih djela i cjelokupnog znanja koje se nalazi u biografskim enciklopedijskim člancima iz područja književnosti čime

se omogućuje opis književnosti bilo kojeg naroda. Dobivena ontologija omogućuje uspostavljanje interoperabilnosti i povezivanje s ostalim strukturiranim izvorima enciklopedijskog znanja na semantičkoj mreži (npr. DBpedia), što će omogućiti povezivanje relevantnog i bogatog znanja HE u „globalnu mrežu znanja“ koja nastaje i razvija se kroz projekte semantičke mreže.

Ključne riječi: enciklopedija, Hrvatska enciklopedija, biografski enciklopedički članak, ontologija, semantički web, FCA, LSA

Extended abstract

The global network, the Internet, is hurriedly transforming into a semantic network by turning from document connection onto data connection, i.e. today's WEB portals with classic information databases are becoming linked data of global computing.

By comparing organisational structures of traditional encyclopedias on paper media with electronic encyclopedias in Web surroundings, you can notice certain differences that come from different types of media which enable new functionalities of searching. The above mentioned changes, through which an encyclopedical work passes through, demands an establishment of a new way of modeling organisation of encyclopedical knowledge in a Web surrounding that will find its foundation on the analysis of the specificity of a structure of an encyclopedical chapter by respecting the basic settings of the Semantic Web, principles of ontological shaping and needs of users so their availability and usability would be ensured. With the aim of preserving encyclopedical usability in today's Web information explosion, there is a need of modifying presentation of its content in a meaningful (semantic way) in a Web surrounding. Semantic interoperability means an existence of infrastructure which will enable mechanical interpretation and conclusion about content on the Web. Therefore, the key term of Semantic Web is presented by ontology, the basic component in enabling semantic interoperability.

The aim of this doctoral thesis is to find out which elements of encyclopedical knowledge organisation can offer support for the development of ontology and develop a method by which ontology will be generated on the basis of encyclopedically organised knowledge. Developed ontology will be used for organisation, searching and browsing data of Croatia's Web Encyclopedia in the selected field of Croatian literature, as well as receiving precise answers on asked questions.

In the introductory part of this work are presented its starting points, goals and methods, as well as the structure of the entire work.

The second chapter deals with a theoretical display and clarification of Semantic Web with the aim of its full understanding. The goal of this chapter is to point out the basic theoretical and technical background of Semantic Web, the meaning of the term Semantic Web is explained, basic difference between the Web we know today and its development toward Semantic Web, disadvantages of today's Web and advantages of Semantic Web are explained, basic terms and the architecture of Semantic Web, review of basic ontologic

definitions and their main goal and role. The chapter gives a detailed review of basic languages used on the semantic web with actual examples (RDF, RDFS; OWL and SKOS).

Some of the more important projects of Semantic Web are shown in the third chapter of the work. During the selection of significant projects, it was considered to choose projects which are significant for better understanding of Semantic Web in the field of encyclopedics. That is why one part of this chapter deals with showing ontologies unavoidable for better understanding of Semantic Web, and the other part of the chapter gives a review of ontologic projects created entirely on encyclopedical knowledge. The mentioned analysis of existing encyclopedical ontological projects shows that no former project did try to connect and research the development of ontology and its constructive elements on the basis of structural organisation of the encyclopedial chapter by researching significance of individual structural elements of an encyclopedical chapter for the development of ontology.

The fourth chapter is an introduction to the development of the ontological model of literature and to the basic settings of Protégé software. Elements of standards and ontologic languages (i.e. vocabularies RDFS, OWL, SKOKS) are shown and applied in the development of ontology of Croatian Encyclopedia in the field of Croatian literature. Pointed out was the possibility of reaching interoperability inasmuch individual ontologic resources, as well as overall ontology, with existing semantic ontologic projects on the Semantic Web which will allow connecting relevant and rich knowledge of *Croatian Encyclopedia* into an „global network of knowledge“ which appears and develops through projects of Semantic Web.

The fifth chapter gave an insight into the historical development of encyclopedia in the world so people could completely realize the context through which encyclopedia had to go through in order to gain today's familiar features of a modern encyclopedic work. The chapter has given basic information about the development of the Croatian central lexicographical institution; Lexicographical institution „Miroslav Krleža“ (LZMK) that does lexicography and encyclopedics of particular interest for the Republic of Croatia. It is shown that this doctoral thesis can contribute to ,with some particular elements , realising the mission and vision of LZMK. Publishing work of basic and expert encyclopedic editions of LZMK is shown, embracing editions on a paper media and those in a web surrounding. By analysing publishing work of web editions, it was found that their substantiation was mostly a matter of adapting traditional organisations of encyclopedic knowledge from paper media to web surroundings. That is why this doctoral thesis will research which uses from applying

ontologic principles of semantic web would LZMK have, as well as users of these valuable knowledge sources.

The sixth chapter has shown a structural organisation of encyclopedical articles, looking back onto basic types of encyclopedic articles and especially on features of biographic encyclopedial article that is the basis of research of this work. An analysis of type of data which contain encyclopedical biographic articles has been done. The mentioned allowed establishing basic ontologic layers i.e. facets by which ontologic relations will classify.

The seventh chapter identifies constructive ontologic elements from encyclopedical biographic article, shows the methodology used in development of ontology, as well as resulting conceptual taxonomy and ontologic relations. The chapter has shown the role of structural elements of encyclopedical biographic article during ontologic development and connected them to corresponding constructive ontologic elements. The results of the research are presented through display of ontologic modules and belonging ontologic relations that can be used in describing a certain term. A display of structure and sequence of elements of encyclopedical biographic article has been shown, with developed ontologic features by which can all types of information in the article be stored, from which constant elements have been determined that can be perceived for development of article infoframes, which is suitable for a quick insight into most important information of individual articles. Final results are shown through application of the gained ontology in description of an individual, as well as through possibility of installing complex semantic questions by unstructured data of encyclopedical biographic articles and through the possibility of organising browsing encyclopedical data, which hasn't been possible until now.

The eight chapter explained the FCA approach applied during building of ontology so a gathering of conceptual features would be established, by which terms in ontology were defined so classification of terms could be made into a hierarchy. Important definitions were emphasized to understand places of formal term analysis in methodology of creating ontology. An actual example of accomplishing FCA approach was shown in 37 articles of Croatian literature in Croatian Encyclopedia, as well as a transformation of a transformational grid into a formal language of first order logics. Chapter shows the advantages of applying FCA analysis because of generating new and unfamiliar terms which could be hardly established only by handiwork of ontology, because texts specific to the domain do not include any kind of noun phrase for labeling these new terms.

The ninth chapter brings out problems of automated indexing methods and information fetching. Theoretically, it shows the LSA method and its application on the example of encyclopedical articles of Croatian Encyclopedia with the goal of learning about its efficiency and goals in building ontology of an certain area. Conducting LSA method on chosen articles shows its utility in assessing the term as a gathering of related terms and affiliations of certain entries of articles (documents) in that matter, by which on the basis of word forms that selected articles consist of allows an automatic classification of articles by individual ontologic classes.

The tenth, also final, chapter of this work is a conclusion which combines theoretical and practical part of the work by giving a short review of research results and showing the possibility of establishing interoperability and connecting *Linked Data* concept of Croatian encyclopedistics with other structural sources of encyclopedical knowledge on the web (e.g. DBpedia, Freebase, etc.).

Keywords: encyclopedia, Croatian Encyclopedia, biographical article, ontology, semantic web, FCA, LSA

Sadržaj

1	UVOD	1
1.1	OSNOVNA POLAZIŠTA RADA	1
1.2	CILJEVI I HIPOTEZA ISTRAŽIVANJA	3
1.3	METODOLOGIJA.....	5
1.4	STRUKTURA RADA.....	6
2	TEORIJSKE OSNOVE SEMANTIČKOG WEBA	11
2.1	SEMANTIČKI WEB	11
2.2	POJMOVI I ARHITEKTURA SEMANTIČKOG WEBA	14
2.3	ONTOLOGIJA	17
2.4	JEZICI SEMANTIČKOG WEBA	21
2.4.1	<i>RDF</i>	21
2.4.2	<i>RDF Shema (RDFS)</i>	33
2.4.2.1	Klase <i>RDF Sheme</i>	35
2.4.2.2	<i>RDF</i> svojstva i atributi.....	36
2.4.3	<i>OWL</i>	39
2.4.3.1	<i>OWL1</i> jezik	43
2.4.3.1.1	Definiranje klasa	43
2.4.3.1.2	Ograničenja svojstava	44
2.4.3.1.3	Složene klase	46
2.4.3.1.4	Svojstva <i>OWL 1</i> jezika	48
2.4.3.2	<i>OWL2</i> jezik	51
2.4.4	<i>SKOS</i>	55
2.4.4.1	Konceptualna klasa (engl. <i>Concept Class</i>).....	56
2.4.4.2	Svojstva za označavanje	56
2.4.4.3	Konceptne sheme (engl. <i>Concept Schemes</i>)	59
2.5	ZAKLJUČAK.....	62
3	DOSADAŠNJI ONTOLOGIJSKI PROJEKTI	63
3.1	UVOD	63
3.2	ONTOLOGIJSKI PROJEKTI	64
3.2.1	<i>SUO</i>	64
3.2.2	<i>SUMO</i>	65
3.2.3	<i>UMBEL</i>	68
3.2.4	<i>Cyc</i>	70
3.2.5	<i>BabelNet</i>	72
3.2.6	<i>CIDOC</i>	73
3.2.7	<i>WordNet</i>	76
3.2.8	<i>Schema.org</i>	83
3.2.9	<i>PROTON</i>	85
3.3	ENCIKLOPEDIJSKI ONTOLOGIJSKI PROJEKTI	89
3.3.1	<i>DBpedia</i>	89
3.3.2	<i>Freebase</i>	95
3.3.3	<i>YAGO</i>	96
3.3.4	<i>YAGO2</i>	99

3.3.5	YAGO3	103
3.3.6	InPhO.....	106
3.4	ZAKLJUČAK.....	112
4	UVOD U ONTOLOGIJSKI MODEL KNJIŽEVNOSTI I PROTEGE PROGRAM	113
4.1	UVOD	113
4.2	DEFINIRANJE DOMENE I PREFIKSA	113
4.3	DEFINIRANJE TAKSONOMIJE KLASA.....	114
4.4	IZGRADNJA INDIVIDUA	116
4.5	DEFINIRANJE ODNOSA	118
4.5.1	<i>Objektni i podatkovni odnosi.....</i>	<i>118</i>
4.5.2	<i>Opisni odnosi</i>	<i>125</i>
4.6	JEDNAKOST I NEJEDNAKOST	130
4.7	VERIFIKACIJA ONTOLOGIJE	131
4.8	ZAKLJUČAK.....	132
5	ENCIKLOPEDIČKA DJELA.....	133
5.1	POVIJESNI RAZVOJ ENCIKLOPEDIJE U SVIJETU.....	133
5.2	LEKSIKOGRAFSKI ZAVOD MIROSLAV KRLEŽA	143
5.3	HRVATSKA OPĆA ENCIKLOPEDIJA.....	146
5.4	OSTALA ENCIKLOPEDIJSKA IZDANJA LZMK	151
5.5	ZAKLJUČAK.....	156
6	ENCIKLOPEDIČKA ORGANIZACIJA ZNANJA	157
6.1	UVOD	157
6.2	ENCIKLOPEDIJSKI ČLANAK	159
6.3	VRSTE ENCIKLOPEDIJSKIH ČLANAKA.....	165
6.4	STRUKTURA I ANALIZA VRSTE PODATAKA BIOGRAFSKOG ENCIKLOPEDIJSKOG ČLANKA.....	167
6.5	ZAKLJUČAK.....	171
7	RAZVOJ ONTOLOGIJE KNJIŽEVNOSTI.....	172
7.1	UVOD	172
7.2	GRADIVNI ONTOLOGIJSKI ELEMENTI ENCIKLOPEDIJSKOG BIOGRAFSKOG ČLANKA IZ PODRUČJA KNJIŽEVNOSTI.....	173
7.3	METODOLOGIJA.....	175
7.3.1	<i>Određivanje ontologijske domene i dosega.....</i>	<i>179</i>
7.3.2	<i>Upotreba postojeće ontologije</i>	<i>181</i>
7.3.3	<i>Označavanje važnih termina u ontologiji.....</i>	<i>181</i>
7.3.4	<i>Definiranje ontologijskih klasa i klasne hijerarhije (pojmovne taksonomije).....</i>	<i>184</i>
7.3.4.1	Agent modul	186
7.3.4.2	Književno djelo modul	189
7.3.4.3	Književni rodovi i vrste modul.....	190
7.3.4.4	Modul Periodizacija književnosti	195
7.3.4.5	Modul Dokument.....	196
7.3.5	<i>Određivanje ontologijskih odnosa.....</i>	<i>199</i>
7.3.5.1	Objektni odnosi	199
7.3.5.2	Svojstva osobe.....	200
7.3.5.3	Svojstva autora	217
7.3.5.4	Svojstva djela	224
7.3.5.5	Svojstva izdavača	234
7.3.5.6	Svojstva za ontologijsko zaključivanje.....	234

7.3.5.7	Podatkovni odnosi	236
7.3.5.8	Opisna svojstva.....	242
7.3.6	<i>Definiranje ograničenja nad relacijama</i>	253
7.3.7	<i>Domena i doseg svojstva</i>	254
7.3.8	<i>Stvaranje individua</i>	255
7.4	REZULTATI.....	257
7.4.1	<i>Prikaz ontologijskih modula s pripadajućim svojstvima</i>	257
7.4.2	<i>Enciklopedijski članak i ontologija</i>	262
7.5	PRIMJENA ONTOLOGIJE.....	266
7.5.1	<i>Primjena ontologije hrvatske književnosti za različite razine pretraživanja</i>	266
7.5.2	<i>Primjena ontologije hrvatske književnosti za pregledavanje informacija</i>	272
7.6	ZAKLJUČAK.....	283
8	FORMALNA POJMOVNA ANALIZA – FCA	284
8.1	UVOD	284
8.2	PREGLED I PRIMJENA FCA.....	285
8.3	PROCES STVARANJA ONTOLOGIJE UPOTREBOM FCA.....	289
8.4	PRIMJER STVARANJA TAKSONOMIJE ONTOLOGIJE NA PODKORPUSU HRVATSKE KNJIŽEVNOSTI IZ HRVATSKE ENCIKLOPEDIJE.....	291
8.4.1	<i>Ekstakcija formalnog konteksta</i>	291
8.4.2	<i>Redukcija informacija</i>	296
8.4.3	<i>Imenovanje pojmova</i>	297
8.4.4	<i>Model logike prvog reda pojmovne rešetke</i>	298
8.4.4.1	Izgradnja jezika	298
8.4.4.2	Transformacija pojmovne rešetke u jezik pravila	299
8.4.4.3	Pravila za aksiome rešetke.....	299
8.4.4.4	Pravila za operacije rešetke	300
8.4.4.5	Predstavljanje ne-taksonomijskih odnosa upotrebom pravila	302
8.4.4.6	Zaključivanje o ontologiji.....	303
8.5	ZAKLJUČAK.....	304
9	LATENTNA SEMANTIČKA ANALIZA	305
9.1	PROBLEMI METODA AUTOMATSKOG INDEKSIRANJA I PRETRAŽIVANJA INFORMACIJA.....	305
9.2	IZRAČUNAVANJE TEŽINE TERMINA	308
9.2.1	<i>Indeksni termini ili ključne riječi</i>	308
9.2.2	<i>Prikaz termina u dokumentu (engl. Term-Document Representation)</i>	309
9.2.3	<i>Težina termina</i>	311
9.2.4	<i>Termin – termin korelacije</i>	313
9.2.5	<i>Frekvencija termina</i>	314
9.2.6	<i>Inverzna frekvencija dokumenta</i>	316
9.2.7	<i>Frekvencija termina i inverzna frekvencija dokumenta - TF-IDF</i>	318
9.3	PRONALAZENJE SLIČNIH DOKUMENATA	320
9.3.1	<i>Model vektorskog prostora (eng. Vector Space Model)</i>	322
9.3.1.1	Latentno semantičko indeksiranje	326
9.3.1.1.1	Upotreba	326
9.3.1.1.2	SVD.....	329
9.3.1.1.3	LSI u izgradnji ontologije hrvatske književnosti na korpusu Hrvatske enciklopedije	332
9.3.1.1.3.1	Uvod.....	332
9.3.1.1.3.2	Primjena LSI na enciklopedijskim člancima	334
9.3.1.1.3.3	Interpretacija i vizualizacija rezultata	338
9.4	ZAKLJUČAK.....	344

10	ZAKLJUČAK.....	345
11	LITERATURA.....	367
12	POPIS TABLICA.....	392
13	POPIS SLIKA	393
14	ŽIVOTOPIS.....	396
15	BIBLIOGRAFIJA.....	397

1 UVOD

1.1 Osnovna polazišta rada

Svjetska globalizacija ne događa se samo u gospodarskom i političkom smislu, već također i u području informacija i znanja. To se prije svega odnosi na razvitak interneta – globalne mreže, koji od povezivanja dokumenata prelazi na povezivanje podataka, tj. ubrzano se transformira u semantički web. Tehnički rečeno, dosadašnji WEB portali s klasičnim bazama informacija i znanja, trebaju postajati povezani podaci (engl. linked data) globalnog oblaka (engl. cloud computing), što se provodi ontologijskim pristupom i novim alatima.

Glavna trenutna prepreka bolje potpore korisnicima je što značenje sadržaja nije računalno razumljivo. Semantička interoperabilnost podrazumijeva postojanje infrastrukture koja će omogućiti strojnu interpretaciju i zaključivanje o sadržajima na webu. Stoga, ključni pojam semantičkog web-a predstavlja ontologija, temeljna komponenta u omogućavanju semantičke interoperabilnosti. Pojam ontologije vezuje se za mogućnost ponovne upotrebe i dijeljenja znanja u nekoj domeni. U svrhu predstavljanja određene domena ontologijom potrebno je obuhvatiti pojmove koji je opisuju i razviti njihovu taksonomiju, attribute pojmova i odnose među pojmovima. Ontologija omogućava preciznu i efikasnu komunikaciju značenja jer je ona ustvari manifestacija zajedničkog dogovorenog razumijevanja domene između više agenata, donoseći prednosti kao što su interoperabilnost, ponovno korištenje i dijeljenje znanja.

Osim uspostavljanja temeljnih postavki razvoja ontologije određene domene, rad će na temelju strukturalne organizacije enciklopedičkih članaka tj. dosadašnje tradicionalne enciklopedičke organizacije znanja dati smjernice za poboljšanje i daljnji razvoj organizacije enciklopedičkog znanja u mrežnom okruženju imajući u vidu njegove specifičnosti i daljnji razvoj prema semantičkom webu.

Razvoj semantičkog weba i ontologijskog pristupa je utjecao i na enciklopedičku organizaciju znanja dajući joj sasvim nove dimenzije. Prednost semantičkog weba i ontologijskog pristupa u organizaciji enciklopedičkog znanja je već prepoznata i primijenjena u velikim enciklopedijskim projektima svjetskih razmjera koje će ovaj rad prikazati. Međutim, ni jedan od tih projekata nije proučavao isključivo strukturu određenog tipa

enciklopedičkog članka kao osnovu razvoja ontologije. Upravo će navedeni pristup biti uže područje ovog rada jer će se utvrditi postoji li poveznica između strukture enciklopedičkog biografskog članka i gradivnih ontologijskih elemenata kako bi se razvoj ontologije određene domene mogao temeljiti na strukturi enciklopedičkog biografskog članka.

Enciklopedička djela su uvijek posjedovala organiziranost (strukturiranost) znanja kojeg prezentiraju. Upravo ta organiziranost (strukturiranost) znanja u enciklopedičkim djelima omogućuje korisniku brzo i lako dolaženje do željene informacije. Enciklopedička organizacija znanja tradicionalnih enciklopedičkih djela rezultirala je njihovim ujednačenim osnovnim svojstvima, načinom izlaganja i organizacijom sadržaja. Pretraživanje željene informacije u enciklopedičkim djelima omogućen je nelinearnom organizacijom, kojom je znanje podijeljeno na manje, pregledne dijelove. Korisnik pristupa znanju poštujući metodički slijed kojim je ono nanizano. Nadalje, slijedeći uputnice kojima su dijelovi znanja povezani korisnik sam određuje slijed sadržaja prema svom interesu. Organizacijska struktura tradicionalne enciklopedije zasniva se na procesu:

- razdioba nepregledne količine znanja u pregledne dijelove tj. članke, a ukoliko su oni još uvijek odveć veliki (nepregledni), dodatno se dijele na naslovljena poglavlja,
- dodjeljivanje naziva člancima – natuknica ili lema, kako bi se traženi dijelovi znanja (članci) mogli pronaći,
- nizanje članaka nekim logičkim slijedom: abecedni redoslijed, konceptualni redoslijed (sustavni, tematski red ili poredak), kronološki redoslijed (vremenski red ili poredak), prostorni redoslijed (red ili poredak) , mješoviti redoslijed,
- uspostavljanje veza među člancima (umrežavanje članaka),
- izrada indeksa.

Usporedbom prethodno navedene organizacijske strukture tradicionalnih s elektroničkim enciklopedijama u web okruženju uočavaju se određene razlike proizašle iz različitih vrsta medija koje omogućuju nove funkcionalnosti pretraživanja. Organizacija mrežnih enciklopedija u počecima svojega razvoja počiva na hipertekstu, za čiju se pretaču na određeni način mogu smatrati upravo enciklopedička i leksikografska djela jer se sadržaj u njima organizira podjelom na dijelove međusobno povezane uputnicama što sasvim nalikuje hipertekstu. U prvim hipertekstovnim enciklopedijama sredinom 1990-ih do željenog članka se pristupalo odabirom sve užega područja kojem članak pripada, odnosno pregledavanjem informacija (engl. *browsing*). Ubrzo se elektroničkim enciklopedijama dodaje mogućnost

pretraživanja čitavoga teksta, popisa članaka ili digitaliziranoga indeksa prema upitu korisnika uspostavljanjem sustava za pretraživanje podataka (engl. *information retrieval*), odnosno tražilica.

Dakle, ukoliko se usporede tradicionalna enciklopedička djela izdavana na papirnatom mediju s elektroničkim enciklopedijama u mrežnom okruženju može se uočiti da je *hipertekstovnost* omogućila zamjenu uputnica elektroničkim poveznicama i olakšala kretanje među sadržajem; *pretraživost* je zamijenila abecedni slijed mogućnošću pretrage naslova članaka ili čitavoga teksta; *neograničenost opsega* je omogućila širenje broja i duljine članaka; *poveznice* su omogućile povezivanje pojedinih članaka na vanjske izvore (npr. digitalizirane sadržaje arhiva, knjižnica, muzeja); *kolaborativnost* je omogućila izravno sudjelovanje korisnika u stvaranju, dopunjavanju i ispravljanju enciklopedičkog djela. Uočljivo je kako su enciklopedijskim djelima tradicionalnoga tipa bili potrebni abecedariji kao svojevrsni kontrolirani rječnici utvrđeni prije početka posla na samim člancima, dok se kod enciklopedijskih djela virtualnoga tipa ističe potreba i za klasifikacijskim sustavom složenije razine, taksonomijom, tezaurusom i ontologijom.

Moderna tehnologija utječe na enciklopedijska djela u procesu njihove izradu, rasparčavanja, korištenja, struktiranosti i pohrane, te istovremeno autorima, stvarateljima i nakladnicima pruža nepojmljive mogućnosti oblikovanja i prikaza sadržaja. Navedene promjene kroz koje prolazi enciklopedičko djelo zahtijeva uspostavljanje načina modeliranja organizacije enciklopedičkog znanja kako bi se osigurala njihova dostupnost i korištenje iz čega proizlazi i potreba uspostavljanja novog načina modeliranja enciklopedičke organizacije koje će svoje utemeljenje pronaći na analizi specifičnosti enciklopedičkog članka, uvažavanjem temeljnih postavki semantičkog weba, principa ontologijskog oblikovanja i potreba korisnika.

1.2 Ciljevi i hipoteza istraživanja

Određene vrste enciklopedijskih članaka koji obrađuju pojedinu struku ili tematsko područje organizirani su na sličan način te njihova struktura treba biti jednako provedena kod svih članaka nekog enciklopedičkog djela.

Cilj ove doktorske disertacije je istražiti koji elementi enciklopedičke organizacije znanja mogu pružiti podršku za razvoj ontologije te razviti metodu kojom će se generirati ontologija na osnovi enciklopedički organiziranog znanja. Na temelju korpusa odabranih

članaka *Hrvatske enciklopedije*¹ Leksikografskog zavoda Miroslav Krleža iz područja hrvatske književnosti i utvrđenih strukturnih elemenata biografskih članaka razviti će se ontologija kojom će se opisati i predstaviti znanje obrađeno u odabranim člancima. Proučavanjem literature, analizom sličnih ontologijskih modela odabranih svjetskih projekata, strukture i znanja pohranjenog u biografskim člancima područja hrvatske književnosti predložit će se ontologijski model koji će na djelotvoran način opisati enciklopedičko znanje navedenog područja Hrvatske enciklopedije. Ovaj glavni cilj se može podijeliti na podciljeve kojima je zajednička svrha postići zadani glavni cilj ovog rada:

- 1) Proučiti relevantnu postojeću literaturu o tematskim područjima bitnim za ostvarivanje glavnog cilja ovog rada s ciljem stjecanja znanja koja će se primijeniti prilikom izrade ontologijskog modela iz enciklopedički organiziranog znanja Hrvatske enciklopedije za područje hrvatske književnosti. Posebna pažnja će se obratiti na proučavanje tematskih područja: semantički web, ontologijski gradivni elementi i njihov razvoj, enciklopedička organizacija znanja, strukturalna organizacija biografskog enciklopedijskog znanja.
- 2) Proučiti i sustavno prikazati dostignuća na području razvoja ontologija u informacijskim znanostima i njihove primjene dobre prakse u svjetskim razmjerima, s naglaskom na problemsku domenu disertacije.
- 3) Iz enciklopedičkih biografskih članaka odabranog područja hrvatske književnosti istražiti i izdvojiti pojmove (entitete), attribute i njihove međusobne odnose.
- 4) Prilikom izgradnje ontologije primijeniti FCA pristup (engl. Formal Concept Analysis, njem. *Formale Begriffsanalyse* – FBA) kako bi se ustanovio skup pojmovnih svojstava kojima su definirani pojmovi u ontologiji te klasificirao pojam s ostalima u hijerarhiju.
- 5) Razviti ontologiju koja će predstavljati odgovarajući rječnik za odabrano područje književnosti Hrvatske enciklopedije.

¹ Članci su preuzeti iz Hrvatska enciklopedija, mrežno izdanje. / glavni urednik Slaven Ravlić. Zagreb: Leksikografski zavod Miroslav Krleža, 2016. Mrežna *Hrvatska enciklopedija* temelji se na knjižnom (tiskanom izdanju) koje je objavljeno u 11 svezaka od 1999. do 2009. godine, glavni urednici kojega su bili Dalibor Brozović (1-3. svezak), August Kovačec (4-7. svezak) i Slaven Ravlić (8-11. svezak). Pošto je 2009. završena knjižna enciklopedija u 11 svezaka, odlučeno je da se daljnji u okviru *Hrvatske enciklopedije* nastavi u obliku računalne baze iz koje bi za nekoliko godina moglo proizaći mrežno izdanje enciklopedije te druga elektronička i multimedijaska izdanja. U jesen 2011. započeo je rad na projektu mrežnog izdanja revizijom abecedarija i pisanjem novih članaka, a u 2012. i ažuriranje i dopune tekstova u računalnoj bazi.

- 6) Protégé softwareom osigurati ontologiju unutar promatrane domene istraživačkog rada, koja će biti korisna kako ljudima, tako i računalima za pristupanje i upravljanje enciklopedičkim informacijama.
- 7) Omogućiti automatsku klasifikaciju članaka pojedinim ontologijskim klasama primjenom latentne semantičke analize (LSA) na temelju razvijene ontologije.

Hipoteza rada glasi: Enciklopedička organizacija znanja može kvalitetno poslužiti razvoju ontologije, koja je temelj interoperabilnosti na semantičkom webu. Na temelju strukture biografskih enciklopedističkih članaka iz područja hrvatske književnosti Hrvatske enciklopedije može se definirati nužan i dovoljan broj pojmova i njihovih međusobnih odnosa koji čine rječnik (ontologiju) za predstavljanje znanja o odabranom području.

1.3 Metodologija

Kako bi se postigli ciljevi ove doktorske disertacije, proučit će se i proanalizirati literatura vezana uz tematiku ovog rada, posebice literatura o semantičkom webu, razvoju ontologija i strukturnim elementima enciklopedijskog članka.

Nakon faze proučavanja literature, pristupit će se analizi postojeće prakse razvoja ontologijskih projekata na enciklopedičkom znanju. Proučit će se i temeljne specifičnosti razvoja ontologija u postojećim projektima te će se prikupljena saznanja iskoristiti pri razvoju ontologije iz Hrvatske enciklopedije za područje hrvatske književnosti.

Slijedi faza sintetiziranja znanja stečenih proučavanjem literature o relevantnim temama i upoznavanja s najboljom praksom odabranih ontologijskih projekata te primjena stečenih spoznaja u razvoju ontologije iz enciklopedički organiziranog znanja. U razvoju ontologije biografskih enciklopedijskih članaka iz područja hrvatske književnosti primijenjena je metodologija analize sadržaja odabranih članaka opće enciklopedije te METHONTOLOGY metoda. Upotrijebit će se Protége software za razvoj ontologije. Uzorak će uključiti korpus Hrvatske enciklopedije tj. stručno područje k10 - hrvatska književnost. Korpus sadrži 1170 enciklopedijskih članaka, koji su ujednačeni u pogledu slijeda podataka o životu osobe i opsega koji zauzimaju imajući u vidu značaj osobe koja se u njima obrađuje. U skladu s ciljem disertacije ontologija biografskih enciklopedijskih članaka Hrvatske enciklopedije iz područja hrvatske književnosti će se razviti njihovom analizom,

identifikacijom termina i odnosa među njima te definiranjem vrsta entiteta koji čine ontologiju. Metodologijski koraci primijenjeni u razvoju ontologije su sljedeći:

1. Identifikacija postojeće ontologije najviše razine apstrakcije je provedena analizom postojećih istraživačkih projekata te je kao najprikladnija odabrana PROTON ontologija.
2. Identificiranje ključnih pojmova u promatranom području – izvršit će se analizom odabranog korpusa tj. biografskih enciklopedijskih članka
3. Stvaranje taksonomije za pojmovne rječnike – primjenit će se pristup od *vrha prema dnu* (engl. top-down) koji počinje s definicijom najopćenitijih pojmova u domeni, a zatim slijedi specijalizacija pojmova.
4. Testiranje za opis znanja izvršeno je preslikavanjem instanci u taksonomiju rječnika, što je izvršeno stvaranjem individua koje se opisuju stvorenim ontologijskim rječnikom.

U cilju uspješnog izvršenja svih prethodno navedenih koraka provedeni su postupci kojima se određuje ontologijska domena i doseg, razmatra se ponovna upotreba neke već postojeće ontologije, određuju se važni termini tj. stvara se terminološki rječnik, definiraju se ontologijske klase i klasna hijerarhija FCA-a pristupom, određuju se ontologijski odnosi i stvaraju se individue.

Na temelju ontologije provest će se LSA metoda kako bi se uočila njezina korisnost u određivanju pojma kao skupa srodnih termina te pripadnost pojedinih natuknica članaka (dokumenata) tom pojmu, čime se omogućuje automatska klasifikacija članaka pojedinim ontologijskim klasama. Ovim pristupom se vrši automatsko klasificiranje članaka na temelju srodnih termina.

Prilikom izrade ove doktorske disertacije koristit će se, dakle, metoda analize, sinteze, apstrakcije, konkretizacije, deskripcije, komparacije, promatranja te metoda modeliranja.

1.4 Struktura rada

Drugo poglavlje ovog rada bavi se teorijskim prikazom i pojašnjenjem semantičkog weba u svrhu njegovog potpunog razumijevanja. Cilj poglavlja je ukazati na osnovnu teorijsku i tehničku pozadinu semantičkog weba. Poglavlje ukazuje na značenje pojma semantički web, temeljnu razliku između weba kakvoga danas poznajemo i njegovog razvoja

prema semantičkom webu, nedostatke današnjeg weba i prednosti semantičkog weba, temeljne pojmove i arhitekturu semantičkog weba, pregled temeljnih ontologijskih definicija, glavnu svrhu i ulogu ontologija na semantičkom webu. Poglavlje daje detaljan pregled osnova jezika primjenjivanih na semantičkom webu uz konkretne primjere. Temeljito su prikazani RDF, RDFS, OWL i SKOS. Poglavlje predstavlja osnovu za dobro razumijevanje primjene ontologija u informacijskim sustavima jer je prvo potrebno razumjeti zamisao semantičkog weba na koji se one oslanjaju.

U trećem poglavlju rada prikazani su neki od važnijih projekata semantičkog weba vezani uz razvoj značajnih ontologija na semantičkom webu. Prilikom izbora značajnih projekata se vodilo računa da se izaberu oni projekti koji su značajni za bolje razumijevanje semantičkog weba na području enciklopedistike. Stoga se jedan dio ovog poglavlja bavi prikazom ontologija nezaobilaznih za bolje razumijevanje semantičkog weba, a drugi dio poglavlja daje osvrt na ontologijske projekte izrasle isključivo na enciklopedijskom znanju. Navedeno poglavlje daje prikaz razvoja značajnih ontologija na semantičkom webu, kao i prikaz razvoja ontologija temeljen isključivo na enciklopedijskom znanju. Iznese činjenice o prikazanim enciklopedijskim ontologijskim projektima u ovom poglavlju ukazuju na to da su navedene ontologije razvijane na temelju strukturiranih podataka enciklopedijskog članka, doprinosa stvaranja strukturiranih podataka od strane samih korisnika iz različitih mrežnih izvora te kombinacije automatskih metoda ekstrakcije i povratne informacije autora i urednika određene domene. Navedena analiza postojećih enciklopedijskih ontologijskih projekata je ukazala na to kako niti jedan dosadašnji projekt nije pokušao povezati i istražiti razvoj ontologije i njezinih gradivnih elemenata na temelju strukturalne organizacije samog enciklopedijskog članka istražujući značaj pojedinih strukturalnih elemenata enciklopedijskog članka za razvoj ontologije, a ne samo značaj strukturiranih podataka enciklopedijskog članka. Poglavlje je ukazalo na nužnost proučavanja te neistražene veze strukturalne organizacije enciklopedijskog članka i razvoja ontologije, te će se upravo tom navedenom problematikom baviti ovaj rad.

Četvrto poglavlje je uvod u razvoj ontologijskog modela književnosti, kao i uvod u temeljne postavke Protégé software. Prikazani su elementi standarda i ontologijskih jezika primijenjenih u razvoju ontologije Hrvatske enciklopedije iz područja hrvatske književnosti, temeljenoj na biografskim enciklopedijskim člancima. Primjena navedenih gradivnih elemenata standarda i ontologijskih jezika upotrijebljenih u razvoju ontologije je prikazana kroz konkretne primjere kako bi se dobio uvod u primijenjene postupke izrade ontologije pomoću Protégé software-a. Poglavlje je dalo kratki uvod u primijenjene standarde

semantičke mreže (tj. rječnike RDFS, OWL, SKOS) prilikom izgradnje ontologije Hrvatske enciklopedije za područje hrvatske književnosti, čime će se dobivena ontologija uklopiti i učiniti sastavnim djelom semantičke mreže podataka. Ukazalo se na mogućnost postizanja interoperabilnosti kako ontologijskih resursa pojedinačno, tako i cjelokupne ontologije sa velikim postojećim semantičkim ontologijskim projektima. Uspostavljanje interoperabilnosti dobivene ontologije s ostalim projektima na semantičkoj mreži omogućit će povezivanje relevantnog i bogatog znanja *Hrvatske enciklopedije* u „globalnu mrežu znanja“ koja nastaje i razvija se kroz projekte semantičke mreže.

Peto poglavlje je dalo uvod u povijesni razvoj enciklopedije u svijetu kako bi se u potpunosti shvatio kontekst kroz koji je enciklopedija morala proći da bi zadobila današnje poznate odlike suvremenog enciklopedičkog djela. Poglavlje je dalo osnovne informacije o razvoju hrvatske središnje leksikografske ustanove, Leksikografskom zavodu „Miroslav Krleža“ (LZMK), koji obavlja djelatnost leksikografije i enciklopedike od osobitog interesa za Republiku Hrvatsku. Ukazalo se kojim pojedinim elementima u ostvarivanju misije i vizije LZMK može pridonijeti ova doktorska radnja. Prikazana je izdavačka djelatnost općih i strukovnih enciklopedijskih izdanja LZMK, obuhvaćajući izdanja na papirnatom mediju i ona u mrežnom okruženju. Analizom izdavačke djelatnosti mrežnih izdanja, utvrđeno je kako se u njihovom ostvarivanju uglavnom radi o prilagodbi tradicionalne organizacije enciklopedičkog znanja s papirnatog medija u mrežno okruženje. Upravo će ova doktorska radnja istražiti koje bi koristi od primjene ontologijskih principa semantičkog weba imao LZMK, kao i sami korisnici ovih vrijednih izvora znanja.

Šesto poglavlje je prikazalo strukturalnu organizaciju enciklopedijskog članka, osvrnulo se na osnovne vrste enciklopedijskih članaka, a posebno na odlike biografskog enciklopedijskog članka koji je temelj proučavanja ovog rada. Izvršena je analiza vrste podataka koje sadrže enciklopedijski biografski članci. Teoretskim opisom strukture i vrsta enciklopedijskih članaka prikazala se postojanost i definiranost ujednačene strukture enciklopedijskih članaka kojih se pridržavaju leksikografi Leksikografskog zavoda Miroslav Krleža prilikom stvaranja Hrvatske enciklopedije. Prikazom odlika enciklopedijskih članaka (posebno biografskih enciklopedijskih članaka) te ukazivanjem na vrstu i oblik podataka koje sadrže biografski članci uspostavljeni su temeljni ontologijski slojevi tj. facete prema kojima će se razvrstavati ontologijski odnosi. Utvrđen je i oblik podataka za pojedine vrste informacija što će se primijeniti prilikom razvoja ontologije Hrvatske enciklopedije za područje hrvatska književnost.

Sedmo poglavlje identificira gradivne ontologijske elemente iz enciklopedijskog biografskog članka, prikazuje metodologiju primijenjenu u razvoju ontologije, kao i rezultirajuću pojmovnu taksonomiju i ontologijske odnose. Dobiveni rezultati istraživanja su predstavljeni i kroz prikaz ontologijskih modula i pripadajućih im ontologijskih odnosa koji se mogu upotrijebiti u opisu određenog pojma. Dobiveni su svi ontologijski odnosi za opis biografskih enciklopedijski članak iz područja hrvatska književnost. Izvršen je prikaz strukture i slijeda elementa enciklopedijskog biografskog članka s razvijenim ontologijskim svojstvima kojima se mogu opisati sve vrste informacija pohranjene u članku, iz čega su utvrđeni stalni elementi koji se mogu sagledati za razvoj infookvira članaka koji su prikladni za brzi uvid u najbitnije informacije o pojedinom članku. Konačni rezultati su prikazani kroz primjenu tako dobivene ontologije u opisu individue, kao i kroz mogućnost postavljanja složenih semantičkih pitanja nestrukturiranim podacima enciklopedijskih biografskih članaka s detaljnijom razinom opisa i dobivenih odgovora te mogućnost organiziranja pregledavanja (engl. browsing) enciklopedičkog znanja, što do sada nije bilo moguće. Poglavlje je ukazalo na ulogu strukturnih elemenata enciklopedijskog biografskog članka prilikom ontologijskog razvoja te ih je povezalo s odgovarajućim gradivnim ontologijskim elementima. Istražilo se i definiralo koji su enciklopedijski elementi članka značajni za razvoj ontologijskih odnosa ili pojmova. Rezultat je ontologija koja je u mogućnosti opisati enciklopedijsko znanje navedenog područja. Sve navedeno je doprinijelo poboljšanju prikaza i pretraživanje enciklopedijskih članka u mrežnom okruženju.

Osmo poglavlje je pojasnilo FCA pristup (engl. *Formal Concept Analysis*, njem. *Formale Begriffsanalyse* – FBA, hrv. *formalna pojmovna analiza*) primijenjen prilikom izgradnje ontologije kako bi se ustanovio skup pojmovnih svojstava kojima su definirani pojmovi u ontologiji te kako bi se izvršila klasifikacija pojmova u hijerarhiju. Poglavlje o formalnoj pojmovnoj analizi je prvenstveno istaknulo bitne definicije za razumijevanje mjesta formalne pojmovne analize u metodologiji izgradnje ontologije. Osim teoretskog objašnjenja same metoda pojmovne analize i procesa njezine primjene prilikom izrade ontologije, prikazan je i konkretan primjer izvršavanja FCA pristupa na 37 članaka područja hrvatske književnosti iz Hrvatske enciklopedije, kao i transformacija pojmovne rešetke u formalni jezik logike prvog reda. Poglavlje ukazuje na prednost primjene FCA analize zbog generiranja novih i nepoznatih pojmova koji bi se teško ustanovili samo ručnom izradom ontologije jer tekstovi specifični za domenu ne uključuju bilo kakvu imensku frazu za označavanje ovih novih pojmova.

Deveto poglavlje iznosi probleme metoda automatskog indeksiranja i pretraživanja informacija, definira i objašnjava osnovne pojmove poput *indeksni termini*, *prikaz termina u dokumentu* i mjerenje *težine termina* u zbirci, mjere pronalaženja sličnih dokumenata, opći model vektorskog prostora. Poglavlje teorijski prikazuje LSA metodu i njezinu primjenu na primjeru enciklopedijskih članaka iz područja hrvatske književnosti Hrvatske enciklopedije kako bi se spoznala učinkovitost i svrha LSA metode u izgradnji ontologije određenog područja. Provedba LSA metode na odabranim dokumentima ukazuje na njezinu korisnost u određivanju pojma kao skupa srodnih termina te pripadnost pojedinih natuknica članaka (dokumenata) tom pojmu, čime se omogućuje automatska klasifikacija članaka pojedinim ontologijskim klasama. Ovim pristupom određivanja pripadnosti pojedinih termina pojedinom pojmu omogućena je automatska klasifikacija članaka pojedinoj ontologijskoj klasi na temelju obrazaca riječi koje sadrže odabrani članci, što je prikazano na konkretnom primjeru.

Deseto i zadnje poglavlje ovog rada je skupni zaključak koji objedinjuje teorijski i praktični dio ovog rada, te na temelju toga i u postojećem društvenom kontekstu zaključuje o važnosti i značenju reorganizacije enciklopedijski organiziranog znanja u mrežnom okruženju koji se temelji na principima semantičkog weba.

2 TEORIJSKE OSNOVE SEMANTIČKOG WEBA

U cilju potpunog razumijevanja uloge ontologija u informacijskim sustavima potrebno je prvo razumjeti zamisao semantičkog weba na koji se one oslanjaju. Cilj ovog poglavlja je ukazati na osnovnu teorijsku i tehničku pozadinu semantičkog weba.

2.1 Semantički web

World Wide Web je u osnovi zapisan u HTML-u (engl. Hyper Text Mark-up Language), koji je osmišljen kako bi bio razumljiv i interpretiran ljudima. Povećanjem količine raspoloživih podataka na trenutnom webu, proces pronalaženja, organiziranja, pristupa i upravljanja informacijama postaje izuzetno težak za korisnike. Stoga se dolazi do zamisli prebacivanja pretraživanja podataka s korisnika na računala, što dovodi do koncepta semantičkog weba kako bi se poboljšale neke slabosti sadašnjeg weba. Osim vrijednih poveznica (linkova, engl. link) uspostavljenih između dokumenata, potrebni su i obvezni alati kao što su pretraživači. Pretraživači bazirani na ključnoj riječi kao što su Yahoo i Google su glavni alati za današnju upotrebu weba. Međutim, njihovom upotrebom pojavljuju se sljedeći problemi:

- Visoki odziv s niskom točnošću.
Ako su glavne stranice i dohvaćene, njihova upotreba je mala ukoliko je dohvaćeno i ostalih 28 758 relevantnih ili nerelavantnih dokumenata. Previše ponuđenih rezultata je jednako loše kao i premalo.
- Niski ili nikakav odziv.
Često se događa da se ne mogu dobiti relevantni odgovori tj. relevantne stranice na postavljene upite. Međutim, ovaj problem se rjeđe pojavljuje s trenutnim pretraživačima.
- Rezultati su visoko osjetljivi na rječnik.
Često početne ključne riječi ne daju željeni rezultat jer relevantni dokumenti upotrebljavaju terminologiju različitu od izvornog upita. Ovo je nezadovoljavajuće jer bi semantički slični upiti trebali dati slične rezultate.
- Rezultati su pojedinačne web stranice. Ukoliko nam je potrebna informacija koja je raširena u različitim dokumentima, potrebno je pokrenuti nekoliko upita za

prikupljanje relevantnih dokumenata te ručno izvući iz njih djelomične informacije i okupiti ih zajedno.

Unatoč poboljšanjima u pretraživačkoj tehnologiji, poteškoće ostaju iste. Čini se da količina web sadržaja nadilazi tehnologijski napredak. Ukoliko je pretraživanje i uspješno, osoba treba sama pregledati odabrane dokumente kako bi izvukla točnu i potrebnu informaciju, što zahtijeva priličnu potrošnju vremena.. Stoga, ne postoji velika podrška za *pretraživanje informacija* (engl. *information retrieval, IR*). Tako je termin *pretraživanje informacija* često upotrebljavan i povezan s pretraživačima, nešto zbog čega nastaju nesporazumi te je bolji nalaznik lokacije (engl. *location finder*). Rezultati web pretraživanja nisu lako dostupni drugim softverskim alatima jer su pretraživači često izolirane aplikacije. Glavna trenutna prepreka za osiguranje bolje potpore korisnicima weba je u tome što značenje tj. smisao web sadržaja nije računalno razumljiv. Postoje alati za pretraživanje i dijeljenje teksta u dijelove, provjeravanje pravopisa, brojanje riječi. Međutim, kada je potrebno interpretirati rečenice i ekstrahirati upotrebljivu informacija za korisnika, mogućnosti sadašnjih softwera su još uvijek ograničene. Alternativni pristup je predstavljanje web sadržaja u obliku koji je mnogo lakše strojno obradiv (engl. *machine-processable*)² i upotreba inteligentnih alata kako bi se unaprijedilo ovo predstavljanje. Ovaj plan preinačavanja weba upućuje na inicijativu semantičkog weba.

Semantički web je podržan od strane *World Wide Web Consortium (W3C)*³, međunarodnog normativnog tijela za web, pod vodstvom Tim Berners-Lee koji je izumio WWW krajem 80-ih godina 20. st. Očekuje se ostvarivanje njegove izvorne slike weba u kojoj značenje informacije ima mnogo važniju ulogu nego što je ima u današnjem webu. Pojam semantičkog Web je uveo izumitelj World Wide Weba, Tim Berners-Lee, čija je temeljna zamisao proširenje mogućnosti trenutnog weba dokumenata i stvaranje weba podataka kojima se može pristupiti direktno ili indirektno te su isti strojno obradivi.^{4,5,6} Dakle,

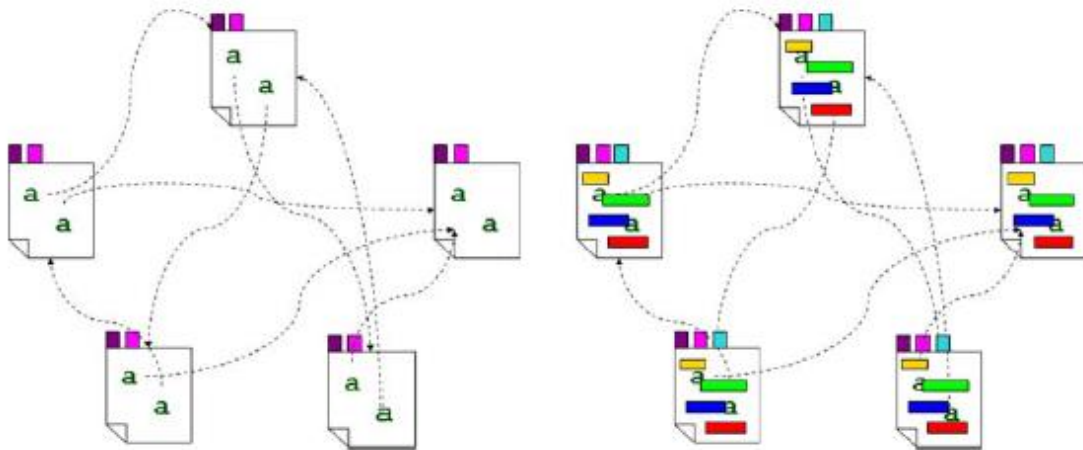
² U literaturi se češće upotrebljava termin strojno razumljiv (eng. *machine-understandable*). To je pogrešna riječ jer daje krivu utisak. Nije nužno za inteligentne posrednike (inteligentne agenta, engl. *intelligent agents*) da razumiju informaciju; za njih je dovoljno da učinkovito obrade informaciju, što katkad kod ljudi uzrokuje shvaćanje da stroj uistinu razumije. // In Antoniou, Grigoris; Van Harmelen, Frank. *A Semantic Web Primer*. 2. ed. Massachusetts London, England: The MIT Press Cambridge, 2008. Str. 3.

³ World Wide Web Consortium (W3C). Dostupno na URL: <http://www.w3.org/> (8.9.2015.)

⁴ Semantic Web. // In World Wide Web Consortium. Dostupno na URL: <http://www.w3.org/standards/semanticweb/> (8.9.2015.)

⁵ Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. // *Scientific American Magazine* 284, 5 (2001), 28–37.

semantički web nije globalni informacijski put paralelan s postojećim World Wide Webom. Upravo suprotno, semantički web je postupni razvoj postojećeg weba. Problem trenutnog WWW je informacijska prezasićenost. Semantički web nudi postojanje metapodataka za strukturiranje. Slika 1 ukazuje na osnovnu razliku između trenutnog WWW i semantičkog weba. Trenutni web se promatra kao skup dokumenata u kojemu postoji nešto metapodataka pridruženih dokumentima (naslov, ključne riječi, itd.) te hiperlinkova među dokumentima (lijeva strana slike). Semantički web dodaje slici semantičke metapodatke na razini dokumenta (pridružene cijelom dokumentu) i na razini sadržaja pridružujući ih određenom dijelu i mjestu u dokumentu. Stoga, desna strana Slike 1 prikazuje semantički web kao mnogo bogatiji, „šareniji“ od sadašnjeg.



Slika 1: Trenutni WWW (lijevo) i semantički web (desno)

Realizacija semantičkog weba ovisi o dostupnosti kritične količine metapodatka za web sadržaj, povezanih s odgovarajućim formalnim znanjem o svijetu. Semantički web dodaje formalne strukture i semantiku (metapodatke i znanje) na web sadržaj u svrhu boljeg upravljanja i pristupa. Krajnji cilj semantičkog weba je osigurati zajednički okvir koji omogućava računalima i ljudima rad u međusobnoj suradnji, dijeljenje i ponovnu upotrebu

⁶ Davies, J.; Lytras, M.; Sheth, A. P. Semantic Web Based Knowledge Management. // IEEE Internet Computing 11, 5(2007), 14-16.

podataka preko aplikacija i zajednica.^{7,8,9} Cilj semantičkog weba je omogućiti napredniji sustav upravljanja znanjem:

- Znanje će se organizirati u pojmovne prostore prema njegovom značenju
- Automatizirani alati će podržati upravljanje provjerom nedosljednog i ekstrakcijom novog znanja
- Pretraga bazirana na ključnim riječima će se zamijeniti s odzivima na upite: zahtijevano znanje će se pretražiti, izlučiti, i predstaviti na čovjeku pristupačan način
- Podržan odziv na upit preko nekoliko dokumenata
- Mogućnost definiranja korisnika koji mogu vidjeti određene dijelove informacija (čak i dijelova dokumenata)¹⁰

2.2 Pojmovi i arhitektura semantičkog weba

Tim Berners-Lee je zamisao semantičkog weba predstavio modelom „*Slojeviti kolač*“ (engl. „*Layer Cake*“), koji je podijeljen na slojeve ukazujući na arhitekturu semantičkog weba (Slika 2). Model ima zadatak standardizirati tehnologiju semantičkog web-a. Na najosnovnijoj razini semantički web ovisi od Unicode i Uniform Resource Identifier (URI), što je jednostavno rečeno web identifikator. Unicode je standard za kodiranje niza znakova, čijom upotrebom svi ljudski jezici mogu biti napisani i čitani na webu. URI je niz znakova koji osiguravaju standardni oblik za identifikaciju izvora.^{11,12,13} Dakle, taj sloj ima zadaću jedinstveno definirati pojmove u globalnoj mreži podataka.

Drugi sloj arhitekture semantičkog weba je XML (engl. Extensible Markup Language) jezik za označavanje, slično kao i HTML. No, za razliku od HTML-a, ne postoji skup predefiniраниh tagova (engl. tag) u XML-u. Umjesto toga, XML je proširiv te tagovi mogu biti definirani i iskoristivi kako to zahtijeva određena aplikacija. Osim navedenih razlika, XML je

⁷ Berners-Lee, T.; Hendler, J.; Lassila, O., loc.cit.

⁸ Davies, J.; Lytras, M.; Sheth, A., loc.cit.

⁹ W3C Semantic Web Activity. Dostupno na URL: <http://www.w3.org/2001/sw/> (9.9.2015)

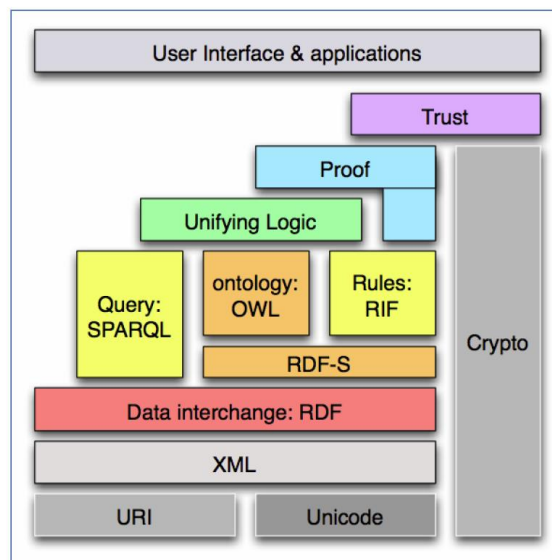
¹⁰ Antoniou, Grigoris; Van Harmelen, Frank., op. cit. (bilj. 2.), str. 4.

¹¹ What is Unicode? Dostupno na URL: <http://www.unicode.org/standard/WhatIsUnicode.html> (10.09.2015.)

¹² URIs, URLs, and URNs: Clarifications and Recommendations: Report from the joint W3C/IETF URI Planning Interest Group // World Wide Web Consortium, 21 September 2001. Dostupno na URL: <http://www.w3.org/TR/uri-clarification/> (10.9.2015.)

¹³ Berners-Lee, T., Fielding, R. RFC3986 - Uniform Resource Identifier (URI): Generic Syntax. // W3C/MIT, 2005. Dostupno na URL: <https://tools.ietf.org/html/rfc3986> (10.9.2015.)

stvoren za prijenos podataka u odnosu na HTML koji se koristi za prikazivanje podataka. Još jedna prednost XML-a je da je samo-opisni jezik (engl. self-descriptive language) koji se može koristiti između različitih platformi i programskih jezika te će i dalje izražavati kompleksne poruke i funkcije.^{14,15,16} XML Schema ograničava strukturu i sadržajne elemente XML dokumenata. World Wide Web Consortium (W3C)¹⁷ preporučuje XML imenski prostor (engl. namespace) koji se primjenjuju za jedinstveno imenovanje elemenata i atributa u XML dokumentu. U semantičkoj web arhitekturi, XML sloj s XML namespace i XML shemom objavljuje da se koristi zajednička sintaksa u semantičkom webu¹⁸. Dakle, on predstavlja temeljnu strukturu zapisa podataka.



Slika 2: Arhitektura semantičkog weba (preuzeto iz: (Antoniou, G.; Van Harmelen, F., 2008.))

Druga važna tehnologija za razvoj semantičkog weba je RDF (engl. Resource Description Framework), koji je u početku bio zamišljen kao meta modela podataka. RDF je okvir za opisivanje informacija i resursa na webu. Stavljajući informacije u RDF datoteke

¹⁴ Berners-Lee, T., Hendler, J., Lassila, O., loc cit.

¹⁵ Nandigam, Jagadeesh; Gudivada, Venkat N; Kalavala, Mrunalini. Semantic Web services. // J. Comput. Small Coll. 21, 1(2005), 50-63.

¹⁶ Web services Architecture / ed. by David Booth ... <et al.>. World Wide Web Consortium, 11 February 2004. Dostupno na URL: <http://www.w3.org/TR/ws-arch/#XML-infoset> (10.9.2015.).

¹⁷ World Wide Web Consortium (W3C), loc. cit.

¹⁸ Obtiko, Marek. Semantic Web Architecture. // Ontologies and Semantic Web, 2007. Dostupno na URL: <http://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html> (10.09.2015.).

računalnim, programima za pretraživanje se čini izvedivim otkrivanje, prihvaćanje, prikupljanje, analiziranje i obrada informacija s interneta. Budući su RDF dokumenti pisani u XML-u, oni mogu lako biti razmijenjeni među različitim vrstama operativnih sustava i aplikacijskih jezika (engl. application language). RDF se temelji na izjavama koje su poznate kao trojke (triple, engl. triple). Svaka trojka sadrži subjekt, predikat i objekt; poput rečenice u prirodnim jezicima. Stoga se svaka trojka može oblikovati kao graf s dva čvorišta (subjekt i objekt), koji su povezani s predikatom (slikovno predstavljenim sa strelicom između čvorova).^{19,20,21,22} Subjekt predstavlja sami resurs i definiran je URI-jem. Predikat, koji predstavlja odnos, također je URI jer svaki odnos ima standardnu definiciju koja je izražena kroz jedinstveni URI. Objekt se može definirati kao URI ili doslovna vrijednost (literalna vrijednost). RDF Schema (RDFS) pruža RDF rječnik kako bi se omogućio opis taksonomije klasa i svojstava, dajući glavne elemente opisa ontologija.^{23,24} Dakle, *RDF* je model koji samo povezuje podatke, a *RDF Schema* daje osnovni smisao i značenje vezama te omogućuje hijerarhiju.

Sljedeći temeljni sloj arhitekture semantičkog weba je web ontologijski jezik OWL (engl. Web Ontology Language), koji se može usporediti s gramatičkim sustavom semantičkog weba. Korijeni OWL-a su u deskriptivnoj logici i njegov temeljni cilj je donijeti snagu zaključivanja deskriptivne logike semantičkom webu. OWL je izgrađen na vrhu RDF-a te pruža dodatni standardizirani rječnik, vrhunsku strojnu interoperabilnost i dobro definiranu sintaksu koja je puno jača od RDF-a.^{25,26} Spomenute specifikacije čine OWL prikladnim jezikom za razvoj ontologije, odnosno prikazivanje pojmova i njihovih odnosa. Na temelju različitih raspona kompatibilnost i ograničenja OWL dolazi u tri različite kategorije: OWL Lite, OWL DL (engl. Web Ontology Language Description Language) i OWL Full. OWL Lite je najjednostavniji i najograničeniji OWL podjezik, koji se može upotrijebiti za razvrstavanje resursa u hijerarhijski oblik. Vrlo je koristan kada se zahtijeva nekomplikiran OWL sustav s

¹⁹ W3C Semantic Web Activity, loc. cit.

²⁰ Obtiko, Marek, loc. cit.

²¹ Resource Description Framework Model and Syntax Specification. / ed. by Ora Lassila, Ralph R. Swick. World Wide Web Consortium, 1999. Dostupno na URL: <http://www.w3.org/TR/PR-rdf-syntax/> (10.9.2015.).

²² RDF 1.1 Primer. / ed by Frank Manola, Eric Miller, Brian McBride. World Wide Web Consortium, February 2014. Dostupno na URL: <http://www.w3.org/TR/rdf-primer/> (10.9.2015.).

²³ Ibid.

²⁴ Shadbolt, Nigel; Hall, Wendy; Berners-Lee, Tim. The Semantic Web Revisited. // J. IEEE Intelligent Systems 21, 3(2006), 96–101.

²⁵ Obtiko, Marek, loc. cit.

²⁶ Antoniou, Grigoris; Van Harmelen, Frank. Web Ontology Language: OWL. Dostupno na URL: <http://www.cs.vu.nl/~frankh/postscript/OntoHandbook03OWL.pdf> (16.09.2015.).

ograničenim resursima ili brzinom. OWL DL je sveobuhvatniji sustav koji dolazi s većim potencijalom i s više fleksibilnosti, dok je još uvijek moguća implementacija i dosljedno korištenje. Stvoren je za puno logičko opisivanje, nudeći sustav zaključivanja. Omogućuje složenije analize, klasifikacije, odnose i svojstva od OWL Lite. OWL Full je teoretski neograničen model, gdje odnosi i klasifikacije mogu biti široko prošireni s mnogo složenosti. Uglavnom je usmjeren na korisnike koji traže maksimum izražajnosti i sintaktičke slobode RDF-a. OWL Full može u potpunosti podržati OWL DL i OWL Lite sustave, iako je to nemoguće pravilno predvidjeti i implementirati. Dakle, ovaj sloj nadograđuje prethodni većim brojem semantičkih veza između podataka. Ontologijski razvojni stručnjaci moraju odabrati OWL vrstu koja najbolje odgovara njihovim potrebama.^{27,28,29,30}

Uloga logičkog sloja (engl. Logic layer) je podržati snažni logični jezik za donošenje zaključaka, koji čine semantički web dovoljno izražajnim. Dakle, sloj logike je zadužen za razumljivo izvlačenje informacija i donošenje odluka od strane aplikacija. Sva semantika i pravila se izvode na slojevima ispod sloja dokaza (engl. Proof) i rezultati će se koristiti za potvrdu zaključivanja. Gornji slojevi arhitekture semantičkog weba izražavaju zajedno i dokaze i pouzdanost podataka kako bi prikazali pouzdanost rezultata. Kriptografija je sredstvo digitalnog potpisa koji se može koristiti za vjerodostojnost. Korisničko sučelje aplikacije se može izgraditi na vrhu svih ovih slojeva.^{31,32} Dakle, *sloj dokaza* traži dokaz kako je nađen podatak točan, a *sloj povjerenja* je zadužen za provjeru mogućnosti vjerovanja resursu da su podaci točni. Ovo je najviši stupanj te će web doseći njegov potpuni potencijal kada korisnici budu imali povjerenje u njegovo djelovanje i u dostavljene informacije³³.

2.3 Ontologija

Pojam ontologije dolazi iz temeljne filozofske discipline koja proučava biće u njegovim temeljnim odredbama, s obzirom na njegovu bit i bitak; nauk o temeljnim uzrocima

²⁷ Obtiko, Marek, loc cit.

²⁸ Antoniou, Grigoris; Van Harmelen, Frank., loc cit.

²⁹ Smith, Michael K.; Welty, Chris; McGuinness, Deborah L. OWL Web Ontology Language Guide. Dostupno na URL: http://www.w3.org/TR/owl-guide/#owl_Class (16.09.2015.).

³⁰ McGuinness, Deborah L.; Harmelen, Frank Van. OWL Web Ontology Language Overview. Dostupno na URL: <http://www.w3.org/TR/owl-features/> (16.09.2015.).

³¹ Obtiko, Marek, loc cit.

³² Palmer, Sean B. The Semantic Web: An Introduction. Dostupno na URL: <http://infomesh.net/2001/swintro/> (16.09.2015.).

³³ Antoniou, Grigoris; Van Harmelen, Frank., op. cit. (bilj. 2.), str. 17-18.

svoga postojećega.³⁴ Ontologije su u području računalnih znanosti definirane kao što slijedi u nastavku teksta:³⁵

- (1) Prema Gruberu razvoj ontologija je jasna formalna specifikacija termina u domeni i odnosa među njima.³⁶ Promatranje kako je svijet izgrađen od određenih objekata koji se mogu grupirati u apstraktne klase temeljene na dijeljenju svojstava je tipično ontologijsko razmišljanje.
- (2) Termin ontologija šire znači jedan ili dva povezana pojma: (a) predstavlja rječnik, često specijaliziran za određenu domenu ili predmet interesa, (b) tijelo znanja koje opisuje neku određenu domenu upotrebom predstavljenog rječnika³⁷.
- (3) Uobičajeno, ontologija sadrži listu termina i odnosa između termina. Termini označavaju važne pojmove (klase objekata) domene. Odnosi uobičajeno uključuju hijerarhiju klase. Osim podklasnih odnosa, ontologije mogu uključivati informacije kao što su:
 - Svojstva (x uči y),
 - Vrijednosna ograničenja (jedino članovi sveučilišta mogu podučavati kolegij),
 - Disjunktne iskaze (sveučilišno i opće osoblje su disjunktne),
 - Određivanje logičnih odnosa između objekata (svaki odsjek mora uključivati barem deset članova fakulteta)³⁸
- (4) Ontologija definira osnovne termine i odnose uspoređujući rječnik tematskog područja kao i pravila za kombiniranje termina i odnosa kako bi se definirala proširenja na rječniku.³⁹

³⁴ Ontologija // Hrvatska enciklopedija, mrežno izdanje / glavni urednik Slaven Ravlić. Zagreb: Leksikografski zavod Miroslav Krleža, 2013-2015. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=45185> (18.11.2015.).

³⁵ Gruber, Thomas; Liu, Ling; Özsu, M. Tamer. Ontology // In Encyclopedia of Database Systems / ed. by Ling Liu and M. Tamer Özsu. New York: Springer-Verlag, 2009. Dostupno na URL: <http://www.dainf.ct.utfpr.edu.br/~tacla/Ontologias/2013/a01-020-Ontology-definitionTomGruber.pdf> (19.09.2015.).

³⁶ Noy, Natalya F.; McGuinness, Deborah L. Ontology Development 101: A Guide to Creating Your First Ontology // Citeseer; 2001. Str. 1. Dostupno na URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.5085> (17.09.2015.)

³⁷ usp. Gašević, Dragan; Djurić, Dragan; Devedžić, Vladan. Model Driven Architecture and Ontology Development. Berlin: Springer-Verlag, 2006. Str. 46.

³⁸ usp. Antoniou, Grigoris; Van Harmelen, Frank., op. cit. (bilj. 2.), str. 10-12.

³⁹ Ibid.

(5) Ontologija je formalna eksplicitna specifikacija zajedničke konceptualizacije, koja se odnosi na apstraktni model nekog fenomena u svijetu identificirajući relevantne koncepte tog fenomena. Eksplicitno znači da su upotrebljene vrste pojmova te da su upotrebljena ograničenja na njima eksplicitno definirana. Formalno se odnosi na činjenicu da ontologija treba biti strojno čitljiva. Zajedničko odražava stajalište da ontologija dohvaća konsenzualno znanje koje nije privatno od nekog pojedinca, nego je prihvaćeno od strane zajednice⁴⁰.

(6) Skup logičkih aksioma stvorenih za sadržavanje namjeravanog značenja rječnika.⁴¹

(7) Ontologija je hijerarhijski strukturirani skup termina za opisivanje domene koja se može upotrebiti kao temelj za bazu znanja.⁴²

Sve navedene definicije izgledaju drukčije u nekom konceptu ili perspektivi definiranja svrhe ontologija, ali sve definicije govore o konceptualizaciji i nedvosmislenosti (odnosno eksplicitnosti). U kontekstu dijeljenja znanja ontologija je eksplicitna specifikacija zajedničke konceptualizacije o domeni interesa. Sa stajališta računalne znanosti i umjetne inteligencije, ontologija predstavlja domenu znanja ili diskurs kao skup pojmova (klasa), njihovih atributa (svojstva), instanci ovih pojmova (individua) i odnosa kojima se klase i individue mogu međusobno povezati.^{43,44} Ontologija je formalan prikaz pojmova s dobro definiranim odnosima među tim pojmovima. Ovi pojmovi su izvor zajedničkih termina koje razumiju i obrađuju strojevi. Tipična ontologija se sastoji od hijerarhijske strukture pojmova i njihovih odnosa u određenoj domeni. Implementacija ontologije je središte prikaza znanja temeljenog na semantičkom webu. OWL je jezik koji je preporučio W3C Semantic Web standard za kodiranje ontologija.⁴⁵ Ontologija modelira znanje na semantičkoj razini, dok shema baze podataka modelira znanje na logičkoj razini. Stoga je ontologija značajna za prikaz određene domene, što omogućuje automatsko zaključivanje i interpretaciju s primjenjivim semantičkim kontekstom. Na temelju svoje neovisnosti od niže razine modela

⁴⁰ Ibid.

⁴¹ Ibid.

⁴² Ibid.

⁴³ W3C Semantic Web Activity, loc cit.

⁴⁴ Gruber, Thomas; Liu, Ling; Özsu, M. Tamer, loc cit.

⁴⁵ McGuinness, Deborah L.; Harmelen, Frank Van, loc cit.

podataka, ontologija je sposobna integrirati i razmjenjivati podatke između heterogenih izvora informacija te definirati sučelja za nezavisne i na znanju utemeljene usluge, kao što je odgovaranje na upite i ponovno korištenje izvora znanja.^{46,47,48} Drugim riječima, ontologija se može koristiti kao način komunikacije između čovjeka i sustava ili sustava sa sustavom, čime je omogućeno semantičko pretraživanje informacija (IR) i upravljanje znanjem.⁴⁹

U kontekstu weba ontologije omogućuju dijeljenje zajedničkog razumijevanja strukture informacija među ljudima ili softverskim agentima (engl. software agent)⁵⁰ određene domene, što je potrebno u prevladavanju razlika u terminologiji. To se postiže pridruživanjem određene terminologije u vidu ontologije za dijeljenje ili definiranje direktnog pridruživanja između ontologija. Ontologije se upotrebljavaju za organiziranje i navigaciju web stranica i povećanje točnosti pretraživanja weba.⁵¹ Primjerice, neka nekoliko različitih mrežnih stranica sadrže medicinske informacije ili osiguravaju medicinske e-komercijalne usluge. Ukoliko ove web stranice dijele i objavljuju istu osnovnu ontologiju termina koju svi upotrebljavaju, računalni agenti mogu ekstrahirati i prikupiti informacije iz ovih različitih stranica. Agenti mogu upotrijebiti sakupljene informacije za odgovaranje na korisnikove upite ili kao ulazne podatke za drugu primjenu. Glavna svrha ontologije nije njezina upotreba u vidu rječnika i taksonomije već *dijeljenje i ponovna upotreba znanja*. Dakle, svrha stvaranja ontologije je:⁵²

⁴⁶ Berners-Lee, T.; Hendler, J.; Lassila, O., loc cit.

⁴⁷ Gruber, Thomas; Liu, Ling; Özsu, M. Tamer, loc cit.

⁴⁸ Gruber, Thomas R. A Translation Approach to Portable Ontology Specifications. // Knowledge Acquisition 5, 2 (1993), 199–220.

⁴⁹ Chena, Rung Ching; Baub, Cho Tsan; Huang, Yun H. Development of Anti-diabetic Drugs Ontology for Guideline-based Clinical Drugs Recommend System Using OWL and SWRL. // In 2010 IEEE World Congress on Computational Intelligence, WCCI 2010 (2010), 5-10.

⁵⁰ Agenti su dijelovi software-a koji rade samostalno i proaktivno. Pojmovno su se razvili iz pojmova objektno rijentiranog-programiranja (engl. object-oriented programming) i sastavnom dijelu-baziranom razvoju software-a (engl. component-based software development). Osobni agent na semantičkom webu će dobivati neke zadatke i sklonosti (prioritete) od osobe, pronalaženje informacija iz web izvora, komuniciranje s ostalim agentima, usporediti informacije o korisničkim zahtjevima i sklonostima (prioritetima), odabiranje točnog izbora i davanje odgovora korisnicima. Treba istaknuti kako agenti neće zamijeniti čovjeka-korisnika na semantičkom webu, niti će oni nužno donositi odluke. U mnogim, ako ne u većini, slučajeva njihova uloga će biti prikupljanje i organiziranje informacija i predstavljanje izbora korisniku na odabir. Agenti semantičkog weba će upotrijebiti svu tehnologiju, kao što je navedeno: (1) metapodaci će biti upotrebljeni za identifikaciju i ekstrakciju informacija iz web izvora, (2) ontologije će se upotrijebiti kako bi pomogle u pretraživanju weba, interpretaciji pretraženih informacija i komunikaciji s ostalim agentima, (3) logika će se upotrijebiti za obradu pretraženih informacija i za izvlačenje zaključaka. Bit će potrebne daljnje tehnologije, kao što su komunikacijski jezici agenata. // Antoniou, Grigoris; Van Harmelen, Frank, op. cit. (bilj. 2.), str. 14-15.:

⁵¹ usp. Antoniou, Grigoris; Van Harmelen, Frank, (bilj. 2.), str. 12.

⁵² Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

- *Dijeljenje zajedničkog razumijevanja sastava informacija među ljudima i softverskim agentima*: informacije se ponavljaju, ista informacija može biti dostupna na mnogo mjesta te ovo dupliciranje može biti varljivo. Uloga ontologije je ukoliko web stranice dijele zajedničku ontologiju, softverski agenati mogu ekstrahirati i spojiti informacije čineći pretraživanje učinkovitijim i ekonomičnijim poštujući resurse.
- *Ponovna upotreba znanja određene domene*: domensko znanje je predmet promjena, no bez obzira koliko se brzo domensko znanje može mijenjati, osnovni koncepti ostaju isti i to je veoma dragocjeno. Kada se ove informacije predstavljaju kao ontologija, bit će u mogućnosti ponovnog korištenja, dijeljenja i sigurnosti.
- *Analiziranje znanja domene*: Analiza i navigacija kroz termine određene domene je jednostavno izrađena kada je predstavljena s ontologijom. Ontologija definira način na koji su ti termini povezani.

2.4 Jezici semantičkog weba

RDF, RDFS i OWL su osnovni jezici za predstavljanje semantičkog web-a, s tim da RDF služi kao temelj.

2.4.1 RDF

RDF (engl. Resource Description Framework, hrv. okvir za opis izvora) je formalni jezik za predstavljanje znanja koji je W3 konzorcij (W3C) prvi put predstavio 1999. godine, a preporučio ga je kao standard za pisanje metapodataka 2004. godine. Predstavlja okvir koji se može upotrijebiti za opis različitih vrsta izvora, dokumenata ili pojmova.⁵³ Kako bi se semantički web implementirao potreban je model podataka koji dozvoljava informacijama da budu distribuirane (dijeljene) na webu, a to je RDF koji rješava temeljno pitanje upravljanja distribuiranim podacima na semantičkom webu te su svi ostali standardi semantičkog web-a izgrađeni na tom temelju distribuiranih podataka.⁵⁴ Stoga je RDF osnovni građevni blok

⁵³ RDF 1.1 primer, loc. cit.

⁵⁴ Allemang, Dean; Hendler, Jim. *Semantic Web for the working Ontologist: Effective modeling in RDFS and OWL. 2nd ed.* Waltham, MA : Morgan Kaufmann/Elsevier, 2011. Str. 27.

semantičkog weba, te za semantički web predstavlja isto što je i HTML za obični web. RDF omogućava međusobnu razmjenu informacija među različitim aplikacijama, što je i osnova semantičkog weba. RDF koristi apstraktni model kako bi raščlanio znanje/informacije u manje djeliće. Taj model je dovoljno jednostavan kako bi se omogućilo da se pomoću njega izrazi bilo koja činjenica, ali i dovoljno složen da se omogući računalnim aplikacijama da iskoriste strukturirano znanje unutar dokumenata⁵⁵ Osnovni koncepti RDF-a su resursi, svojstva i izjave.

Resurs (engl. resources) predstavlja bilo koji pojam iz stvarnog svijeta koji se može identificirati i imenovati. U semantičkom webu se na stvari u svijetu referira kao na resurse. Stoga, RDF gleda na svijet kao sastav dviju vrsta resursa: individua i odnosa (relacija). *Resurs* može biti bilo što o čemu netko može željeti razgovarati, kao npr. Shakespeare, Split, „vrijednost X“, „sve ribe u Jadranskom moru“. Resurs je sve što može imati URI. Dakle, RDF koristi URI-je da identificira resurse. Katkada se kao alternative riječi za resurs mogu koristiti *entitet* ili *stvar*. U svakom slučaju, *resurs* je riječ koja se koristi u standardima semantičkog weba te će se stoga upotrebljavati i u ovom radu.⁵⁶ Osim pojmova koji se žele povezati u semantičkom webu, pod resursima se podrazumijevaju i sami odnosi, odnosno veze između tih pojmova.^{57,58} Dakle, sve što u RDF trojki objekt ili subjekt opisuju, ili na što se pojam odnosi, naziva se resurs. Resurs je sve ono što se opisuje pomoću RDF izjava. To mogu biti knjige, izdavači, mjesta, ljudi, hoteli, razni upiti, proizvodi i slično. Svaki resurs se jednoznačno određuje dodjeljivanjem URI-a (engl. Uniform Resource Identifier), kako bi se mogao identificirati. Na taj način se razlikuju resursi sa istim imenom, ali različitim značenjem poput homonima. Osnovna funkcija URI-a je točno definiranje resursa o kojem se radi, za razliku od URL-a koji je bez sadržaja u pozadini beskorisna i nefunkcionalna poveznica (engl. link). Istom resursu može biti dano različito ime. Također, moguće je da dva ili više različitih dokumenata koriste ista imena za različite resurse. Kako bi se ovaj problem riješio, svakom resursu se daje posebno ime, koje je identificirano URI-jem. To ime treba biti jedinstveno na globalnoj razini kako se ne bi dogodilo da dva ili više resursa imaju isto ime ili da se različita imena koriste za isti resurs. Dakle, resurs je sve što se može definirati pomoću URI-ja, npr. <http://www.enciklopedija.hr>. RDF opisuje te resurse pomoću njihovih svojstava

⁵⁵ Liyang, Yu. *A Developer's Guide to the Semantic Web*. Heidelberg ; New York : Springer, 2011. Str. 53.

⁵⁶ Allemang, Dean; Hendler, Jim, (bilj. 54.) op. cit., str. 27-28.

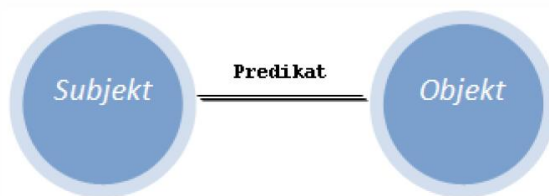
⁵⁷ Antoniou, Grigoris; Van Harmelen, Frank., op. cit. (bilj. 2.), str. 63-64.

⁵⁸ Allemang, Dean; Hendler, Jim, (bilj. 54.) loc. cit.

(odnosa) i vrijednosti svojstava. *Odnos* predstavlja resurs koji ima ime poput „izdaje“ i sl., a vrijednost navedenog svojstva bi bila npr. „Leksikografski zavod Miroslav Krleža“.

Kombinacija resursa, svojstva i vrijednosti toga svojstva se naziva *izjava* (*iskaz*, engl. *statement*). Svaka informacija ili znanje je predstavljeno skupom izjava, a svaka od tih izjava se sastoji od *subjekta-predikata-objekta* i taj redoslijed se ne smije mijenjati. Subjekt je ono u čemu se u izjavi govori i mora biti predstavljen URI-jem. Odnosi se na određeni resurs u svijetu. Subjekt izjave se također naziva početni čvor izjave. Predikat u RDF-u označava vezu između subjekta i objekta. RDF model daje prednost korištenju URI-ja za identificiranje predikata, nasuprot stringova ili riječi jer se predikat može tretirati kao resurs u drugoj izjavi, što znači da se može proširiti dodatnim informacijama o predikatu. Objekt ili vrijednost svojstva, također mora biti identificiran pomoću URI-ja, no razlika je u tome što objekt može biti i doslovna (literalna) vrijednost ili bilo koji drugi podaci. Doslovne vrijednosti (literali) također mogu imati URI koji objašnjava tip podataka koji se nalazi u tim podacima, kako bi se stroju olakšalo tumačenje toga podatka. Još jedno svojstvo subjekta i objekta jest to što mogu poprimiti praznu vrijednost, tj. vrijednost koja ne mora biti identificirana URI-jem. Ova vrijednost se naziva anonimni čvor (eng. *anonymous node*) ili *bnode*. Budući da se svaka izjava sastoji od ova tri dijela, izjava se također naziva i *trojka* (*triplet*, engl. *triples*) te predstavlja temeljni građevni blok za RDF. Svaka izjava ili trojka predstavlja određenu činjenicu, skup ovih izjava ili trojki predstavlja neku informaciju ili znanje, i može se predstaviti RDF grafom. Subjekt i objekt se mogu odnositi na konkretne stvari, poput imena i prezimena autora, ali i apstraktne stvari poput boje očiju. Dakle, resurs se naziva sve što objekt ili subjekt opisuju, ili na što se pojam odnosi, tj. resurs je sve ono što se opisuje pomoću RDF izjava. To mogu biti knjige, izdavači, mjesta, ljudi, razni upiti, proizvodi i sl. Na semantičkom webu, odnosno webu informacija, svatko može doprinosti znanju o resursu. Subjekt i objekt predstavljaju imena stvari, predikat predstavlja vezu među tim stvarima, npr. *Miroslav Krleža* (subjekt) je *autor* (predikat) djela koje se zove *U agoniji* (objekt). Dakle, potrebno je povezati (spojiti, dovesti u vezu) individue i odnose koji ih povezuju, a u svrhu uspostavljanja ove veze RDF omogućuje *iskaze*, odnosno trojke. Dakle, trojka je minimalna jedinica diskursa u RDF-u i sastoji se od: (a) subjekta (engl. *subject*), koji identificira resurs o kojemu je izjava, (b) svojstva (engl. *predicate*), odnosno URI-a koji identificira svojstvo subjekta i (c) objekta, koji identificira resurs povezan svojstvom sa subjektom. Unarne relacije su u RDF-u predstavljene klasama (engl. *classes*), a binarne relacije su predstavljene svojstvima (engl. *properties*). Same klase i odnosi su resursi. Ovim redoslijedom je potrebno načiniti dvije vrste veza: (1) povezati resurse na unarnu relaciju kojoj pripadaju (resurs

Miroslav Krleža pripada unarnoj relaciji *Književnik*) i (2) povezati parove resursa na binarnu relaciju kojoj oni pripadaju (par *Miroslav Krleža*, *U agoniji* pripada relaciji *autor*). Slika 3 prikazuje RDF trojku.

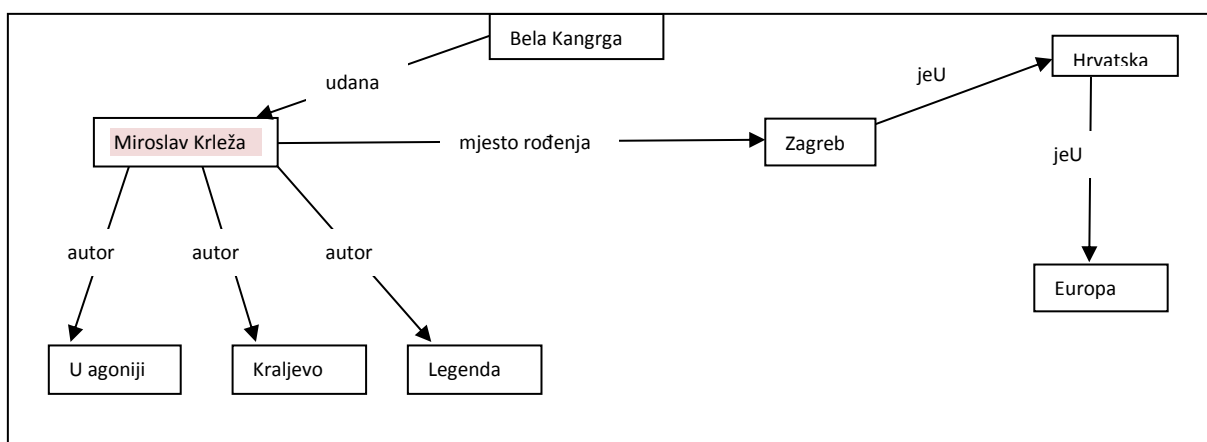


Slika 3: RDF trojka

Trojke su interesantne kada se više od jedne trojke referira na isti entitet, kao što je u Tablici 1. Trojke iz Tablice 1 su prikazane kao direktan graf u kojemu je svaka trojka prikazana kao veza (strelica) od subjekta do objekta sa svojstvom imenovanim na strelici, kao što je prikazano na Slici 4. Graf vizualizira i izražava istu informaciju predstavljenu u Tablici 1, ali sve što se zna o Miroslavu Krleži (bilo subjekt ili objekt) je prikazano na jednom čvoru.

Tabela 1: Jednostavne trojke

Subjekt	Predikat	Objekt
Miroslav Krleža	autor	U agoniji
Miroslav Krleža	autor	Kraljevo
Miroslav Krleža	autor	Legenda
Bela Kangrga	udana	Miroslav Krleža
Miroslav Krleža	mjesto rođenja	Zagreb
Zagreb	jeU	Hrvatska
Hrvatska	jeU	Europa



Slika 4: Grafički prikaz trojki iz Tablice 1

Ovaj model je distribuiran te višestruki sudionici mogu činiti izjave o izvoru doprinoseći *globalnom grafu* (engl. *global graph*), što predstavlja cjelokupno dostupno znanje i informacije o određenom izvoru.⁵⁹ Ova zamisao informacijskog grafa uključuje Open World Assumption (OWA), u kojemu se ne može pretpostaviti kako su podaci kompletni jer "uvijek postoji više informacija koje mogu biti poznate".⁶⁰ Slično hipertekstualnom Web-u, u kojemu bilo koji sudionik može stvarati link između dokumenata, semantički web proširuje ovaj isti smisao sveobuhvatnosti šireći ukupno znanje dostupno o danom izvoru kroz izjave doprinijete od strane različitih sudionika (akademske, komercijalne itd.).⁶¹ Upotreba zajedničkih metapodatkovnih rječnika^{62,63} omogućuje sudionicima u ovoj okolini da čine doprinos grafu objavljujući podatke o bilo kojem izvoru i definirajući dodatne odnose između ostalih izvora. Web rječnici definirani upotrebom RDF-a i RDF Schema-e (RDFS) definiraju pojmove i odnose koji povezuju ove izvore na semantičkom Web-u⁶⁴ te se mogu koristiti za ugrađivanje strukturiranih podataka u HTML. Budući da metapodatkovni rječnici predstavljaju različite stupnjeve složenosti, veći rječnici mogu predstaviti mnoge korisne stavke za korisnike. Svrha i primarna vrijednost rječnika je opis i razmjena podataka - korisnost metapodatkovnih rječnika se može odnositi na sposobnost potencijalnih korisnika da uče i razumiju taj rječnik, uspješno primjenjuju rječnik u svom poslu te uspješno primjenjuju termine kako bi se postigli očekivani rezultati. U studiji⁶⁵ o ResearchCyc ontologiji identificira se klasifikacija problema iskoristivosti, koji je povezan s upotrebom veće ontologije gornje razine opisujući opće pojmove i imajući široko područje potencijalne primjene. Ova klasifikacija problematičnih područja uključuje: (a) razumljivost, (b) učenje i (c) operabilnost. Autori su napravili brojne prijedloge za poboljšanje ontologija gornjih razina, uključujući stvaranje bolje dokumentacije

⁵⁹ Antoniou, Grigoris; Van Harmelen, Frank., op. cit. (bilj. 2.), str. 62.

⁶⁰ Allemang, Dean; Hendler, Jim, (bilj. 54.) op. cit., str. 13.

⁶¹ Berners-Lee, T.; Hendler, J.; Lassila, O., loc cit.

⁶² Metapodatkovni rječnik osigurava zajednički skup elemenata, kojega izdavači podataka mogu koristiti za opis informacija. Skup metapodatkovnih elemenata ili schema je model za organiziranje informacija i može također definirati pravila ili standarde za opis izvora. and may also define rules or standards. Metapodatak, odnosno podatak koji opisuje podatak, služi nizu funkcija, uključujući otkrivanje izvora, organiziranje elektroničkih izvora, interoperabilnost, digitalna identifikacija, arhiviranje i čuvanje. // Understanding metadata / by National Information Standards Organization. Bethesda, MD: NISO Press, 2004. Dostupno na URL: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf> (10.02.2014.).

⁶³ Termin "rječnik" i "ontologija" se često koriste naizmjenično kada se objašnjava pojam web rječnika, iako ontologija može definirati mnogo kompleksnije odnose između pojmova i ostalih metapodatkovnih elemenata, kao što su klase i svojstva. // W3C. Vocabularies. Dostupno na: [URL:http://www.w3.org/standards/semanticweb/ontology](http://www.w3.org/standards/semanticweb/ontology) (26.10.2015.).

⁶⁴ W3C. Vocabularies. Dostupno: [URL:http://www.w3.org/standards/semanticweb/ontology](http://www.w3.org/standards/semanticweb/ontology) (26.10.2015.).

⁶⁵ Conesa, Jordi; Storey, Veda C.; Sugumaran, Vijayan. Usability of upper level ontologies: The case of ResearchCyc. // Data and Knowledge Engineering 69, 4(2010), 343-356. Dostupno na URL: <http://www.sciencedirect.com/science/article/pii/S0169023X09001177> (15.06.2015.).

i popratne vizualizacije.⁶⁶ Neka od ovih pitanja upotrebljivosti utvrđena u prethodno navedenoj studiji, kao što je razumljivost, značajno su se ublažila implementirajući rječnike specifičnih domena, gdje će se termini identificirati prema prepoznatljivosti s obzirom na korisničku upoznatost s domenom. U raspravi o sustavima srodne terminologije i rječnika prema očekivanjima korisnika, Nielsen naznačava „treba paziti da se ne koriste riječi u nestandardnim značenjima“⁶⁷ ili koje su inače nepoznate. Čak i veliki domenski rječnici namijenjeni određenoj korisničkoj zajednici, popraćeni opsežnom dokumentacijom, mogu stvoriti slična pitanja o upotrebljivosti za korisnike koji žele implementirati taj rječnik. Jedan pristup za poboljšanje korisnosti metapodatkovnih rječnika je metapodatkovni aplikacijski profil (engl. application profile - AP), koji predstavlja sastavljeni skup elemenata metapodataka iz jedne ili više metapodatkovnih schema koja izražava funkcionalne zahtjeve aplikacije ili zajednice ostajući kompatibilan s osnovnim rječnicima iz kojih su izvedeni.⁶⁸ Koncept AP-a je uspješno primijenjen na različite opće rječnike, posebno Dublin Core (DC), koji osigurava temeljnu shemu za DC-Library Application Profile (DC-Lib), koji predstavlja sastavljeni skup elemenata koji razjašnjava upotrebu DC metapodataka u knjižnicama.⁶⁹

Kada se žele koristiti podaci preko nekoliko izvora, morat će se zajedno sjediniti (engl. merge) ti izvori. Vrijednost prikazivanja trojkama je lakoća s kojom se ova vrsta sjedinjavanja može ostvariti. Budući su informacije predstavljene kao trojke, sjedinjenje informacije iz dva grafa je tako jednostavno, kao što je oblikovanje grafa svih trojki iz svakog pojedinog grafa uzimajući ih zajedno. Suština spajanja svodi se na davanje odgovora na pitanje: "Kada je čvor u jednom grafu isti čvor kao čvor u drugom grafu?" U RDF-u je to pitanje riješeno kroz upotrebu URI-ja (engl. Uniform Resource Identifiers). Čvorovi i svojstva na semantičkom webu ne mogu se imenovati samo imenom jer to nije dovoljna informacija jesu li dva čvora uistinu ista. Ovdje se može postaviti pitanje „Zar ne postoji samo jedna stvar u svemiru za koju se svi slažu referirajući se na Shakespeare?“ Kada se poziva na dogovor na webu, vrijedi „nikad ne reci sve“. Netko će se npr. referirati na Shakespeara književnika, a netko na lik iz filma „Shakespeare in Love“ koji po karakteristikama vrlo malo podsjeća na stvarnu osobu Shakespeara. Sličan primjer je Washington koji može predstavljati još različitije resurse. Kako bi se ostvarilo uspješno spajanje grafova u okruženju semantičkog weba treba se jasno

⁶⁶ Ibid.

⁶⁷ Nielsen, Jakob. Usability engineering. Boston : Academic Press, 1993. Str. 123.

⁶⁸ Duval, Erik... [et al.]. Metadata principles and practicalities. // D-Lib Magazine 8, 4(2002). Dostupno na URL: <http://www.dlib.org/dlib/april02/weibel/04weibel.html> (26.10.2015.).

⁶⁹ Dublin Core Metadata Initiative (DCMI). *Library application profile*. 10 September 2004. Dostupno na URL: <http://dublincore.org/documents/library-application-profile/> (26.10.2015.).

razlikovati na koje se značenje riječi Shakespeare ili Washington misli? Dakle, RDF posuđuje svoje rješenje za ovaj problem iz temeljne web tehnologija, preciznije, kao rješenje uzima URI. Sa sintaksom i oblikom URI-ja upoznati su čak i povremeni korisnici interneta zbog URL –a, na primjer:

<http://www.WorkingOntologist.org/Examples/Chapter3/Shakespeare#Shakespeare>.

No, značaj URI-ja kao globalnog identifikatora za web resurs često nije uočena. URI pruža globalnu identifikaciju za resurs, koja je ista preko cijelog Web-a. U slučaju da se dva agenta na web-u žele referirati na isti resurs preporučena je praksa da pristanu na zajednički URI za taj resurs. Ovo nije ustanovljeno posebno za semantički web nego za web općenito te globalna imenovanja dovode do globalnog mrežnog uspjeha. Zašto onda Web ima obje ove zamisli? Pojednostavljeno, URI je identifikator s globalnim dosegom (tj. "World Wide" u "World Wide Web" smislu) jer se bilo koje dvije web aplikacije u svijetu mogu referirati na istu stvar upućivanjem na isti URI, ali sintaksa URI-ja koristi sve informacije u URI-ju (koji navodi stvari kao što su ime poslužitelja/servera, protokol, broj porta, naziv datoteke, itd.) kako bi se locirala npr. datoteka (ili mjesto u datoteci) na Webu. Kada svi ti dijelovi rade, može se kazati kako URI nije samo URI, nego je također i URL. Dakle, URI nije isto što i URL. Kako bi se pronašla određena stranica na web-u koristi se URL te stranice predstavljajući njezinu lokaciju. Dakle, URL služi za dohvaćanje informacije u digitalnom obliku. Za razliku od njega, URI služi da identificira bilo što, a ne samo lokaciju. Tako URI može identificirati OIB, telefonski broj, zemljopisnu lokaciju, pojam itd. Navedeno upućuje na to kako URI može predstavljati bilo što (čemu se može, a i ne mora pristupiti na web-u), bez obzira je li se tome može pristupiti direktno kao kod URL-a. Dakle, kada se u web preglednik upiše URL, uvijek se vrati lokacija stranice ukoliko ista postoji. No, ako se u web preglednik utipka URI, ne mora se nešto vratiti. Upravo ova razlika čini RDF model iznimno fleksibilnim jer svatko može govoriti o bilo čemu (osobe, predmeti, odnosi među njima, zemljopisne lokacije i sl.), a ne samo o dokumentima koji se nalaze na webu.

RDF primjenjuje pojam URI-ja za rješavanje problema identiteta u sjedinjavanjima grafova. Primjena je vrlo jednostavna: Čvor iz jednog grafa je spojen s čvorom iz drugog grafa, ako imaju isti URI. S jedne strane, može se činiti nepoštenim "rješavanje" problema identiteta čvora oslanjajući se na drugi standard. S druge strane, budući da se pitanja identiteta pojavljuju na web-u u cjelini, a ne samo na semantičkom web-u, bilo bi smiješno da ne koriste istu strategiju za rješavanje problem u oba slučaja. Korištenje URI-ja kao standarda za

globalne identifikatore omogućuje upućivanje diljem svijeta na bilo koji simbol, odnosno bilo kojih dvoje ljudi bilo gdje u svijetu mogu upućivati na istu stvar. Ovo svojstvo URI-ja pruža jednostavan način organizacijama standarda (poput W3C⁷⁰) da odrede značenje (npr. W3C standardi osiguravaju definicije pojmova kao što su `type`, `subClassOf`, `Class`, `inverseOf` i tako dalje). Standardi se primjenjuju na globalnoj razini preko semantičkog weba, tako da se odnose na ove rezervirane riječi na isti način na koji se odnose na bilo koji drugi resurs na semantičkom web-u, tj. kao URI.

W3C je definirao niz standardnih namespace-a za korištenje s web tehnologijama, uključujući `xsd:` za definiciju XML sheme; `xmlns:` za XML namespace-e itd. Semantički web je postupio na isti način s namespace definicijama glavnih slojeva semantičkog web-a. Dakle, RDF koristi URI-je kako bi dao imena resursima i svojstvima. Skupina tih URI-ja koji imaju određenu svrhu i koji se koriste pri opisu određenog dokumenta ili resursa naziva se rječnik (eng. *vocabulary*). Svi URI-ji koji koriste taj rječnik sadrže istu oznaku (koja se u svrhu kraćeg pisanje URI-ja označava nekim prefiksom). Dakle, URI-ji koji koriste određeni rječnik sastoje se od teksta (ili prefiksa) iza kojeg slijede lokalna imena resursa. Najvažniji (korijenski) rječnik je *rdf:* rječnik, koji se XML sintaksi skraćeno pišemo *rdf:*
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> = *rdf:*

Sljedeći standardnu praksu W3C-a koristit će se *qnames*⁷¹ za referiranje na te termine, koristeći sljedeće definicije za standardne namespace-e:

- *rdf:* ukazuje na identifikatore koji se koriste u RDF-u. Skup identifikatora definiranih u standardu je poprilično mali i koristi se za definiranje vrsta i svojstva u RDF-u. Globalni URI za *rdf* namespace je <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
- *rdfs:* ukazuje na identifikatore koji se koriste za RDF Schema jezik (RDFS). Globalni URI za *rdfs* namespace je <http://www.w3.org/2000/01/rdf-schema#>.

⁷⁰ World Wide Web Consortium (W3C), loc. cit.

⁷¹ Problem koji se javlja kod URI-ja su njegova relativno duga imena. Rješenje tomu nudi XML sintaksa, odnosno *XML QName*. *Qname* se sastoji od prefiksa, iza kojega slijedi dvotočka i lokalno ime resursa, npr. ako imamo prefiks (*dbpedia*) i *namespace* (<http://dbpedia.org/page/>), onda se puni URI <http://dbpedia.org/page/Split> može kraće napisati: *dbpedia: Split*
Dakle, *namespace*-ovi služe kako bi se skratila imena URI-ja.

- *owl*: ukazuje na identifikatore koji se koriste web ontologijski jezik OWL (engl. Web Ontology Language). Globalni URI za owl namespace je <http://www.w3.org/2002/07/owl#>.

Ovi URI-ji osiguravaju dobar primjer interakcije između URI-ja i URL-ja. Za svrhe modeliranja, bilo koji URI u jednom od tih prostora (npr. <http://www.w3.org/2000/01/rdf-schema#> *subClassOf* ili ukratko *rdfs: subClassOf*) odnosi se na određeni termin o kojemu je W3C načinio neke izjave u RDFS standardu. No, termin također može biti dereferenciran tj. ako se pogleda na server/poslužitelja www.w3.org, postoji stranica na lokaciji [2000/01/rdf-schema](http://www.w3.org/2000/01/rdf-schema) o *subClassOf*, dajući dodatne informacije o ovom resursu.⁷² Osnovni pojmovi RDF rječnika mogu se razvrstati u četiri kategorije:

- *Sintaksna imena*: *rdf:RDF*, *rdf:Description*, *rdf:ID*, *rdf:about*, *rdf:parseType*, *rdf:resource*, *rdf:li*, *rdf:node*, *rdf:datatype*;
- *Imena klasa*: *rdf:Seq*, *rdf:Bag*, *rdf:Alt*, *rdf:Statement*, *rdf:Property*, *rdf:XMLLiteral*, *rdf:List*;
- *Imena svojstava*: *rdf:subject*, *rdf:predicate*, *rdf:object*, *rdf:type*, *rdf:value*, *rdf:first*, *rdf:rest*, *_n* (*n* je decimalni broj veći od 0, s tim da početna znamenka ne smije biti 0);
- *Imena resursa*: *rdf:nil*

URI-ji ovih pojmova se pišu u RDF-u tako da se spoji RDF *namespace* sa imenom pojma, npr. URI *rdf:type* je izveden od <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

Osnovni pojmovi koji se koriste u RDF/XML-u su *rdf:RDF*, *rdf:Description*, *rdf:about*, i *rdf:resource*. Slijedi njihovo kratko pojašnjenje:

- Element *rdf:RDF* omogućuje aplikacijama prepoznavanje RDF dokumenta, njegov su sadržaj brojni opisi te završava tagom `</rdf:RDF>`, odnosno rdf dokument počinje i završava s `<rdf:RDF>....</rdf:RDF>`.
- Budući je svaka izjava u RDF modelu opis resursa u stvarnom svijetu, koristi se izraz *rdf:Description* kako bi se opisao taj resurs, odnosno subjekt izjave. Sadržaj *rdf:RDF* taga je niz izraza *rdf:Description* od kojih svaki označava početak opisa resursa. Unutar toga elementa stvaraju se izjave, koje se mogu kreirati na jedan od tri načina:

⁷² Allemang, Dean; Hendler, Jim, (bilj. 54.) op. cit., str. 38.

-pomoću *rdf:about* atributa koji se odnosi na već postojeći resurs,

-pomoću *rdf:ID* atributa koji stvara novi resurs,

-bez atributa, stvarajući anonimni resurs.

Budući je `<rdf:Description>` početak izjave, onda `</rdf:Description>` označava njezin kraj.

Stoga, svaka izjava ima oblik:

```
<rdf:Description
rdf:about="..">
....
</rdf:Description>
```

Dakle, pomoću izjave se definira subjekt, predikat (svojstvo) i objekt (vrijednost svojstva) izjave na sljedeći način:

```
<rdf:Description rdf:about="URI subjekta izjave">
<URIpredikata rdf:resource="URI objekta izjave"/>
</rdf:Description>
```

- *rdf:about* je atribut elementa *rdf:Description*. U značenju je ekvivalentan ID atributu, ali se često upotrebljava kako bi se ukazalo na to kako je predmet o kojemu je načinjena izjava već negdje definiran.
- *rdf:resource* atribut unutar nekog svojstva predstavlja objekt trojke. Njime se upućuje na URI resursa koji je korišten kao objekt svojstva, a taj objekt se može opisati pomoću *rdf:Description* čvora, tako da izjava poprimajući sljedeći oblik:

```
1: <?xml version="1.0"?>
2: <rdf:RDF
3: xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4: xmlns:linkbiz="http://www.linkbiz.com/ucenici#">
5: <rdf:Description
6: rdf:about="http://www.linkbiz.com/ucenici#ucenik1">
7: <rdf:type
8: rdf:resource="http://www.linkbiz.com/ucenici#osnovnoskolac"/>
9: <linkbiz:razred
10: rdf:resource="http://www.linkbiz/ucenici #peti"/>
```

```

11: <linkbiz:prolaznost
rdf:resource="http://www.linkbiz.com/ ucenici #uspjeh"/>
12: </rdf:Description>
13: <rdf:Description
14: rdf:about="http://www.linkbiz.com/ ucenici #uspjeh ">
15: <linkbiz:ocjena> (odlican (5))</linkbiz:ocjena>
16: </rdf:Description>
17: </rdf:RDF>

```

U prethodnom primjeru je uočljivo korištenje *rdf:resource* atributa unutar *linkbiz:prolaznost* svojstva, gdje ovaj atribut upućuje na URI resursa koji je korišten kao objekt ovoga svojstva. Taj objekt je posebno opisan pomoću *rdf:Description* čvora (linije 13-16). Ovakav uzorak se može ponovno koristiti sve dok se ne opišu svi resursi, tj. ako resurs *linkbiz:uspjeh*, koristi neki drugi resurs kao vrijednost svoga svojstva (tj. kao objekt), umjesto vrijednosti *ocjena* u liniji 15, taj resurs može ponovno biti opisan unutar odgovarajućeg svojstva, sve dok se ne opišu svi resursi koji se žele opisati. Na navedenom primjeru je uočljivo u retku 1 kako je dokument u XML formatu te je navedena verziju 1.0 XML dokumenta koja se koristi u danom dokumentu. Redak 2 stvara *rdf:RDF* element koji završava sa tagom *</rdf:RDF>*, kako bi aplikacije mogle prepoznati da je riječ o RDF dokumentu. Redak 3 u *rdf:RDF* element uključuje i XML namespace koristeći *xmlns* (što je kratica za XML *Namespace*) atribut koji označava da se prefiks *rdf:* odnosi na URI <http://www.w3.org/1999/02/22-rdf-syntax-ns#>, odnosno:

```
xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
```

Unutar *rdf:* elementa moguće je definirati bilo koji URI pomoću *xmlns* atributa, pa tako u liniji 4 imamo definiran novi URI <http://www.linkbiz.com/ucenici#>, koji je označen sa prefiksom *linkbiz:*

```
xmlns:linkbiz="http://www.linkbiz.com/ucenici#">.
```

Kada se završilo definiranje URI-ja koristi se „>“ znak, kako bi se zatvorila *rdf:RDF* oznaka. Izraz je *rdf:type* identificira vrstu danog resursa (npr. *truba* je vrsta puhačkih instrumenata).

U nastavku će se ukratko spomenuti i ostali pojmovi RDF rječnika:

- *rdf:parseType* i *rdf:value* se koriste kada se želi izraziti vrijednost nekoga svojstva pomoću vrste toga svojstva.

- *rdf:datatype* atribut se koristi kako bi se objasnilo koji se tip podataka nalazi u resursima tj. radi li se o broju, tekstu, mjeri u kilogramima, centimetrima i sl., kako bi računalo znalo koja je vrsta podatka. Upravo zbog tog razloga XML sintaksa unutar rdf-a nudi mogućnost označavanja vrste podataka, koji se koristi kada se želi izraziti vrijednost nekog svojstva i vrste podatka koji se nalazi u tome svojstvu. Budući da RDF/XML sintaksa nema određene tipove podataka, koriste se tipovi iz XML Schema-e.

RDF/XML sintaksa omogućava i rad sa spremnicima, koji se koriste kada postoji veći broj resursa ili atributa koje se želi opisati kao cjelinu (npr. razred učenika, glazbeni bend, poduzeće i sl.) tako što će se unutar njega npr. dodati učenici koji idu u određeni razred ili će se opisati djelatnosti kojima se bavi određeno poduzeće. Spremnik je resurs koji unutar sebe sadrži grupu stvari, a te stvari se zovu članovi (eng. members) toga spremnika. Tri su osnovna elementa koja se koriste kao spremnici, a to su:

- *rdf:Bag* element koji se koristi kada poredak stvari u spremniku nije važan, tj. kada lista vrijednosti unutar spremnika ne mora imati određen redoslijed te može sadržavati istovjetne (duplicirane) vrijednosti.,
- *rdf:Seq* element koji se koristi kada je redoslijed članova važan (npr. kada želimo članove poredati po abecedi ili važnosti unutar poduzeća te može sadržavati istovjetne (duplicirane) vrijednosti.
- *rdf:Alt* element koji se koristi kada želimo predstaviti skup alternativa, a korisnik može odabrati samo jednu od njih (npr. prijevod dokumenta ili nekog pojma na druge jezike).
- Atribut `<rdf:parseType="Collection">` se koristi ukoliko se želi ograničiti dodavanje drugih članova u spremnik. Ovaj element zatvara spremnik i omogućava da se unutar spremnika nalaze samo određeni članovi, bez mogućnosti dodavanja drugih članova. Ovaj atribut koristi unaprijed definirani rječnik s elementima *rdf:first*, *rdf:rest*, *rdf:nil*. Članovi unutar spremnika će u grafičkom prikazu biti povezane pomoću *rdf:first* i *rdf:rest* elementa, s tim da je zadnji element označen sa *rdf:nil* resursom. Kako drugi RDF dokumenti ili aplikacije ne mogu pristupiti članovima ovoga spremnika (zato jer

ne mogu koristiti prazne elemente spremnika kako bi došli do elemenata ispod njih), pomoću ove sintakse se može zatvoriti dani spremnik.

Ukoliko se želi izraziti npr. kako je resurs „osoba xy“ napisao članak *Linked Data* u 2011. godini, navedeno se ne može opisati pomoću jedne trojke jer je trojka jedino dovoljna za izraziti da je osoba xy (subjekt) napisao (predikat) članak *Linked Data* (objekt), dok je za dodatne izjave potrebna druga trojka. RDF nudi rješenje za ovaj problem u obliku reifikacije tj. dodavanja dodatne izjave o nekoj izjavi. Tako se za opis RDF izjave pomoću RDF-a, koriste `<rdf:statement>`, `<rdf:subject>`, `<rdf:predicate>` i `<rdf:object>` elementi. RDF je apstraktni model prikazivanja podataka, koji ne određuje način na koji će RDF biti reprezentiran. Osim RDF/XML-a, postoje i drugi formati za predstavljanje RDF-a, kao što su *Notation3 (N3)*, *Turtle* i *N-triples*.

2.4.2 RDF Shema (RDFS)

Resurs se u RDF-u opisuje predikatima koji daju dodatne informacije o njemu. Predikati nisu proizvoljni, nego su unaprijed definirani pomoću zajedničkoga jezika (npr. upotreba bilo kojeg izraza Dublin Core rječnika⁷³, kao što je `dc:Creator`, `dc:title`, `dc:Date`). Svi izrazi koji imaju prefiks *rdfs:* su definirani pomoću RDFS-e, koja omogućuje korištenja zajedničkoga jezika za stvaranje klasa, podklasa, svojstava te veza među tim klasama i svojstvima. RDFS proširuje RDF na način što uvodi skup karakterističnih resursa koji opisuju

⁷³ Dublin Core (DC) ili Dublinski osnovni skup elemenata metapodataka (Dublin Core Metadata Element Set – DCMES) standard je za opis informacijskih izvora različitih područja informacije. Dakle, Dublin Core je osnovni standard metapodataka za opis mrežnih izvora radi pronalazjenja i razmjene informacija. Za razliku od drugih, namijenjenih potrebama određene strukovne zajednice (poput Encoding Archive Initiative za arhive, Computer Interchange of Museum Information (CIMI) za muzeje ili (UNI)MARC za knjižnice), DCMES je skup općenito razumljivih opisnih elemenata koji služe razmjeni među različitim zajednicama korisnika. DCMES je rezultat međudisciplinarnog i međunarodnog dogovora na kojem su zajednički radili od 1995. knjižničari, istraživači u području digitalnih knjižnica, raspačavatelji bibliografskih i sličnih podataka te stručnjaci za označavanje elektroničkih tekstova. Registriran je kao norma Američkog instituta za nacionalne standarde ANSI/NISO pod oznakom Z39.85-2001. Taj skup od 15 elemenata metapodataka po svojoj se zadaći dijeli u tri skupine: elementi za sadržajni opis (naslov, tema i ključne riječi, opis, tip, izvor, odnos, obuhvat), elementi za određivanje vlasničkih i autorskih prava (stvaratelj, nakladnik, suradnik, prava) i elementi za određivanje oblika pojavnosti u datom vremenu i prostoru (datum, jezik, format, identifikatori). Redoslijed DCMES elemenata nije određen standardom i nije obvezno navesti sve opisne elemente, a svaki element može se pojaviti više puta. Elementi za opis nekog podatka mogu se izabrati iz nadziranog rječnika, koji predstavlja ograničen skup često korištenih i precizno definiranih izraza. // Sarić, Ivana; Magdić, Antonio; Essert, Mario. Sheme metapodataka značajne za knjižničarstvo s primjerom implementacije OpenURL standarda / Vjesnik bibliotekara Hrvatske 54, 1/2(2011), 138-139.

metapodatke. Dok RDF opisuje resurse i podatke pomoću metapodataka, RDF Schema omogućava opisivanje tih metapodataka. RDFS je u biti jezik pomoću kojeg se stvara rječnik o određenoj temi ili području pomoću klasa, podklasa i svojstava RDF resursa. Primjer rječnika stvorenog pomoću RDFS-a je Dublin Core rječnik.

Može se govoriti o dvije specifikacije RDF-a: sintaksi RDF-a, koja pokazuje kako se RDF elementi povezuju jedan s drugim pomoću XML-a, turtle-a, N3-a i dr. (npr. elementi kao što su `rdf:type` i `dc:Creator` se koriste kako bi se opisao određeni resurs, te daju informaciju o tome koji je tip resursa i tko je autor danog resursa), i shemi RDF-a pomoću koje se definiraju ti elementi (*namespace* prefiksi ispred svakog elementa, npr. *rdf* i *dc*, predstavljaju shemu pomoću koje je taj određeni element definiran). Dakle, pomoću sheme se definiraju rječnici koji se odnose na neko područje, a ti rječnici definiraju elemente i načine na koji se ti elementi povezuju jedan s drugim. Elementi jednog rječnika se mogu povezati s elementima drugog rječnika. RDF Schema je definirana pomoću RDF-a, tj. sve informacije u RDFS-u su definirane pomoću RDF trojki. Elementi RDF Scheme se označavaju sa sljedećim *namespace*-om unutar dokumenta:

`xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#`

Unutar RDF Scheme je definirana grupa klasa i svojstava koji se koriste za opis određenih RDF elemenata, a te klase i svojstva su temeljni dio sheme. Prilikom rada s RDFS-om, važna su sljedeća pravila za svojstva i klase: instance klasa nasljeđuju svojstva klasa, podsvojstva nasljeđuju vrijednosti svojstava koja su iznad njih te se svojstva i klase u RDFS-i definiraju odvojeno - tj. nakon što se definiraju klase (i veze među klasama), moraju se zasebno definirati svojstva te zatim povezati ta svojstva sa željenim klasama. Ako se ne poveže određeno svojstvo sa nekom klasom, onda je to svojstvo nezavisno, tj. može stajati samo za sebe i služiti da se opiše bilo koja klasa. Zbog ovoga svatko može govoriti o bilo kojemu resursu jednostavno tako što mu dodaje još svojstava. Ovo je posebno važno, budući da se dodavanjem tih svojstava ne mijenja aplikacija koja se napravila prije nego su svojstva bila dodana, tj. ona se ne treba doradivati svaki put kada netko doda novo svojstvo. Bez zasebnog definiranja klasa i svojstava ovo ne bi bilo moguće.

2.4.2.1 Klase RDF Sheme

Klasa je skup sličnih stvari te je dopušteno nasljeđivanje među njima, a njihovo postojanje omogućava definiranje resursa RDF-u kao instance jedne ili više klasa i hijerarhijsku organizaciju klasa. Primjerice, ako je klasa A podklasa klase B, a klasa B je opet podklasa klase C, svaki resurs klase A također pripada i klasi C. Također će klasa A naslijediti sva svojstva klase B, koja će naslijediti sva svojstva klase C; a sve tri ove klase su podklase klase *Class*.

Svaki resurs unutar RDF dokumenta ima zajedničku klasu *rdfs:Resource*, a svi novi resursi, uključujući i resurs *rdfs:Resource*, definirani su pomoću *rdfs:Class* izjave. Potrebno je uočiti kako *rdf:resource* iz prethodnog odjeljka nije isto što i *rdfs:Resource* jer je *rdf:resource* XML atribut koji se upotrebljava s određenim elementom kako bi se definiralo svojstvo nekog elementa (npr. koristi se sa *rdf:type* elementom kako bi se pokazalo da je vrijednost toga svojstva neki drugi element), a *rdfs:Resource* je klasa definirana u RDF Shemi (zbog toga je *rdf:resource* *case sensitive* te će validator izbaciti grešku ako se napiše *rdf:Resource*) te je već određena kao zadana u RDF-u. Stoga se ne koristi često unutar dokumenta kojim se definira novi rječnik jer je svaki resurs unutar RDF dokumenta početno podklasa *rdfs:Resource* klase, koja je zajednička klasa svih resursa, budući se nalazi iznad svih klasa koje se definiraju te je kao takvu nije potrebno zasebno definirati.

rdfs:Container je nadklasa svih spremnika: *rdf:Bag*, *rdf:Seq* i *rdf:Alt*, dok se *rdfs:ContainerMembershipProperties* klasa sastoji od samih elemenata spremnika.

rdf:Bag, *rdf:Seq* i *rdf:Alt* se koriste kako bi se grupirali članovi spremnika.

rdfs:Datatype je klasa svih tipova podataka, te je podklasa od *rdfs:Literal* klase, koja se koristi kako bi se opisale literalne vrijednosti (kao što je npr. *rdfs:Comment*).

Skup klasa koje se koriste u RDFS-i je prikazan u Tablici 2.

Tabela 2: Klase RDF Scheme

RDF klase			
Element	Klasa	Podklasa	Opis
<i>rdfs:Class</i>	Klasa		Vrsta ili kategorija resursa
<i>rdfs:Datatype</i>	Tipova podataka	<i>Class</i>	Informacija o tipu podataka
<i>rdfs:Resource</i>	Svih resursa	<i>Class</i>	Svi resursi unutar RDF-a su članovi ove klase
<i>rdfs:Container</i>	Spremnika	<i>Resource</i>	Superklasa svih klasa

			spremnik
rdfs:Literal	Literarnih vrijednosti (tekstualnih i numeričkih)	Resource	Literarne vrijednosti u RDF-u, kao što su brojevi, stringovi i sl.
rdf:List	Lista	Resource	Klasa svih RDF lista
rdf:Property	Svojstva	Resource	Resursi koji su RDF svojstva (properties)
rdf:Statement	Izjava	Resource	Klasa svih RDF izjava
rdf:Alt	Spremnik alternativa (zamjena)	Container	
rdf:Bag	Neuređeni spremnici	Container	
rdf:Seq	Uređeni spremnici	Container	
rdf:ContainerMembershipProperties	Spremnik svojstava članova	Property	U ovoj klasi su sadržani članovi spremnika
rdf:XMLLiteral	XML Literalne vrijednosti	Literal	Literalne vrijednosti u RDF-u koje koriste XML sintaksu

2.4.2.2 RDF svojstva i atributi

Svojstvo *rdfs:subClassOf* identificira klasu koja je podklasa neke druge klase. Koncept nasljeđivanja unutar RDF Sheme omogućava definiranje ovakvih tzv. nadklasa. Nove podklase klase tada nasljeđuju svojstva i ograničenja nadklase. Dakle, instanca klase uvijek nasljeđuje svojstva svoje klase. Povezujući klase jednu s drugom na ovaj način dobiva se mogućnost da se pomoću RDFS-e opišu veze među elementima koji su opisani pomoću RDF trojki. Međutim, ukoliko se želi dati značenje podacima treba se moći izraziti nešto više o samim elementima i o svojstvima pomoću kojih ih se povezuje, tj. o predikatima RDF trojki, što omogućava *rdfs:subPropertyOf* svojstvo u RDF Shemi. Svojstvo *rdfs:subPropertyOf* se koristi kada se žele dodati dodatna pojašnjenja drugom svojstvu. Može se dodati i više *subPropertyOf* svojstava. Tako se za jedno svojstvo može kazati da je podsojstvo (*subPropertyOf*) nekoga drugog svojstva pomoću trojke P *rdfs:subPropertyOf* R. Ako postoji trojka koja koristi svojstvo A (koje je podsojstvo svojstva R)

A P B.

onda se može zaključiti i da je svojstvo te trojke i R (budući da je P podsojstvo od R, te nasljeđuje vrijednosti od R)

A R B.

Svojstvo *rdfs:seeAlso* se koristi kako bi se identificirao drugi resurs koji sadrži neke dodatne informacije o resursu koji se opisuje te se također koristi kako bi se povezao

dokument u kojemu se opisao rječnik sa nekim drugim dokumentom. Uz pomoć njega se u shemu mogu dodati neki dodatni elementi, bez da ju se treba prepravljati. Svojstvo *rdfs:seeAlso* se može nadograditi pomoću *rdfs:subPropertyOf* svojstva kako bi se dale dodatne informacije o načinu na koji jedan resurs daje dodatne informacije o drugom resursu. Svojstvo *rdfs:isDefinedBy* određuje namespace resursa, tako što daje URI za određenu shemu, npr. unutar RDFS rječnika klasa *rdf:Statement* je definirana kao namespace unutar RDF sintakse, dok je *rdfs:Literal* definiran kao *namespace* unutar RDF Scheme.

Dva svojstva koja omogućavaju pojašnjenja za ljudske čitatelje RDF modela su *rdfs:comment* i *rdfs:label*. Svojstvo *rdfs:comment* služi za dokumentiranje resursa, dok *rdfs:label* služi za pisanje imena resursa kako bi bilo bolje razumljivo ljudima (ovom elementu se može dodati atribut *xml:lang* kako bi se dale oznake za korištene jezike). Ovo je posebno korisno jer se može dogoditi da rječnik i schema nemaju povezanu dokumentaciju ili da link između dokumentacije i rječnika nije održavan. Dajući komentare i čitljive oznake (eng. *label*), korisniku se daje informacije o tome što se definira. Ovo je također korisno budući da ponekad sama imena klasa ili svojstava ne daju dovoljno informacija o tome koja im je svrha. Npr. ako se pogleda *Tablicu 2.* s klasama ili *Tablicu 3.* sa svojstvima stupac „Element“ ne daje mnogo informacija o klasi ili svojstvu, dok stupac „Opis“ daje dodatne informacije koje otkrivaju svrhu danog elementa. Svi opisi u tome stupcu u RDF-u su dani uz pomoć *rdfs:comment* svojstva.

Svojstvo *rdf:type* definira vrstu resursa. Prethodno je već utvrđeno kako svi resursi pripadaju vrsti *rdfs:Resource*. Svojstvo *rdf:type* ukazuje kako je resurs koji se opisuje instanca te klase. Ovo svojstvo se podrazumijeva unutar RDF/XML dokumenta i nije izričito dano, tj. možemo ga koristiti, ali je to nepotrebno.

Svojstva *rdf:subject*, *rdf:predicate* i *rdf:object* se koriste kod reifikacije kako bi se dale dodatne informacije o RDF izjavi.

Svojstva *rdf:first* i *rdf:next* se koriste kako bi se odredile definicije veza unutar kolekcije (*Collection* atribut je korišten u RDF odjeljku, kod spremnika). Kada se opisuju termini koji se koriste za podatke, ponekad je korisno iskazati kako neko svojstvo može biti predstavljeno u odnosu na definiranu klasu, tj. želi se pomoću nekoga svojstva iskazati kako je subjekt unutar trojke u kojoj se koristi to svojstvo dolazi iz jedne klase, dok objekt dolazi iz druge. U tu svrhu se koristi *rdfs:domain* i *rdfs:range*. U matematičkom smislu, domena (eng. *domain*) neke funkcije je skup vrijednosti za koje je ta funkcija definirana, dok je doseg (eng. *range*) te funkcije skup vrijednosti koje ta funkcija može poprimiti. Npr. funkcija kvadratni korijen kao svoju domenu ima sve pozitivne brojeve

(budući da u skupu realnih brojeva negativni brojevi nemaju kvadratni korijen), dok su doseg te funkcije svi realni brojevi (budući da postoji i pozitivan i negativan kvadratni korijen).⁷⁴

Isto tako, u RDFS-i svojstvo ili predikat P se može izraziti pomoću trojki

P rdfs:domain D

P rdfs:range R

što znači da se relacija ili predikat P odnosi na sve vrijednosti od klase D do klase R.

U RDFS-u *domain* znači sljedeće: ako P (predikat) *rdfs:domain* pripada domeni klase D i postoji trojka subjekt P Objekt, onda će i subjekt pripadati klasi D, tj. subjekt *rdf:type D*.

Isto tako *range* znači sljedeće: ako P (predikat) *rdfs:range* ima doseg do klase R i postoji trojka subjekt P Objekt, onda će objekt pripadati klasi R, tj. objekt *rdf:type R*. Ovo objašnjava kako se može koristiti svojstvo P. Trojka koja koristi P kao predikat, domena je subjekt te trojke dok je doseg (*range*) objekt. Kada se kaže da svojstvo P ima domenu D, zapravo se govori da kada se god koristi to svojstvo P, subjekt te trojke u kojoj se koristi pripada klasi D. Isto tako, kada se kaže da svojstvo P ima doseg R, zapravo se govori da kada se god koristi to svojstvo P, objekt te trojke u kojoj se koristi pripada klasi R.⁷⁵ Svojstva *rdfs:range* i *rdfs:domain* se koriste kako bi se dodala neka ograničenja u rječnik, koja definiraju koje se klase mogu koristiti za određena svojstva. Svojstvo *rdfs:domain* se koristi da se odredi koja svojstva mogu biti korištena unutar neke klase, tj. da se poveže svojstvo sa resursom koji se opisuje (subjekt trojke). Svojstvo *rdfs:range* se koristi da se odrede koje klase neko svojstvo može imati za svoje vrijednosti (objekt trojke). Npr. nekada se *naslov* može koristiti kao naslov nečega kao što je Članak, kao i za naslov nečega kao što je npr. Knjiga, Slika ili bilo što drugo. Međutim, ukoliko se želi ograničiti da se svojstvo *naslov* koristi samo unutar klase Knjiga, onda se koristi *rdfs:range* koji govori kako npr. vrijednost svojstva *naslov* može biti samo nešto što se nalazi u klasi Knjiga.

Skup svih svojstava (eng. *property*) i atributa RDF Scheme je prikazan Tablicom 3.

⁷⁴ Allemang, Dean; Hendler, Jim, (bilj. 54.) op. cit., str. 130.

⁷⁵ Ibid.

Tabela 3: Svojstva i atributi RDFS Sheme

RDF svojstva			
<i>Element</i>	<i>Domena</i>	<i>Doseg</i>	<i>Opis</i>
rdfs:domain	Property	Class	Domena resursa
rdfs:range	Property	Class	Doseg resursa
rdfs:subPropertyOf	Property	Property	Svojstvo koje je podsvojstvo svojstva
rdfs:subClassOf	Class	Class	Resurs je podklasa klase
rdfs:comment	Resource	Literal	Opis resursa za ljudske čitatelje
rdfs:label	Resource	Literal	Ime (oznaka) resursa za ljudske čitatelje
rdfs:isDefinedBy	Resource	Resource	Definicija (određivanje) resursa
rdfs:seeAlso	Resource	Resource	Dodatne informacije o resursu
rdfs:member	Resource	Resource	Član danog resursa
rdf:first	List	Resource	
rdf:rest	List	List	
rdf:subject	Statement	Resource	Subjekt danog resursa u RDF izjavi
rdf:predicate	Statement	Resource	Predikat danog resursa u RDF izjavi
rdf:object	Statement	Resource	Objekt danog resursa u RDF izjavi
rdf:value	Resource	Resource	Svojstvo koje se koristi za vrijednost
rdf:type	Resource	Class	Resurs je instanca klase
RDF Atributi			
<i>Element</i>	<i>Opis</i>		
rdf:about	Definira resurs koji se opisuje		
rdf:Description	Spremnik za opis toga resursa		
rdf:resource	Definira resurs kako bi se identificiralo svojstvo		
rdf:datatype	Definira tip podataka u elementu		
rdf:ID	Definira ID elementa		
rdf:li	Definira listu		
rdf:_n	Definira node (čvor)		
rdf:nodeID	Definira ID noda elementa		
rdf:parseType	Definira kako bi element trebao biti analiziran (parsiran)		
rdf:RDF	Korjen RDF dokumenta		
rdf:base	Definira bazu XML-a		
rdf:lang	Definira jezik sadržaja elementa		

2.4.3 OWL

Ontologijom se definiraju pojmovi koji služe za opis neke domene⁷⁶, odnosno opisivanje i predstavljanje znanja nekog područja. Ontologija se sastoji od pojmova i veza među tim pojmova. Pojmovi se nazivaju klase ili koncepti. Veze među klasama su

⁷⁶ Domene su određena područja koja se žele opisati, poput poslovanja, medicine, obrazovanja i slično.

hijerarhijske prirode te postoje nadklase (super –klase) koje predstavljaju koncepte veće razine i pod –klase koje predstavljaju specifičnije koncepte, a koji sadrže sve attribute i značajke koncepata više razine. Osim veza među klasama, postoje i svojstva tih klasa, koja opisuju razne značajke i attribute koncepata te se koriste za vezivanje ili udruživanje klasa. Ontologija se definira prema Thomas Gruber-u kao „jasna specifikacija konceptualizacije“. U ovom kontekstu, konceptualizacija predstavlja objekte, pojmove i entitete unutar domene i apstraktan pogled odnosa koji postoje između njih.⁷⁷

Najpoznatiji jezik za označavanje ontologija jest OWL (engl. Web Ontology Language). Svrha OWL-a je definirati ontologije koje uključuju klase, svojstva i veze među njima kako bi se opisalo određena područje te se ti pojmovi koriste u RDF dokumentima za opis nekog resursa, a aplikacija koja koristi te dokumente temelji svoje zaključivanje na ontologijama otkrivajući nove ili već dane činjenice o resursu koji je opisan. Međutim, za razliku od RDFS-e, OWL omogućava kompleksnije i bogatije veze među tim pojmovima te se mogu izgraditi aplikacije s naprednijom razinom zaključivanja. OWL nudi razne sintakse za kreiranje ontologija (npr. Manchester sintaksa, OWL/XML sintaksa, Functional Style sintaksa i sl.), međutim jedina obvezna sintaksa koja mora imati podršku u svim OWL alatima je RDF/XML sintaksa. OWL mora biti u mogućnosti prikupiti informacije iz distribuiranih resursa. To je jednim dijelom omogućeno tako što se dopušta ontologijama da se povežu jedna s drugom, tj. da uvezu informacije iz drugih ontologija. OWL pretpostavlja otvorenost resursa, što znači da resursi nisu ograničeni za jednu arhivu (eng. *file*) ili područje. Ako se definira neka klasa u jednoj ontologiji, ta klasa se može nadopunjavati u drugoj ontologiji, ali te informacije pomoću kojih se nadopunila ta klasa ne mogu poništiti prethodnu informaciju koja je već definirana, iako ta informacija može biti proturječna. Dakle, činjenice se u ontologiju mogu dodavati, ali se nikada ne mogu izbrisati. RDFS je rječnik za oblikovanje podataka (engl. data-modeling vocabulary) koji dodaje dodatne mehanizme RDF modelu za opisivanje odnosa između grupe izvora te se može upotrijebiti za izgradnju metapodatkovnih rječnika. To je potpuno izvedeno kroz očitovanja klasa, podklasa i svojstava za opis pojmova ili pozivanje na postojeće komponente drugih rječnika bez potrebe da ih se definira.⁷⁸ Ova metodologija je posebno korisna kada se implementiraju opći RDFS rječnici (Dublin Core (DC), schema.org), gdje je potrebno dodatno pročišćavanje ili definiranje složenijih odnosa

⁷⁷ Gruber, T. Toward principles for the design of ontologies used for knowledge sharing. // International Journal of Human-Computer Studies 43, 5-6(1995), 907-928.

⁷⁸ W3C. *RDF Schema 1.1*. 25 February 2014 Dostupno na URL: <http://www.w3.org/TR/rdf-schema> (26.10.2015.).

između pojmova i entiteta za bolje predstavljanje podataka određene domene znanja. Dok RDFS osigurava osnovno sredstvo za opisivanje veza između izvora, OWL osigurava proširene mogućnosti za bogatije rječničke definicije, oblikovanje složenijih odnosa između pojmova i unaprjeđenje semantičkog zaključivanja.

Ontologije pružaju mnogo više mogućnosti od RDFS-e te se za stvaranje ontologija najčešće koristi OWL jezik. Neke od prednosti ontologija naspram RDFS-e su:

- RDFS-om se ne može iskazati da dva ili više svojstva imaju jednaku vrijednost,
- RDFS ne nudi mogućost izražavanja znanja kako je neko svojstvo jedinstveno ili je važnije od nekog drugog svojstva. Ponekad se želi kazati da određeno svojstvo može imati samo jednu vrijednost u određenoj instanci klase,
- RDFS-om se može izraziti vrijednosti određenoga svojstva, ali se ne može reći npr. da je to zatvoren skup svojstava,
- RDFS-om se ne mogu iskazati negativne vrijednosti, npr. ne može se reći da *Osoba1 nije ženskog spola*,
- Element *owl:AllDifferent* je definirana OWL klasa, koja ima svojstvo *owl:distinctMembers*, koje povezuje instancu *owl:AllDifferent* sa listom svih pojedinih članova kolekcije. Značenje ove izjave je da su svi članovi različiti jedan od drugih, što se ne može izraziti RDFS-om.

OWL je semantički web standard koji izražava formalno definirano značenje kroz očitovanje klasa, svojstava i instanci.⁷⁹ Osigurava unaprjeđenje semantičkog zaključivanja kroz izražavanje dosega (engl. range) domene i ograničenja, kao i mogućnost izvođenja zaključaka o odnosima koji postoje unutar podataka.^{80,81} W3C utvrđuje kako ne postoji jasna razlika između termina ontologije i rječnici u odnosu na Semantički web, te ontologije mogu izraziti pojmove i odnose o domeni zbog raznolikih stupnjeva složenosti ili jednostavnosti, ovisno o potrebama primjene.⁸² Korisnost upotrebe ontologije kao strukture za definiranje rječnika metapodataka je jer formalna definicija termina i taksonomije razvijena kroz odnos

⁷⁹ W3C. *OWL 2 Web ontology language document overview (second edition)*. 12 December 2012. Dostupno na URL: <http://www.w3.org/TR/owl2-overview> (26.10.2015.):

⁸⁰ Segaran, Toby; Evans, Colin; Taylor, Jamie. *Programming the Semantic Web*. Beijing ; Sebastopol, CA : O'Reilly, 2009. Str. 135.

⁸¹ W3C. *OWL 2 Web ontology language document overview (second edition)*, loc. cit.

⁸² W3C. *Vocabularies*, loc. cit.

klasa i svojstava predstavlja precizniji način komuniciranja značenja i razumijevanja između sustava.⁸³

U arhitekturi semantičkog weba iznad RDFS-e je izgrađen je moćniji OWL sloj, kojim je moguće opisati značenje i semantiku. Svrha mu je bilo koji logički princip izraziti tako da aplikacije mogu izvući zaključke na temelju te opisane logike, a glavno obilježje ovoga sloja jest povjerenje. Budući je OWL izgrađen kao dodatni sloj na RDFS-u, svi izrazi koji se koriste u RDFS-i se mogu koristiti i pri stvaranju OWL dokumenata. Stoga se može zaključiti kako se OWL sastoji od RDFS-e i novih izraza za bolje izražavanje. Postoje dvije verzije OWL jezika, a to su-OWL1 (standard od 2004.g.)⁸⁴ i OWL2 (standard od 2008.g.)⁸⁵. OWL 2 je u biti nadogradnja na OWL 1, gdje su dodane neke nove značajke, stoga se OWL 1 može smatrati podvrstom OWL-a 2 jer će sve ontologije stvorene pomoću OWL 1 jezika biti razumljive i aplikacijama koje koriste OWL 2 jezik. OWL koristi skup pojmova ili izraza pomoću kojih se definiraju klase i svojstva. Ti izrazi imaju URI

<http://www.w3.org/2002/07/owl#>, a kao prefiks se koristi *owl* tako da u RDF/XML dokumentu se koristi XML namespace `xmlns:owl=http://www.w3.org/2002/07/owl#`

Osnovni pojmovi koje se koriste u OWL-u su *aksiomi* (eng. *axiom*), *entiteti* (eng. *entity*), *izrazi* (eng. *expresion*) i *IRI imena* (eng. *IRI names*). Aksiom je osnovna izjava u OWL ontologiji koja predstavlja osnovni dio znanja. Npr. izjava „*Klasa Clanak je podklasa klase WebStranica*“ je aksiom. Na svaku ontologiju se može gledati kao na skup aksioma. Pretpostavka je da su svi aksiomi unutar neke ontologije istiniti (točni). To znači sljedeće: ako postoji aksiom „*Antun Gustav Matoš je napisao pripovijetke Iverje*“ i ako se postavi upit „*Je li Antun Gustav Matoš napisao pripovijetke Novo Iverje*“ odgovor na postavljeni upit će biti „*Možda ili nepoznato*“, a odgovor na upit neće biti „*Ne, nije*“ budući da je *Antun Gustav Matoš* mogao napisati oba djela.⁸⁶ Svaki aksiom se mora sastojati od određenih dijelova, tj. aksiom (kao i izjava) mora imati neku klasu, neko svojstvo i (nekada) neku vrijednost ili individu. Ti dijelovi se zovu *entiteti*. Stoga će za prethodni aksiom „*Iverje je napisao Antun Gustav Matoš*“ postojati trojka `< Iverje : napisao Antun Gustav Matoš >` gdje je individua *Antun Gustav Matoš* entitet koji se naziva *objekt*, *Iverje* je entitet koji se naziva *kategorija*, dok je svojstvo `: napisao` *relacija*. Može postojati i neki drugi aksiom te ga kombinirati s

⁸³ Segaran, Toby; Evans, Colin; Taylor, Jamie, op. cit. (bilj. 80), str. 146.

⁸⁴ W3C. OWL Web Ontology Language Guide. 10 February 2004. Dostupno na URL: <http://www.w3.org/TR/owl-guide/> (26.10.2015.).

⁸⁵ W3C. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. 02 December 2008. Dostupno na URL: <http://www.w3.org/TR/2008/WD-owl2-syntax-20081202/> (26.10.2015.).

⁸⁶ usp. Pollock, Jeffrey T. *Semantic Web for Dummies*. Indianapolis: Hoboken, N.J. : Wiley, 2009. Str. 202.

ovim. Jedna od ključnih značajki OWL-a je kombiniranje kategorija i/ili relacija kako bi se dobili novi, kompleksniji opisi novih entiteta koji se nazivaju *izrazi* (engl. *expression*) te su oni jedan od glavnih razloga zbog kojeg OWL ima napredniju moć izražavanja naspram RDFS-e.

IRI (engl. *Internationalized Resource Identifiers*) imena imaju istu svrhu kao i URI-ji, ali se za pisanje IRI-ja može koristiti veći broj znakova. Dok URI-ji koriste *ASCII* način kodiranja znakova, koji se temelji na engleskoj abecedi i sastoji se od 127 znakova, IRI-ji koristi *Unicode* standard za razmjenu podataka koji koristi preko milijun znakova, tako da se mogu imati resursi koji sadrže znakove napisane na svim jezicima i pismima (npr. znakovi *č,ć,š,đ,ž* u hrvatskom jeziku, ćirilično pismo i sl.); što je i bila glavna motivacija za IRI-je. IRI-ji se mogu konvertirati u URI-je i obratno, što je korisno kada se koriste IRI-ji u protokolu poput HTTP-a, koji prihvaća samo URI-je.

Svaka ontologija se nalazi unutar *rdf:RDF* elementa, a ontologija započinje sa deklaracijom *namespace*-ova koji se koriste u njoj. Grupa tvrdnji u ontologiji se može ispisati unutar *owl:Ontology* oznake, koja omogućava opis ontologije pomoću komentara, oznaka te uvoz druge ontologije. Zaglavlje ontologije počinje sa *<owl:Ontology>* , a zatvara sa *</owl:Ontology>* oznakom.

Element *rdf:about* dodjeljuje ime ontologiji, a *rdfs:comment* opisuje ontologiju. Element *owl:priorVersion* omogućava pristup prijašnjim verzijama ontologije (ako postoje).

Element *owl:imports* uvozi ontologiju koja je definirana sa *rdf:resource* atributom. To nije isto kao i definiranje te ontologije u zaglavlju pomoću *namespace*-a, budući da *owl:imports* zapravo uvozi tu određenu ontologiju, ali uvozi i sve druge ontologije koje su uvezene u tu drugu ontologiju.

2.4.3.1 OWL1 jezik

Elementi OWL 1 jezika se odnose na klase, svojstva i instance tih klasa, te veze među njima.

2.4.3.1.1 Definiranje klasa

Klase se u OWL1 jeziku definiraju sljedećim elementima. Zadana osnovna klasa u OWL jeziku je *owl:Thing* (kao što je i *rdfs:Resource* u RDFS-u), što znači kako je svaka individua u ontologiji član *owl:Thing* klase te je svaka klasa koja se definira podklasa

owl:Thing klase (uključujući i *rdfs:Resource* klasu). Element *owl:Class* služi za definiranje klase u OWL 1 ontologiji te je podklasa klase *rdfs:Class*. Prethodno navedeno se može prikazati na sljedeći način, sljedeći princip od „najmanjeg“ prema „najvećemu“ (tj. od podklase prema nadklasi):

owl:Class->*rdfs:Class*->*rdfs:Resource*->*owl:Thing*

Definiranje prazne klase je omogućeno s *owl:Nothing*. Nove korijenske klase se mogu definirati jednostavnim deklariranjem imena klase, npr. <*owl:Class* *rdf:ID*="Autor">.

Pomoću *rdf:ID* atributa se klasi dalo ime, a nakon toga se toj klasi mogu dodavati svojstva. Način pozivanja te klase je pomoću *rdf:about* atributa jer on omogućava pozivanje definicije klase koja je već negdje definirana te joj se dodaju neka svojstva bez mijenjanja glavnog dokumenta u kojemu je ta klasa definirana.

Element *rdfs:subClassOf* povezuje neku klasu sa nekom općenitijom klasom koja je iznad nje, tj. ako je *X rdfs:subClassOf Y*, onda je i svaka instanca klase X također i instanca klase Y (budući da je Y iznad X). Poveznica *rdfs:subClassOf* je prijelazna i na druge klase. Ako je A podklasa od B, a B je podklasa od C onda je i A podklasa od C. Definiranje klase se sastoji od davanja imena klasi te dodavanja restrikcija toj klasi (u ovom slučaju jedina restrikcija je dana pomoću *subClassOf* elementa, ali može se dodavati još restrikcija).

2.4.3.1.2 Ograničenja svojstava

Osim određivanja karakteristike svojstava, moguće je dodatno ograničiti doseg (engl. range) svojstava u određenim kontekstima na različite načine. U OWL-u se to postiže s ograničenjima svojstava (engl. property restrictions). Različiti oblici ograničenja svojstava opisani u nastavku se mogu koristiti samo u kontekstu *owl:Restriction*, tj. OWL 1 izraza pomoću kojega se može opisati neka anonimna klasa, a ta se klasa definira tako što se dodaju neke restrikcije na određeno svojstvo. Sve instance ove klase moraju biti pod tom restrikcijom. Restrikcija se sastoji od dva dijela, u prvom dijelu se govori o svojstvu na koje se restrikcija odnosi, što je učinjeno pomoću *owl:onProperty* svojstva koji označava ograničeno svojstvo. Drugi dio se odnosi na samu restrikciju, drugim riječima na ono na što se restrikcija odnosi. *owl:Restriction* je korišten kako bi se definirala anonimna klasa.

U OWL su dozvoljene dvije vrste ograničenja za svojstva. *Value constraints* tj. ograničenja na vrijednosti svojstva koja stavljaju ograničenja na doseg svojstva, te *cardinality*

constraint pomoću kojih se stavljaju ograničenja na broj vrijednosti koje neko svojstvo može poprimiti.

Vrijednosna ograničenja (engl. Value constraints) su *owl:allValuesFrom*, *owl:someValuesFrom* i *owl:hasValue*. Svojstva *allValuesFrom* and *someValuesFrom* su lokalna na njihove sadržane klasne definicije.

Svojstvo *owl:allValuesFrom* iskazuje kako vrijednost svojstva koje se ograničava mora biti iz točno određene klase ili mora imati točno određeni doseg. Primjerice, ograničenje koje se nametnulo ovom klasom se odnosi na svojstvo „napisao“ (što se izriče koristeći *owl:onProperty* oznaku), a sve vrijednosti svojstva „napisao“ moraju biti instance klase KnjiževnoDjelo (što je rečeno koristeći *owl:allValuesFrom* oznaku). Dakle, ovo ograničenje zahtijeva za svaku instancu klase da ima instance određenog svojstva, vrijednost svojstva su svi članovi klase naznačeni s *owl:allValuesFrom* klauzulom.

Ukoliko se želi kazati kako samo neke od vrijednosti svojstva moraju biti iz klase koju se ograničava, onda se koristi *owl:someValuesFrom* ograničenje te tada barem jedna od vrijednosti ovoga svojstva mora biti instanca klase. Razlika između ove dvije formulacije je razlika između egzistencijalnog i univerzalnog kvantifikatora.

Ukoliko se želi naznačiti kako samo jedna vrijednost svojstva mora biti iz klase koja se ograničava, onda se koristi *owl:hasValue* ograničenje. Ovo ograničenje omogućuje određivanje klasa na temelju postojanja određene vrijednosti svojstva. Individua će biti član takve klase svaki put kada barem jedna vrijednost njezinog svojstva je jednaka od *hasValue* *resursu*.

Dakle, pomoću *owl:allValuesFrom* se naznačava kako sve vrijednosti svojstva moraju biti iz određene klase, a može biti i da ovo svojstvo uopće nema vrijednost. Ograničenjem *owl:someValuesFrom* svojstvo mora imati barem jednu vrijednost iz određene klase, ali može imati i vrijednost iz neke druge klase. Pomoću *owl:hasValue* se kazalo da određeno svojstvo ima barem jednu određenu vrstu vrijednosti.

Cardinality constraints su *owl:maxCardinality*, *owl:minCardinality* i *owl:cardinality*. U OWL-u se pretpostavlja se kako bilo koja instanca klase može imati proizvoljan broj (nula ili više) vrijednosti za određeno svojstvo. Kako bi se to svojstvo u nekoj klasi učinilo obveznim, ili kako bi se broj vrijednosti nekoga svojstva ograničio, ili pak kako bi se reklo da svojstvo ne može imati nikakve vrijednosti u sebi, koriste se *cardinality constraints* (ograničenja bitnosti nekoga svojstva). Pomoću ovih ograničenja ograničava se bitnost svojstva na temelju klasa koje to svojstvo koriste.

Svojstvo *owl:min:Cardinality* se koristi kada se želi opisati klasu svih individua koje imaju najmanje X semantički različitih vrijednosti za neko svojstvo (gdje je X broj ograničenja). Primjerice, želi se kazati kako određena knjiga ima najmanje jednu osobu koja ju je napisala. Svojstvo *owl:maxCardinality* se koristi kada se želi npr. iskazati neki maksimalan broj osoba koje su napisale neku knjigu.

Kombiniranjem *minCardinality* i *maxCardinality* svojstava može se reći kako neka klasa može imati točno jednu vrijednost zadanoga svojstva, npr. da klasa Knjiga mora imati točno 1 vrijednost svojstva „*napisao*“ tj. neku knjigu je napisala točno 1 osoba. Tada se u *owl:minCardinality* i *owl:maxCardinality* ograničenje stavlja broj , a to se može napisati i kraće pomoću *owl:cardinality* svojstva. Važno je napomenuti kako svojstvo „*napisao*“ može imati i veći broj vrijednosti nego je to rečeno sa *owl:cardinality* izrazom (tj. više od 1 vrijednosti), ali onda te vrijednosti moraju biti jednake tako da je točan broj različitih vrijednosti jednak 1.

2.4.3.1.3 Složene klase

OWL 1 daje mogućnost i korištenja skupa operatora pri stvaranju klasa. Ti operatori daju mogućnost izražavanja Boolove logike, tj. predstavljaju AND, OR i NOT operatore, odnosno presjek (*intersectionOf*), uniju (*unionOf*) i komplement (*complementOf*).

Svojstvo *owl:intersectionOf* opisuje presjek dviju ili više klasa koji uključuje sve individue koje su zajedničke svim klasama u presjeku.

Svojstvo *owl:unionOf* povezuje neku klasu sa listom klasa koje ju opisuju, *owl:unionOf* izjava opisuje klasu koja se sastoji od dvije ili više klasa koje su članovi te klase koja se opisuje.

Svojstvo *owl:complementOf* opisuje klasu čije individue ne pripadaju nekoj drugoj klasi. Prema tome, kada se iskaže kako neka individua ne pripada toj klasi, onda se podrazumijeva da ona pripada svim drugim klasama koje nisu iste kao i ta klasa. Kada se ovo svojstvo koristi sa *intersectionOf* (presjekom) može biti korisno, npr. želi se kazati kako je „Amater“ Autor koji nije ProfesionalniAutor.

Ukoliko se koriste prethodno navedeni operatori, aplikacija će moći izvući sljedeće zaključke:

- Ako je klasa A presjek (*owl:intersectionOf*) klasa B, C i D, onda je klasa A podklasa svim trima klasama B, C i D.

- Ako je klasa A presjek nekih klasa, a klasa B presjek nekih drugih klasa, klasa A će biti podklasa klase B ako su svi sastavni dijelovi klase A podklase jednog sastavnog dijela klase B.
- Ako je klasa A unija (*owl:unionOf*) klasa B, C i D, tada su sve tri klase B, C i D podklase klase A.
- Ako je klasa A unija nekih klasa, a klasa B unija nekih drugih klasa, klasa B će biti podklasa klase A ako su svi sastavni dijelovi klase A nadklase jednom sastavnom dijelu klase B.
- Ako je klasa A komplement (*owl:complementOf*) klase B, onda će sve podklase klase A biti različite od klase B.

OWL 1 pri definiranju klasa pruža mogućnost nabiranja klasa, izražavanja jednakosti i nejednakosti klasa pomoću izraza *owl:oneOf*, *owl:equivalentClass* i *owl:disjointWith* respektivno.

Definiranje klasa nabiranjem se postiže pomoću *owl:oneOf* svojstva. Vrijednost ovoga svojstva mora biti skup instanci klase koju želimo definirati, što omogućava da neka klasa bude definirana nabiranjem instanci od kojih se sastoji. Klasa definirana sa *owl:oneOf* svojstvom sadrži točno onoliko individua koliko je nabrojano, te se ne mogu dodavati dodatne instance toj klasi pomoću neke aplikacije ili u nekoj drugoj ontologiji. Skup individua te klase se obično predstavlja pomoću *rdf:parseType="Collection"* izraza, koji je zgodna kratica za pisanje skupa elemenata u nekoj listi. Npr. ako postoji klasa TemaKnjiževnogDjela koja opisuje koja tema djela može biti, i želi se kazati kako su moguće teme: prijateljstvo, ljubav, socijalna osviještenost i sl. koristiti će se *owl:oneOf* element. Nabiranje je osobito korisno kada postoji velik broj instanci neke klase koje treba nabrojati.

Nejednakosti klasa se definiraju pomoću *owl:disjointWith*, a izražava kako dvije klase nisu jednake, tj. neki član jedne klase ne može biti član neke druge klase koja je različita od prve. Svojstvo *owl:disjointWith* je simetrično, što znači kada se kaže da je <klasa1 *owl:disjointWith* klasa2>, to znači da je klasa1 različita od klasa2, ali isto znači i da je klasa2 različita od klase1.

Jednakost klasa se definiraju pomoću *owl:equivalentClass* definira, a izražava kako su dvije klase jednake, tj. svaka individua koja je član jedne klase je također član i druge klase koja je jednaka prvoj. Drugim riječima, dvije klase imaju potpune iste instance. Ovo svojstvo se također koristi kada se želi iskazati kako su dvije klase u dvije različite ontologije jednake.

Tako će primjerice, kada se u nekom sustavu postavimo upit odgovor će uključiti i klase obiju ontologija u jednome upitu.

Na temelju ovih svojstava aplikacija će moći zaključiti sljedeće:

- Ako klasa A ima svojstvo *owl:oneOf* pomoću kojega su nabrojane instance te klase B, C i D onda će svaka od tih instanci imati *rdf:type* koji je dan klasi A
- Ako je klasa A *owl:equivalentClass* klasi B onda će između tih dviju klasa biti ubačena veza *<A owl:sameAs B>*
- Ako je klasa A *owl:disjointWith* s klasom B, onda će i bilo koja podklasa klase A biti *owl:disjointWith* s klasom B.

2.4.3.1.4 Svojstva OWL 1 jezika

Osim klasa, OWL nudi i neke konstrukte za svojstva. Aplikacije većinu zaključivanja temelje na „razumijevanju“ svojstava koja se koriste u RDF dokumentima. Zbog toga je bitno dati veći izbor pri obilježavanju svojstava nego je to dano u RDFS-i koja većinom koristi tri izraza za definiranje svojstava: *rdfs:domain*, *rdfs:range* i *rdfs:subPropertyOf*. Nadalje u RDFS-i se definiraju svojstva i onda se koriste kako bi se pomoću njega povezale dvije stvari, s tim da to svojstvo može povezivati jedan resurs s drugim ili može povezivati resurs s klasificiranom (string, integer i sl.) ili neklasificiranom (literalnom) vrijednošću. Oba slučaja koriste svojstvo *rdf:Property*. Međutim, u OWL 1 jeziku se razlikuju dvije glavne skupine svojstava koja se žele definirati, a to su *Objektna svojstva* (engl. *ObjectProperty*) koja se definiraju sa *owl:ObjectProperty* kojim se povezuje jedan resurs sa drugim resursom i *Podatkovna svojstva* (engl. *DatatypeProperty*) koja se definiraju s *owl:DatatypeProperty* kojim se povezuje resurs sa *rdfs:Literal* (literalnom) vrijednošću ili jednom od klasificiranih vrijednosti XML Scheme (string, integer i slično).

owl:ObjectProperty i *owl:DatatypeProperty* su podklase RDF klase *rdf:Property*.

OWL podržava sljedeće konstrukte za aksiome:

- RDF Schema konstrukti: *rdfs:subPropertyOf*, *rdfs:domain* i *rdfs:range*
- povezivanje s drugim svojstvima: *owl:equivalentProperty* i *owl:inverseOf*
- globalne *cardinality constraints*-e: *owl:FunctionalProperty* i *owl:InverseFunctionalProperty*
- logičke karakteristike svojstava: *owl:SymmetricProperty* i *owl:TransitiveProperty*

owl:DatatypeProperty se koriste kada se želi definirati neko *podatkovno* svojstvo, npr. svojstvo „brojKnjizevnihDjelaAutora“ i želi se da ono bude pozitivan broj. U tom slučaju se koristi *datatype xsd:nonNegativeInteger*. URI *reference* za sva *datatype* svojstva je *xsd:http://www.w3.org/2001/XMLSchema* budući su ta svojstva definirana u XML Schemi. Podatkovna svojstva (engl. Datatype property) su svojstva definirana u XML Schemi, a preporučena za korištenje u OWL su dana u Tablici 4.

Tabela 4: Podatkovna svojstva XML Scheme

xsd:string	xsd:normalizedString	xsd:boolean	
xsd:decimal	xsd:float	xsd:double	
xsd:integer	xsd:nonNegativeInteger	xsd:positiveInteger	
xsd:nonPositiveInteger	xsd:negativeInteger		
xsd:long	xsd:int	xsd:short	xsd:byte
xsd:unsignedLong	xsd:unsignedInt	xsd:unsignedShort	xsd:unsignedByte
xsd:hexBinary	xsd:base64Binary		
xsd:dateTime	xsd:time	xsd:date	xsd:gYearMonth
xsd:gYear	xsd:gMonthDay	xsd:gDay	xsd:gMonth
xsd:anyURI	xsd:token	xsd:language	
xsd:NMTOKEN	xsd:Name	xsd:NCName	

Svojstva u OWL-u mogu biti simetrična, prijenosna (tranzitivna), funkcionalna, obrnuto funkcionalna i suprotna.

Svojstvo je simetrično kada veza između subjekta i objekta vrijedi u oba smjera. Ako *Resurs1* predstavlja subjekt u izjavi te je povezan s objektom izjave *Resurs2* sa simetričnim svojstvom *S*, onda je i *Resurs2* (koji je u ovoj izjavi subjekt) povezan svojstvom *S* sa *Resurs1* (koji je u ovoj izjavi objekt). *owl:SymetricProperty* je podklasa od *owl:ObjectProperty*, pa prema tome *rdfs:range* simetričnog svojstva može biti samo resurs, a ne *literal* ili *datatype*.

Svojstvo je tranzitivno ako postoji *Resurs1* povezan s *Resursom2* pomoću svojstva *T*, a *Resurs2* je povezan s *Resurs3* sa istim svojstvom, onda je i *Resurs1* povezan sa *Resurs3* s tim svojstvom *T*. Primjer tranzitivnog (prijenosnog) svojstva je „veće od“ u slučaju ako je A veće od B, a B je veće od C, onda je i A veće od C. *owl:TransitiveProperty* je podklasa od *owl:ObjectProperty*, pa prema tome *rdfs:range* prijenosnog svojstva može biti samo resurs, a ne *literal* ili *datatype*, kao i kod simetričnog svojstva.

Funkcionalno svojstvo omogućava izraziti kako za neku instancu postoji najviše jedna vrijednost toga svojstva. Za dvije izjave koje su povezane funkcionalnim svojstvom *F* može se reći $\langle x F y \rangle$ i $\langle x F z \rangle$ vrijedi da je $y=z$, čime se izrazilo da može postojati više instanci *rdfs:domain-a*, ali za sve njih postoji najviše jedna *rdfs:range* vrijednost. Funkcionalno svojstvo radi istu stvar kao i *owl:cardinality*. Ako se želi kazati kako neka klasa može imati

točno jednu vrijednost zadanoga svojstva, npr. da klasa KnjiževnoDjelo mora imati točno 1 vrijednost svojstva „*napisao*“ tj. da je neko djelo napisala točno 1 osoba, onda se može koristiti *owl:FunctionalProperty*, i dobit će se isto značenje kao i korištenjem *owl:cardinality* svojstva. *owl:FunctionalProperty* je podklasa od *rdf:Property*, te prema tome *rdfs:range* funkcionalnog svojstva može biti resurs, *literal* ili *datatype*, za razliku od simetričnog i tranzitivnog svojstva.

Obrnuto funkcionalno svojstvo je suprotno od funkcionalnog svojstva. Dok je u funkcionalnom svojstvu za svaki *rdfs:domain* bio jedan *rdfs:range*, u obrnuto funkcionalnom, postoji jedan *rdfs:domain* za svaki *rdfs:range*. Dakle ako postoje izjave $\langle y \ F \ x \rangle$ i $\langle z \ F \ x \rangle$, onda je $y=z$. *owl:InverseFunctionalProperty* je podklasa od *rdf:Property*, stoga *rdfs:range* obrnuto funkcionalnog svojstva može biti resurs, *literal* ili *datatype*, isto kao i kod funkcionalnog svojstva. Također je moguće definirati neko svojstvo kao funkcionalno i obrnuto funkcionalno u isto vrijeme. Većina ID svojstava je ujedno i funkcionalna i obrnuto funkcionalna. Ako se želi objasniti razliku između obrnuto funkcionalnog i funkcionalnog svojstva dobra analogija je *e-mail* adresa. *E-mail* je obrnuto funkcionalno svojstvo budući da jedan *e-mail* pripada samo jednoj osobi, tj. obično je jedan *e-mail* u vlasništvu jedne osobe. Međutim, *e-mail* nije funkcionalno svojstvo budući da jedna osoba može posjedovati više *e-mail* adresa. Isto tako, godina rođenja je funkcionalno svojstvo, budući da svaka osoba je rođena na određeni dan ima i jedinstven dan rođenja. Međutim, to nije obrnuto proporcionalno svojstvo, budući da više osoba može imati isti dan rođenja. Da bi godina rođenja bila obrnuto funkcionalno svojstvo onda bi samo jedna osoba u svijetu mogla biti rođena na jedan datum u godini, što nije slučaj.⁸⁷

Svojstvo je inverzno ako postoji *Resurs1* koji je povezan sa *Resursom2* pomoću inverznog svojstva *I*, onda će to svojstvo povezati *Resurs2* sa *Resursom1*. Primjerice, ako postoji svojstvo *napisao* kojim želimo reći da je *Autor napisao* neko *KnjiževnoDjelo*. Međutim, želi se reći i da je *KnjiževnoDjelo napisanoOd* strane nekoga *Autora*. Budući je pretpostavka kako su „*napisao*“ i „*napisanoOd*“ suprotna svojstva, tu se može koristiti *owl:inverseOf*, npr. u svojstvu „*napisao*“ vrijednosti *rdfs:domain*-a i *rdfs:range*-a su suprotne od vrijednosti *rdfs:domain*-a i *rdfs:range*-a u svojstvu „*napisanoOd*“. *owl:inverseOf* je svojstvo, a ne klasa. To znači da se ono ne može koristiti da se neki *rdfs:domain* poveže sa *rdfs:range*-om, nego služi da se zada ograničenje pri definiranju nekoga drugog svojstva (npr. „*napisao*“ i „*napisanoOd*“).

⁸⁷ Liyang, Yu, op. cit. (bilj. 55), str. 122.

2.4.3.2 OWL2 jezik

OWL 2 jezik daje naprednije mogućnosti izražavanja od OWL 1 te će ih ovaj odjeljak prikazati, kategorizirajući ih u kategorije navedene u nastavku.⁸⁸

Učinkovitija sintaksa omogućuje lakše stvaranje nekih uobičajenih izjava. Ovo obilježje ne omogućuje neke nove mogućnosti zaključivanja, nego samo nudi olakšan način korištenja jezika dajući neke nove izraze koji omogućavaju skraćeno i učinkovito pisanje izjava naspram OWL 1 jezika. Slijedi njihovo kratko pojašnjenje.

U OWL 1 jeziku će se upotrebom *owl:disjointWith* izraz ukazati na nejednakost dviju klasa. Međutim, ukoliko se želi reći da je više od dvije klase različito treba se koristiti veći broj *owl:disjoint* izraza. OWL 2 pruža skraćenicu za ovaj slučaj upotrebom *owl:AllDisjointClasses* konstrukta, što je mnogo kraće i preglednije. Još jedno svojstvo slično ovome jest *owl:disjointUnionOf* koje se koristi kada se želi kazati kako se neka klasa sastoji od više podklasa koje su međusobno različite.

OWL 2 nudi i mogućnost iznošenja negativnih tvrdnji u izjavama, koje su korisne jer pomoću njih se može izreći kako neka tvrdnja nije istinita. Budući su *owl:ObjectProperty* i *owl:DatatypeProperty* dva moguća tipa danog svojstva, OWL 2 nudi konstrukte za iznošenje dva negativna tipa tvrdnji, a to su *owl:NegativeObjectPropertyAssertion* i *owl:NegativeDataPropertyAssertion*. Ovim se može reći kako neke dvije individue nisu povezane nekim svojstvom, što OWL 1 nije mogao. Slično tomu, pomoću *owl:NegativeDataPropertyAssertion* svojstva može se reći da jedan resurs nema određenu vrijednost za dano svojstvo.

Novi izrazi povećavaju izražajnost svojstava OWL 2 jezika te uključuju nova svojstva i nova ograničenja pomoću kojih to postižemo. Nova svojstva u OWL 2 jeziku su navedena u nastavku. Izraz *owl:hasSelf* se koristi kada se želi reći kako je neka klasa povezana sama sa sobom pomoću danog svojstva. Konstrukt *owl:hasSelf* ima tip *rdf:Property*, a njegov *rdfs:range* može biti bilo koji resurs. Ukoliko se *owl:hasSelf* restrikcijom neki izraz definira kao klasu, onda se dobiva klasu svih objekata koji su povezani sami sa sobom pomoću nekog svojstva. OWL 2 restrikcija na kardinalnost svojstava (eng. *cardinality restraints*) konstruktom *owl:minQualifiedCardinality* (tzv. „najmanje“ ograničenje) izražava da klasa Književnik može imati najmanje jedno svojstvo „*napisao*“. Konstrukt *owl:maxQualifiedCardinality* (tzv. „najviše“ ograničenje) izražava da klasa Književnik može napisati najviše jedno djelo, a

⁸⁸ Ibid.

owl:qualifiedCardinality (tzv. „točno“ određeno ograničenje) izražava da klasa *Književnik* može napisati točno jedno *KnjiževnoDjelo*.

Kako bi se definirala refleksivna, nerefleksivna i asimetrična svojstva OWL 2 jezik koristi *owl:ReflexiveProperty*, *owl:IrreflexiveProperty* i *owl:AsymmetricProperty*. Ova svojstva su podklasa od *owl:ObjectProperty* te njihov *rdfs:range* može biti samo resurs, a ne *literal* ili tip podataka. Nerefleksivno svojstvo govori da se neki resurs ne može odnositi sam na sebe, tj. suprotno je od refleksivnoga svojstva. Svojstvo je asimetrično ako povezuje resurs A sa resursom B, ali ne povezuje resurs B sa resursom A. Dakle, asimetrično svojstvo je suprotno od simetričnog (koje povezuje A sa B ali i B sa A). Npr. ako postoji svojstvo „napisao“ i želi se izreći da je <Autor napisao KnjiževnoDjelo>, ali također se želi ograničiti to svojstvo tako da se ne možemo reći da je <KnjiževnoDjelo napisao Autor>, to svojstvo će se definirati kao asimetrično.

Prethodno navedeni konstrukti su važni iz sljedećeg razloga. Ako neka izjava nije dana u RDF dokumentu to ne znači da ta izjava nije točna, nego jednostavno da nije napisana, što ostavlja velik prostor za nepoznanice te aplikacija nije učinkovita kao što bi mogla biti ako bi se uklonile te nepoznanice. Na postavljene upite će se moći odgovoriti sa većom sigurnošću što pridonosi većem povjerenju u izjave koje aplikacija može prikupiti.

OWL 2 jezik nudi sljedeće izraze za izražavanje nejednakosti svojstava. Ukoliko se želi reći da su dva svojstva međusobno različita koristi se *owl:propertyDisjointWith*, domena i doseg ovoga svojstva su tipa *rdf:Property*. Budući su *owl:ObjectProperty* i *owl:DatatypeProperty* podklase *rdf:Property*, *owl:propertyDisjointWith* se može koristiti da se odredi nejednakost i tipovi podataka (eng. *datatype*) i objektnih (eng. *object*) svojstava. Ukoliko se želi reći kako je grupa svojstava međusobno različita koristi se *owl:AllDisjointProperties*.

OWL 2 omogućuje definiranje svojstva koje se sastoji od lanca nekih drugih svojstava koja povezuju resurse. Lanac nekih drugih svojstava mogu biti samo objektna svojstva, a ne i tipovi podataka (*string*, *integer* i sl.). *OWL2 Reference*⁸⁹ nudi sljedeći primjer pri objašnjavanju ovoga svojstva. Definira se poveznica „imaStrica“ pomoću koje se želi reći da neka osoba ima strica. Ta poveznica se sastoji od lanca koji sadrži svojstva „imaOca“ pomoću kojeg se želi reći da neka osoba ima oca i „imaBrata“ pomoću kojega se želi reći da taj otac ima brata:

⁸⁹ W3C. OWL 2 Web Ontology Language: Quick Reference Guide (Second Edition). 11 December 2012. Dostupno na URL: <http://www.w3.org/TR/owl2-quick-reference/> (26.10.2015.).

```

<owl:ObjectProperty rdf:about="http://www.primjer.com/ontologija#imaOca">
</owl:ObjectProperty>
<owl:ObjectProperty
rdf:about="http://www.primjer.com/ontologija#imaBrata">
</owl:ObjectProperty>
<rdf:Description rdf:about="http://www.primjer.com/ontologija#imaStrica">
<owl:propertyChainAxiom rdf:parseType="Collection">
<owl:ObjectProperty
rdf:about="http://www.primjer.com/ontologija#imaOca"/>
<owl:ObjectProperty
rdf:about="http://www.primjer.com/ontologija#imaBrata"/>
</owl:propertyChainAxiom>
</rdf:Description>

```

Sada, ako postoje izjave

```

<http://www.primjer.com/osoba#Ivan> primjer:imaOca
<http://www.nesto.com/resurs#Pero>.
<http://www.nesto.com/resurs#Pero> primjer:imaBrata
<http://www.ime.com/resurs#Hrvoje>.

```

aplikacija će moći zaključiti da

```

<http://www.primjer.com/osoba#Ivan> primjer:imaStrica
<http://www.ime.com/resurs#Hrvoje>.

```

Ovo je korisno jer ne treba definirati da neka osoba ima strica, nego aplikacija to sama zaključuje i izvlači izjavu.

OWL 2 omogućuje definiranje tzv. ključeva za neku klasu pomoću *owl:hasKeys* konstrukta, koji se koristi kada se želi reći da je svaka instanca neke klase identificirana određenim svojstvom ili skupom svojstava. Npr. za klasu Autor može se imati ključ „*autorID*“ (ovdje se koristi *rdf:parseType* atribut kako bi se naglasila mogućnost ubacivanja i više od jednoga svojstva kao ključa kod *owl:hasKey* konstrukta). Pomoću ovoga ključa se može identificirati bilo koja instanca klase *Autor*, tj. može se identificirati bilo kojega autora pomoću njegovog *autorID* broja. Ukoliko postoje bilo koje dvije instance klase koje imaju iste vrijednosti ovoga ključa (ili više njih), iako mogu imati različite URI-je, onda će te dvije individue biti iste. Ovo je posebno korisno ako su dvije instance neke klase dobivene iz dva različita dokumenta na Web-u. Primjerice, ako je neki autor ima istu vrijednost ključa *autorID* u dva različita dokumenta na Webu, onda aplikacija može zaključiti da je to zapravo isti *Autor*.

OWL 2 pruža povećanu podršku za tipove podataka od OWL 1 jezika, nudeći veći broj preporučenih tipova podataka, mogućnost dodavanja ograničenja na tipove podataka i mogućnost definiranja vlastitih tipova podataka. Tipovi podataka ugrađeni u OWL2 su prikazani u Tablici 5.

Tabela 5: Podržani tipovi podataka u OWL 2 jeziku

Vrsta	Tip podataka
Cijeli brojevi i decimalni brojevi	xsd:decimal, xsd:integer, xsd:nonNegativeInteger, xsd:nonPositiveInteger, xsd:positiveInteger, xsd:negativeInteger, xsd:long, xsd:int, xsd:short, xsd:byte, xsd:unsignedLong, xsd:unsignedInt, xsd:unsignedShort, xsd:unsignedByte
Brojevi s pokretnim zarezom (engl. float-point numbers)	xsd:double, xsd:float
Niz podataka (engl. string)	xsd:string, xsd:normalizedString, xsd:token, xsd:language, xsd:Name, xsd:NCName, xsd:NMTOKEN
Boolove vrijednosti	xsd:boolean
Binarni podaci	xsd:hexBinary, xsd:base64Binary
IRI-ji	xsd:anyURI
Instance vremena	xsd:dateTime, xsd:dateTimeStamp
XML Literal	rdf:XMLLiteral
Novi ugrađeni tipovi u OWL 2 jezik	
Tip podataka	Objasnjenje
owl:real	Svi realni brojevi
owl:rational	Svi racionalni brojevi, podskup owl:real-nih brojeva, sadrži i vrijednosti xsd:decimal

xsd:minInclusive i *xsd:maxInclusive* određuju granične vrijednosti koje važe za određeni doseg (eng. *range*), a vrijednosti koje su izvan toga dosega su određene pomoću *xsd:minExclusive* i *xsd:maxExclusive* konstrukta. Ovi se konstrukti mogu koristiti samo na numeričkim tipovima podataka, dok drugi tipovi podataka imaju svoje vlastite konstrukte.

OWL 2 ima mogućnost određivanja nekih ograničenja na tipove podataka, *owl:onDatatype* i *owl:withRestrictions* su glavni konstrukti za nametanje ovih ograničenja te se pomoću njih mogu definirati novi tipovi podataka. Primjerice, može se kreirati novi tip podataka „*dobKnjiževnika*“, koji će imati donju granicu 18 godina i gornju granicu 65 godina ukoliko se želi da *dobKnjiževnika* bude između te dvije granice. Stvoreni podatak *dobKnjiževnika* se može koristiti kao *rdfs:range* vrijednost za neko svojstvo ili sa bilo kojim drugim tipom podataka. OWL 2 pruža i mogućnost kombiniranja tipova podataka kako bi se dobio novi tip pomoću konstrukata *owl:datatypeComplementOf*, *owl:intersectionOf* i *owl:unionOf*.

Jednostavnije mogućnosti modeliranja metapodataka te proširene mogućnosti pisanja zabilješki (komentara, napomena) čime se omogućava dodavanje zabilješki aksiomima, dodavanje informacija o domeni i dosegu tim zabilješkama i dodavanje zabilješki o

zabilješkama. OWL 2 omogućava dodavanje anotacije samoj ontologiji, entitetima (klasama, svojstvima, individuama), anonimnim individuama, aksiomima ali i samim anotacijama. U svrhu dodavanja informacija aksiomima koriste se konstrukti *owl:Axiom*, *owl:annotatedSource*, *owl:annotatedProperty* i *owl:annotatedTarget*.

2.4.4 SKOS

Naziv SKOS (engl. Simple Knowledge Organization System) je izabran kako bi se naglasio cilj pružanja jednostavnog, ali moćnog okvira za izražavanje sustava za organizaciju znanja (engl. Knowledge Organization System, KOS) na strojno razumljiv način. Temeljni SKOS (engl. SKOS Core) osigurava model za izražavanje osnovne strukture i sadržaja konceptualnih shema kao što su tezaursi, klasifikacijske scheme, popisi predmetnih odrednica (engl. subject heading lists), taksonomije, folksonomije i ostale vrste kontroliranih rječnika te također konceptualne scheme ugrađene u glosarije i terminologiju. Konceptualna schema je ovdje definirana kao skup koncepata, izborno uključujući izjave o semantičkim odnosima između tih pojmova. SKOS temeljni rječnik (engl. SKOS Core Vocabulary) je skup RDF svojstava i RDFS klasa, koji se može koristiti za izražavanje sadržaja i strukture konceptualne sheme kao RDF grafa.

Upotreba SKOS temeljnog rječnika (engl. SKOS Core Vocabulary) omogućuje izražavanje konceptualne sheme kao RDF grafa te povezivanje i/ili spajanje podataka s ostalim RDF podacima semantičkih web aplikacija, što u praksi omogućuje dijeljenje izvora podataka preko weba na neki novi i neočekivani smisleno sastavljeni i integrirani način od strane aplikacija.⁹⁰

SKOS (engl. Simple Knowledge Organization System) je preporuka W3C-a te osigurava sredstvo za predstavljanje sustava za organizaciju znanja (uključujući kontrolirane rječnike, tezaurse, taksonomije i folksonomije) na distribuiran i povezan način. Sustavi za organizaciju znanja su uglavnom formalno već dugo vremena zastupljeni kao dio knjižnične znanosti, ali ne kao sredstvo za predstavljanje i razmjenjivanje u računalnim mrežama. S obzirom na postojanje nekoliko tezaurskih standarda, moglo bi se zapitati zašto je bilo potrebno stvoriti još jedan. Ključna razlika između postojećih tezaurskih standarda i SKOS-a

⁹⁰ W3C. SKOS Core Guide. Working Draft 2 November 2005. Dostupno na URL:
<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/> (26.10.2015.).

je njegova osnova u semantičkom web-u. Za razliku od postojećih standarda, SKOS je od početka osmišljen kako bi se omogućio stvaranje usklađene organizacije znanja koje se može ponovo koristiti i na koje se može upućivat preko weba. SKOS nije stvoren da zamijeni bilo koji tezaursni standard, nego ih naglašava dovodenjem distribuirane prirode semantičkog weba tezaursima i kontroliranim rječnicima. Cilj stvaranja SKOS-a je mogućnost pridruživanja (mapiranja, engl.mapping) bilo kojeg tezaursnog standarda na SKOS na prilično jednostavan način.

2.4.4.1 Konceptualna klasa (engl. Concept Class)

Konceptualna klasa *skos:Concept* omogućuje tvrdnju kako je neki resurs ustvari konceptualni resurs tj. sami resurs je koncept te se pri tome koristi *rdf:type* svojstvo za trojke prilikom opisa resursa kao konceptualnog resursa.

2.4.4.2 Svojstva za označavanje

Ovaj odjeljak opisuje SKOS temeljna svojstva za označavanje (*engl. Labelling Properties*) kao što su *skos:prefLabel*, *skos:altLabel*, *skos:hiddenLabel*, *skos:prefSymbol* i *skos:altSymbol*. Ovdje označavanje (*engl. labelling*) znači dodjeljivanje nekih vrsta izraza resursu, gdje je izraz namijenjen za obilježavanje (oznaka) resursa u prirodnom jeziku diskursa i / ili u prikazima namijenjenim ljudima.

Svojstva *skos:prefLabel* i *skos:altLabel* omogućuju dodjeljivanje preferiranih i alternativnih leksičkih oznaka resursima. Tako se opisno svojstvo *skos:altLabel* upotrijebljava za označavanje sinonima. Kada se označava resurs tipa *skos:Concept*, nije potrebno ograničiti preferiranu i alternativnu leksičku oznaku kako bi se precizirali sinonimi. Svojstva *skos:prefLabel* i *skos:altLabel* su pod-svojstva *rdfs:label*. Skraćenice i akronimi se mogu koristiti za označavanje koncepata te je izbor njihovog korištenja kao preferiranih ili alternativnih termine slobodan. Vrijednost svojstava *skos:prefLabel* i *skos:altLabel* treba biti čisti literal (*engl.plain literal*), tj. niz znakova s dodatnim tagom jezika, a tag jezika se može upotrijebiti kako bi se ograničio opseg leksičkoj oznaci na određeni jezik. Vrijednosti

dopuštene kao jezični tagovi su dane u [XMLLANG]⁹¹. Kada se koncept višejezično označava, treba imati na umu kako suptilne razlike u smislu prevedenih riječi ili fraza između jezika mogu rezultirati različitim interpretacijama pojmova od strane različitih jezičnih zajednica. Pojam ne bi trebao imati više od jedne preferirane leksičke oznake po jeziku. Preporučljivo je da dva koncepta u istoj konceptualnoj shemi nemaju istu preferiranu oznaku na bilo kojem jeziku. Međutim, pogrešno napisane riječi su obično uključene među skrivene leksičke oznake (engl. hidden lexical labels), koje predstavljaju leksičku oznaku za resurs kako bi niz znakova bio dostupan aplikacijama koje obavljaju tekstualno indeksiranje i postupke pretraživanja, ali se ne želi da oznake bude vidljiv na drukčiji način. Skrivena leksička oznaka se resursu pridružuje upotrebom *skos:hiddenLabel* svojstvom. Najčešća upotreba skrivenih oznaka uključuje pogrešno napisane varijante ostalih leksičkih oznaka. Svojstvo *skos:hiddenLabel* je pod-svojstvo od *rdfs:label*.

Simboličko označavanje (engl. Symbolic labelling) znači označavanje koncepta sa slikom. Za dodjeljivanje preferiranog i alternativnog simboličkog označavanja konceptu upotrebljavaju se *skos:prefSymbol* i *skos:altSymbol* svojstva. Preporučljivo je da dva koncepta u istoj konceptualnoj shemi nemaju istu preferiranu simboličku oznaku.

SKOS temeljna dokumentacijska svojstva (engl. Documentation Properties) su raspoređena kao hijerarhija svojstva, kao što je prikazano na Slici 5.

```
skos:note
|
|-- skos:definition
|
|-- skos:scopeNote
|
|-- skos:example
|
|-- skos:historyNote
|
|-- skos:editorialNote
|
|-- skos:changeNote
```

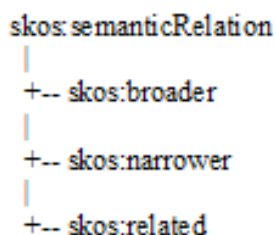
Slika 5: SKOS temeljna dokumentacijska svojstva (preuzeto iz: (SKOS Core Guide, 2005.))

Tako su *skos:definition*, *skos:scopeNote*, *skos:example*, *skos:historyNote*, *skos:editorialNote*, *skos:changeNote* je također *skos:note*. Svojstvo *skos:note* se koristi za

⁹¹ IETF (Internet Engineering Task Force). RFC 3066: Tags for the Identification of Languages / ed. H. Alvestrand. 2001. Dostupno na URL: <http://www.ietf.org/rfc/rfc3066.txt> (26.11.2015.).

opću dokumentaciju u bilo koju svrhu. Svojstvo *skos:definition* se koristi za potpuno objašnjenje namjeravanog značenja pojma. Svojstvo *skos:scopeNote* se koristi za neke, moguće djelomične informacije o namjeravanom značenju i/ili upotrebi koncepta, koji može biti izražen kao informacija o tome što je ili nije uključeno u značenje (opsegu) koncepta. Svojstvo *skos:example* donosi primjer korištenja koncepta. Svojstvo *skos:historyNote* daje upute za korisnike sheme, posebno se odnosi na značajne promjene objašnjenja pojma. Svojstvo *skos : editorialNote* donosi informacije koje pomažu u upravljanju uredničkim poslom, npr. uredničke napomene o mogućim promjenama koje se još mogu napraviti. Svojstvo *skos:changeNote* se koristi za dokumentiranje detaljnih promjena koncepta, za potrebe administracije i održavanja.

SKOS temeljni rječnik uključuje sljedeća svojstva za izjavljivanje semantičkih (paradigmatskih) odnosa između pojmova: *skos:semanticRelation*, *skos:broader*, *skos:narrower* i *skos:related*. Svojstva su hijerarhijski poredana, kao što je prikazano na Slici 6.



Slika 6: Hijerarhija semantičkih svojstava (preuzeto iz: (SKOS Core Guide, 2005.))

Domena i doseg svojstva *skos:semanticRelation* je klasa *skos:Concept*, stoga upotreba bilo kojeg SKOS temeljnog semantički odnosnog svojstva podrazumijeva da su subjekt i objekt trojke resursi vrste *skos:Concept*.

Ukoliko se želi tvrditi da je jedan pojam je šireg značenja (smisla, općenitiji) nego drugi, gdje opseg (značenja) jednog pojma u potpunosti pada u opseg značenja drugog pojma, koristi se *skos: broader* svojstvo. Inverzna tvrdnja, odnosno za izražavanje kako je jedan pojam užeg značenja (specifičnij) od drugoga koristi se *skos:narrower* svojstvo. Dakle, svojstva *skos: broader* i *skos:narrower* su međusobno inverzna svojstva te su oboje tranzitivna svojstva.

Za uspostavljanje asocijativnog odnosa (engl. Associative Relationship) između dva pojma koristi se *skos:related* svojstvo, koje je simetrično svojstvo.

2.4.4.3 Konceptne sheme (engl. Concept Schemes)

Uobičajeno se pojmovi definiraju u odnosu na druge pojmove, kao dio interno koherentne konceptne sheme. Kao što je spomenuto u uvodu, konceptualna shema je definirana kao skup koncepata, izborno uključujući izjave o semantičkim odnosima između tih pojmova. Klasa *skos:ConceptScheme* omogućuje tvrdnju kako je resurs konceptnih shema. Koncepti ne moraju biti dio konceptne sheme, oni mogu biti definirani i deklarirani kao samostalni resursi. Međutim, ukoliko se želi tvrditi da je koncept dio određene konceptne sheme, koristi se *skos:inScheme* svojstvo. Koncept može sudjelovati u bilo kojem broju konceptnih shema.

Za konceptne sheme gdje su pojmovi poredani kao nadređeni ili podređeni (odnosno širi ili uži) u generaliziranoj hijerarhiji, koristi se *skos:hasTopConcept* svojstvo kako bi se uspostavila poveznica između konceptne sheme i pojmova koji su pojmovi najviše razine u generaliziranoj hijerarhiji. Preporučljivo je koristiti *skos:hasTopConcept* svojstva jer to učinkovit način lociranja pojmova gornje razine za danu shemu. Kada se očituje konceptna shema, mogu se također koristiti svojstva iz drugih RDF rječnika kao što su DCMI Terms⁹² za opis meta-svojstava sheme. Subjektivno indeksiranje⁹³ (engl. Subject Indexing) informacijskih resursa na webu se ostvaruje svojstvima *skos:subject*, *skos:isSubjectOf*, *skos:primarySubject* i *skos:isPrimarySubjectOf*. Svojstvom *skos:subject* se uspostavlja tvrdnja da je koncept subjekt (tj. tema) informacijskog resursa. Može se upotrijebiti svojstvo *skos:isSubjectOf* kako bi se uspostavila tvrdnja u suprotnom smjeru, iz čega slijedi zaključak kako su *skos:subject* i *skos:isSubjectOf* inverzna svojstva. Svojstvo *skos:subject* je pod-svojstvo od *dc:subject*. Informacijski resurs može imati bilo koji broj predmeta. Kako bi se izrazilo da je jedan od ovih predmeta primarni predmet (primarni subjekt) dokumenta, upotrebljava se *skos:primarySubject* svojstvo, a njegovo inverzno svojstvo je *skos:isPrimarySubjectOf*.

⁹² Dublin Core Metadata Initiative (DCMI). DCMI Metadata Terms. [Dostupno na URL: http://dublincore.org/documents/dcmi-terms/](http://dublincore.org/documents/dcmi-terms/) (26.11.2015.)

⁹³ Indeksiranje je intelektualna analiza predmeta dokumenta kako bi se utvrdili pojmovi predstavljeni u njemu te dodijelili dogovarajući preferirani termini (engl. preferred term) u svrhu omogućavanja pretraživosti informacija.// Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval. Dostupno na URL: <http://www.willpowerinfo.co.uk/glossary.htm> (26.11.2015.)

Svojstvo *skos:primarySubject* je pod-svojstvo od *skos:subject*. Svojstvo *skos:isPrimarySubjectOf* je pod-svojstvo od *skos:isSubjectOf*. Resurs ne bi trebao imati više od jednog primarnog predmeta (subjekta) bilo koje konceptualne sheme.

Značajke SKOS-a dopuštaju definiranje smislenih grupacija ("zbirki") pojmova (engl. Meaningful Collections of Concepts), stoga SKOS rječnik ima poseban rječnik za obradu zbirke, koji je opisan u nastavku. Kako bi se definirale smislene konceptualne zbirke koristi se *skos:Collection* klasa i *skos:member* svojstvo. Za dodavanje leksičkih oznaka na zbirke (Označene zbirke, engl. Labelled Collections), upotrebljava se *rdfs:label* svojstvo. Zbirka se može definirati kao slijepi čvor (engl. blank node), tj. ne treba joj biti dodijeljen URI. Najčešća upotreba označene zbirke je hijerarhijski prikaz te se na taj način mogu opisati uži i širi odnosi između pojma i zbirke.

Klasa *skos:CollectableProperty* podržava opći mehanizam kojim se zbirke mogu uključiti u semantičke odnose. Svojstva *skos:narrower*, *skos:broader* i *skos:related* su očitovana kao sakupljajuća svojstva (engl. collectable properties).

Kako bi se definirala poredana zbirka pojmova (engl. Ordered Collections), upotrebljava se *skos:OrderedCollection* klasa sa *skos:memberList* svojstvom. Poredana zbirka može također imati oznaku (upotrebljava se *rdfs:label*). Mogu se koristiti sa semantičkim odnosnim svojstvima na isti način kao neporedane zbirke (engl. unordered collections). Klasa *skos:OrderedCollection* je pod-klasa od *skos:Collection*.

Zbirke mogu također biti ugnježdene (engl. Nested Collections) unutar druge zbirke. Poredane zbirke mogu također biti ugnježdene, a mogu biti ugnježdene i unutar neporedanih zbirki i obrnuto.

Objavljeni predmetni pokazatelji (engl. Published Subject Indicators) se omogućuje *skos:subjectIndicator* svojstvom, kojim se postavlja poveznica između koncepta i čovjeku čitljivog dokumenta koji osigurava potpuni, konačni opis tog pojma. Svojstvo *skos:subjectIndicator* je *owl:InverseFunctionalProperty*, a to znači ukoliko bilo koja dva čvora u RDF grafu imaju istu vrijednost za ovo svojstvo, oni su ustvari isti resurs. Stoga se može koristiti *skos:subjectIndicator* svojstvo za indirektno identificiranje pojma upućivanjem na URI dokumenta koji je objavljeni predmetni indikator za taj pojam.

Za resurs tipa *skos:Concept*, bilo koje svojstvo tog resursa (npr. creator, date of modification, source itd.) treba biti interpretirano kao *svojstva pojma*, a ne kao svojstva nekih „stvarnih stvari svijeta“ kojih taj resurs može biti konceptualizacija.

Ovaj neizravni SKOS sloj omogućuje izražavanje podataka tezaurusa kao RDF grafa. Pojmovni sadržaj bilo kojeg tezaurusa može također biti preoblikovan kao RDFS/OWL

ontologija. SKOS predavljanje tezaurusa se pridružuje prilično izravno na izvorne strukture podataka, te može biti stvoren bez skupih preoblikovanja i analiza. Temeljni SKOS je neovisan za osiguranje stabilnog kodiranja tezaurusnih podataka u grafičkom RDF formalizmu, kao i migracijskog puta za istraživanje dobiti prebacivanja iz tezaurusnih u RDFS/OWL oblikovanih formalizama.

Svojstvo *owl:sameAs* se ne bi trebalo koristiti za izražavanje činjenice kako dva konceptualna resursa (npr. resursi tipa *skos:Concept*) dijele isto značenje. Svojstvo *owl:sameAs* znači da su dva resursa identična u svakom pogledu (oni su ustvari isti resurs). Iako dva konceptualna resursa mogu imati isto značenje, oni mogu imati različite stvaratelje, oznake, dokumentaciju, povijest i budućnost. Međutim, bilo koji dio može postavljati pridružujući odnos između bilo koja dva konceptualna resursa, što se izražava SKOS pridružujućim rječnikom (engl. SKOS Mapping vocabulary , SKOSMAP⁹⁴) koji sadrži brojna RDF svojstva za izražavanje pridružujućih odnosa između konceptualnih resursa. SKOS pretpostavlja da na značenje konceptualnog resursa ne utječu bilo koje RDF ili prirodne jezične izjave o njemu jer stvaratelj koncepta određuje značenje pojma i izjave o konceptu pružajući informativni skup izjava o konceptu, kako bi drugi ljudi ispravno protumačili značenje i dijelili isto razumijevanje. U praksi, to znači da se pojam može koristiti u bilo kojem kontekstu i da će značenje pojma ostati nepromijenjeno. Međutim, treba imati na umu da je skup semantičkih odnosa u koje je uključen koncept važan dio objašnjenja značenja tog pojma, te će utjecati na interpretaciju od strane drugih ljudi. Gdje koncept ima malo dokumentacije, ljudi će se teško oslanjati na semantički kontekst koncepta kako bi oblikovali njegovo značenje. Stoga, širina objašnjenja značenje koncepta preko oznaka i dokumentacije će odrediti koliko je taj koncept koristan izvan prvotno predviđenog konteksta.

⁹⁴ W3C. SKOS Mapping Vocabulary Specification. Dostupno na URL: <http://www.w3.org/2004/02/skos/mapping/spec/2004-11-11.html> (26.11.2015.).

2.5 Zaključak

Poglavlje *Teorijske osnove semantičkog weba* bavi se teorijskim prikazom i pojašnjenjem semantičkog weba u svrhu njegovog potpunog razumijevanja. Cilj poglavlja je bio ukazati na osnovnu teorijsku i tehničku pozadinu semantičkog weba. Objasnjeno je značenje pojma semantički web, temeljna razlika između weba kakvoga danas poznajemo i njegovog razvoja prema semantičkom webu, nedostaci današnjeg weba i prednosti semantičkog weba, temeljni pojmovi i arhitektura semantičkog weba. Poglavlje je dalo pregled temeljnih ontologijskih definicija te je ukazalo na glavnu svrhu i ulogu ontologija na semantičkom webu. Detaljno su prikazane osnove RDF, RDFS, OWL i SKOS jezika primjenjivanih na semantičkom webu uz konkretne primjere. Cijelo poglavlje predstavlja osnovu za dobro razumijevanje primijene ontologija u informacijskim sustavima jer je prvo potrebno razumjeti teoretsku zamisao semantičkog weba na kojoj se temelji.

3 DOSADAŠNJI ONTOLOGIJSKI PROJEKTI

3.1 Uvod

Razvoj semantičkog weba i ontologije imao je veliki utjecaj na predstavljanja znanja, što je vrlo značajno jer se znanje može primijeniti za zaključivanje kao dio sustava za potporu odlučivanju.⁹⁵ Enciklopedistika nije iznimka te razvoj semantičkog weba može imati velika poboljšanja za ovo područje jer može pomoći u organizaciji znanja, dobivanju preciznih odgovora na postavljene upite te brzom i temeljitom razvoju velike količine informacija u ovom području.

Enciklopedijski rad je izložen velikom broju sveučilišta, škola i akademskih ustanova. Enciklopedije taksonomiziraju ljudsko znanje i pružaju vrijedno polazište znanja iz svih akademskih disciplina za znanstvenike i studente, pokrivajući njihovu tematsku građu sa širokog i detaljnog aspekta. Dok je eksplozija informacija učinila njihovo vjerno i kratko sažimanje još nedostižnijim, enciklopedije su možda nikada relevantnije. Stavljen je naglasak na razvoj i održavanje referentnih radova koji nude dostupnost i ažuriranje sažetaka. Ukratko, znanstvenici i studenti ne trebaju samo referentna djela, njima su potrebna i sredstva za učinkovito traženje i navigaciju. U svrhu očuvanja korisnosti enciklopedije u današnjoj informacijskoj eksploziji potrebno je poboljšati sposobnost predstavljanja njihovog sadržaja na smisleni (semantički) način. Dinamična priroda i povećanje razmjera digitalnih referentnih radova čine tradicionalne uredničke metode okupljanja i organiziranja izvora metasadržaja (indeksi, unakrsne uputnice, tablice sadržaja) toliko intenzivnim i neučinkovitim da su praktički neprimjenjivi na elektroničke enciklopedije u web okruženju. Korisniku koji proučava izvor može se ponuditi "vođeno razgledavanje" kroz sadržaj ostvarivanjem pojmovnog i semantičkog navigacijskog sučelja. Kako bi se riješili navedeni problemi na odgovarajući način, potrebne su sofisticiranije tehnike generiranja metasadržaja velikih i neistovremeno ažuriranih tekstualnih korpusa, što zahtijeva tehnološke eksperte za dizajniranje formalnog prikaza i eksperta domene za dohvaćanje semantičkog sadržaja korpusa.

⁹⁵ Jovic, Alan; Prcela, Marin; Gamberger, Dragan. Ontologies in Medical Knowledge Representation. // Proceedings of the *ITI 2007* 29th International Conference on Information Technology Interfaces, IEEE 2007, Cavtat, str. 535–540. Dostupno na URL: <https://bib.irb.hr/datoteka/301358.iti2007jovicprcelagamberger.pdf> (27.11.2015.).

Poglavlje će prikazati neke od važnijih projekata semantičkog weba vezane uz razvoj značajnih ontologija te će se prilikom njihovog izbora posebno voditi računa o onima koji su značajni za bolje razumijevanje semantičkog weba za područje enciklopedistike. Stoga se jedan dio ovog poglavlja bavi prikazom ontologija nezaobilaznih za bolje razumijevanje semantičkog weba, a drugi dio poglavlja daje osvrt na ontologijske projekte izrasle isključivo na enciklopedijskom znanju.

3.2 Ontologijski projekti

3.2.1 SUO

Prepoznajući potrebu za velikim ontologijama i otvorenim pristupom stvaranja slobodnog i javnog standarda javila se potreba za suradničkom izradom standardne ontologije gornje razine (engl. Standard Upper Ontology - SUO), koja će osigurati definicije za termine opće namjene i biti će temelj za specifičnije domenske ontologije. Potreba za standardnom ontologijom gornje razine je bila pokretačka snaga zbog koje je IEEE⁹⁶ organizirao Radnu grupu za ontologiju gornje razine (engl. Standard Upper Ontology Working Group, SUO WG⁹⁷) koja će raditi na njezinom razvoju pod projektom P1600.1⁹⁸. *SUO* ima različite svrhe, a neke od njih su:

- stvaranje novih baza znanja i podataka jer programeri mogu izrađivati novo znanje i definirati nove elemente podataka terminima opće ontologije te stoga postići određeni stupanj interoperabilnosti s drugim usklađenim sustavima;
- integracija baza znanja jer se elementi podataka iz postojećih sustava mogu pridružiti (mapirati) na opću ontologiju;
- integracija ontologija određenih domena, čime se takve ontologije (ako su u skladu sa SUO-om) biti interoperabilne (do određenog stupnja) na temelju zajedničkih pojmova i definicija.

⁹⁶ IEEE (engl. Institute of Electrical and Electronics Engineers) je međunarodna neprofitabilna organizacija zadužena za unapređivanje tehnologije. IEEE se odlikuju s visoko citiranom literaturom, konferencijama, tehnologijskim standardima te profesionalnim i edukacijskim aktivnostima. Osnovana je 1963. godine, a sjedište joj je u Piscataway, New Jersey. // IEEE – About. Dostupno na URL: <http://www.ieee.org/about/index.html> (3.06.2016.).

⁹⁷ SUO WG je organizirao IEEE za razvoj standard koji će specificirati ontologiju gornje razine koja će podržati interoperabilnost, pretraživanje i dohvaćanje informacija. // Standard Upper Ontology Working Group (SUO WG). Home Page. Dostupno na URL: <http://web.archive.org/web/20071010030010/http://suo.ieee.org/index.html> (3.06.2016.).

⁹⁸ IEEE P1600.1 Standard Upper Ontology (SUO) Working Group. Dostupno na URL: <http://suo.ieee.org/> (3.06.2016.).

Rezultat rada navedene Radne grupe za SUO ontologiju gornje razine je ontologija ograničena na metakoncepte, opće, apstraktne i filozofske pojmove koji su dovoljno općeniti za usmjeriti na široki opseg domenskih područja. Namjera je osigurati strukturu i skup općih pojmova pod kojim će se domenske ontologije (npr. medicina, financije, strojarstvo itd.) moći oblikovati. Radna grupa je započela rad SUO ontologije na temelju nekoliko ontologija gornje razine (npr. SUMO, CYC, ontologija gornje razine opisana od strane Russell i Norvig-a i John Sowa ontologije gornje razine) pokušavajući prilagoditi većinu njihovih pojmova i odnosa u zajedničku ontologiju. Procijenjeno je da SUO konačno sadrži između 1000 i 2500 termina te približno deset definicija za svaki termin.⁹⁹

3.2.2 SUMO

SUMO (engl. Suggested Upper Merged Ontology) je opsežan ontologijski projekt otvorenog pristupa¹⁰⁰ te je ontologija prvi put objavljena u prosincu 2000, uključujući ontologiju gornje razine. Izvorno je razvijena od Teknowledge Corporation, a trenutno je održava Articulate Software¹⁰¹. Adam Pease, tadašnji tehnički urednik standarda, autorizirao je izvjestan broj publikacija koje predstavljaju njezina ključna svojstva.¹⁰²¹⁰³ SUMO se može pregledavati,¹⁰⁴ preuzimati i besplatno koristiti. Kao što je slučaj sa Cyc¹⁰⁵-om, temeljna struktura njegove ontologije je teorija skupova, uključujući i aksiome i pravila. Međutim, za razliku od Cyc-a, ne uključuje program za zaključivanje. SUMO se izvorno bavi konceptima meta-razine tj. općenitim entitetima koji ne pripadaju određenoj problemskoj domeni.

⁹⁹ Gašević, Dragan; Djurić, Dragan; Devedžić, Vladan, op. cit. (bilj.37), str. 79.

¹⁰⁰ Pease, Adam; Niles, Ian; Li, John. The suggested upper merged ontology: A large ontology for the semantic web and its applications. Dostupno na URL: <http://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf> (27.11.2015.).

¹⁰¹ Articulate Software. Dostupno na URL: <http://www.articulatesoftware.com/> (27.11.2015.).

¹⁰² Niles, Ian; Pease, Adam. Towards a Standard Upper Ontology // Proceedings of the 2nd International Conference on Formal Ontology in Information Systems 2(2001), str. 211-277. Dostupno na URL: <http://projects.tecknowledge.com/HPKB/Publications/FOIS.pdf> (27.8.2015.)

¹⁰³ Pease, Adam; Niles, Ian. IEEE Standard Upper Ontology : a progress report. // Knowledge Engineering Review 17(2002), str. 65-70.

¹⁰⁴ Pogledaj: Suggested Upper Merged Ontology (SUMO). Dostupno na URL: www.ontologyportal.org (27.8.2015.).

¹⁰⁵ Cycorp. Dostupno na URL: <http://www.cyc.com/> (27.8.2015.).

Proširen je na ontologiju srednje razine (engl. Mid-Level Ontology, MILO)¹⁰⁶ i dvanaest domenskih ontologija uz potpunu podešenost skupova poveznica na WordNet.¹⁰⁷ Inače, WordNet je velika leksička baza podataka engleskog jezika, koja se započela razvijati 1986. godine na Princeton University. To je velika semantička mreža međusobno povezanih riječi i skupina riječi pomoću leksičkih i konceptualnih odnosa. WordNet-ovi gradivni blokovi su sinonimski skupovi podataka. Svaki član određenog sinonimskog skupa podataka izražava isti koncept, iako svi članovi određenog sinonimskog skupa nisu međusobno zamjenjivi u svim kontekstima.¹⁰⁸ MILO su autori pokušali prodati konstruktorima semantičkog weba,¹⁰⁹ no ova je ontologija samo manja verzija istog osnovnog oblikovanja kao i Cyc te je podložna istim problemima, stoga je jednako slabo prihvaćena.¹¹⁰ Dostupna formalna ontologija sadrži oko 25,000 termina i 80,000 aksioma kada se kombiniraju sve domenske ontologije.¹¹¹ Definirana su pridruživanja (mapiranja) na WordNet sinonimne grozdove te je objavljen velik broj sjedinjenja SUMO-a s projektima koji uključuje milijune Wikipedijinih činjenica tj. s DBpedia i YAGO sadržajima na razini instanci o osobama, kao i početna povezivanja Mondial geographical podataka.¹¹² Također su izvršena pridruživanja Open Biomedical Ontologies¹¹³ na SUMO.¹¹⁴

SUMO je nastao sjedinjenjem javno dostupnih ontološki sadržaj¹¹⁵ u sveobuhvatnu i jedinstvenu strukturu uključujući knjižnicu ontologija dostupnih na Ontolingua serveru¹¹⁶ i ontologija razvijenih na ITBM-CNR¹¹⁷ institutu, John Sowa¹¹⁸ ontologiju gornje razine,

¹⁰⁶ Niles, Ian; Allan, Terry. The MILO : A General-purpose, Mid-level Ontology. // Proceedings of the International Conference on Information and Knowledge Engineering, Las Vegas: CSREA Press, 2004. Str. 5-19.

¹⁰⁷ WordNet. Dostupno na URL: <https://wordnet.princeton.edu/> (3.06.2016).

¹⁰⁸ U daljnjem tekstu ovog poglavlja je detaljnije opisan WordNet projekt.

¹⁰⁹ Pease, Adam; Niles, Ian; Li, John, loc. cit.

¹¹⁰ Legg, Catherine. *Ontologies on the semantic web // Annual Review of Information Science and Technology* 41(2007), Chapter 9, 437.

¹¹¹ Vidjeti Suggested Upper Merged Ontology (SUMO), loc. cit.

¹¹² The Mondial Database. Dostupno na URL: <http://datahub.io/dataset/mondial> (27.8.2015.).

¹¹³ The OBO Foundry. Dostupno na URL: <http://www.obofoundry.org/> (27.8.2015.).

¹¹⁴ Suggested Upper Merged Ontology (SUMO), loc. cit.

¹¹⁵ Niles, Ian; Pease, Adam. Origins of the IEEE Standard Upper Ontology. // *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*. Seattle, 2001, str. 37- 42. Dostupno na URL: <http://www.adampease.org/OP/pubs/IJCAI2001.pdf> (27.8.2015.).

¹¹⁶ KSL Ontology Server Projects. Dostupno na URL:

<http://www.ksl.stanford.edu/software/ontolingua/ontology-server-projects.html> (27.8.2015.).

¹¹⁷ ITB CNR Institute for Biomedical Technologies. Dostupno na URL: <http://www.itb.cnr.it/> (27.8.2015.).

¹¹⁸ Sowa, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA, 2000. Dostupno na URL: <http://www.jfsowa.com/krbook/> (27.8.2015.).

različite mereotopološke teorije, Russell i Norvig¹¹⁹ ontologiju gornje razine, James Allen vremenske aksiome¹²⁰, formalna teorija cjelina Casatija i Varzija¹²¹, Smithova ontologija granica^{122,123}, Nicola Guarino formalna mereotopologija^{124,125}, različita formalna predstavljanja planova i procesa uključujući CPR¹²⁶ i PSL.¹²⁷ Nakon što je utvrđen cijeli relevantni sadržaj i povezan sa SUO web site¹²⁸, preveden je na SUO-KIF jezik tj. izvršeno je tzv. sintaktičko sjedinjenje (engl. syntactic merge). Uslijedilo mnogo teže semantičko povezivanje (engl. semantic merge), odnosno kombiniranje svih različitih ontologija u jedan konzistentan i opsežan okvir. Ontologije su prvo podijeljene u dvije klase. Prva klasa ontologija sadrži John Sowa te Russell i Norvig ontologije gornje razine, a druga klasa sadrži sve ostale ontologije. Dvije ontologije gornje razine su spojene u jednu konceptualnu strukturu. Budući su oba izvora ontologija vrlo kompaktna i sadrže značajnu količinu preklapanja sadržaja, ovo sjedinjenje nije predstavljalo ozbiljne praktične ili teorijske probleme. Ovako sjedinjena ontologija je upotrijebljena kao osnova za svrstavanje svih drugih sadržaja pretvorenih u SUO-KIF jezik.¹²⁹

¹¹⁹ Russell, Stuart; Norvig, Peter. *Artificial Intelligence: A Modern Approach* // Artificial Intelligence 1-2(1995), 932.

¹²⁰ Allen, James F. Towards a general theory of action and time // *Artificial Intelligence*, 23(1984), 123-154.

¹²¹ Casati, R.; Varzi, A. *Holes and Other Superficialities*. Cambridge, Mass. : MIT Press, 1994. Str. 45.

¹²² Simons, Peter M. *Parts: a Study in Ontology*. Oxford : Clarendon Press ; New York : Oxford University Press, 1987. Str. 63.

¹²³ Smith, Barry. Fiat Objects // Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, Proceedings of the 11th European Conference on Artificial Intelligence on *Artificial Intelligence*, Amsterdam, 8 August 1994, Amsterdam: European Coordinating Committee for *Artificial Intelligence* / ed. by N. Guarino, L. Vieu and S. Pribbenow. 1784(1994), 15-23. Dostupno na URL: <http://link.springer.com/article/10.1023%2FA%3A1017948522031> (8.9.2015.).

¹²⁴ Borgo, Stefano; Guarino, Nicola; Masolo, Claudio. An Ontological Theory of Physical Objects // *Proceedings of Eleventh International Workshop on Qualitative Reasoning (QR'97)* / ed. by L. Ironi. Cortona (Italia), 3-6 Giugno, 1997. Str. 223-231.

¹²⁵ Borgo, Stefano; Guarino, Nicola; Masolo, Claudio. A Pointless Theory of Space Based on Strong Connection and Congruence // *Principles of Knowledge Representation and Reasoning (KR96)* / ed. by L. C. Aiello and J. Doyle, Morgan Kaufmann, 1996. Dostupno na URL: <http://www.loa.istc.cnr.it/old/Papers/KR96.pdf> (27.11.2015.).

¹²⁶ Pease, Adam R.; Carrico, Todd M. JTF-ATD Core Plan Representation: A Progress Report // *Proceedings of the AAAI-97 Spring Symposium on Ontological Engineering*. 1997. Dostupno na URL: <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-012.pdf> (27.11.2015.).

¹²⁷ *The Process Specification Language (PSL): Overview and Version 1.0 Specification* / Schlenoff, Craig ... <et al.>. NISTIR 6459, National Institute of Standards and Technology, Gaithersburg, MD, 2000. Dostupno na URL: <http://www.mel.nist.gov/msidlibrary/doc/nistir6459.pdf> (27.11.2015.).

¹²⁸ SUO. Dostupno na URL: <http://suo.ieee.org/refs.html> (27.11.2015.).

¹²⁹ SUMO ontološki jezik (logička izražajnost punoga prvog reda) za predstavljanje znanja je KIF verzija (*Stanford Knowledge Interchange Format*) nazvana SUO-KIF. Specifikacija trenutne SUO-KIF verzije se mogu pogledati na URL: <http://suo.ieee.org/suo-kif.html> (27.11.2015.).

3.2.3 UMBEL

Projekt *UMBEL* (engl. *Upper Mapping and Binding Exchange Layer*) je ekstrahirani podskup OpenCyc-a,¹³⁰ osiguravajući Cyc podatke u RDF ontologiji temeljene na SKOS-u i OWL 2. Prvi put je objavljen u srpnju 2008, održava ga Structured Dynamics.¹³¹ UMBEL je stvoren kako bi pridonio sadržajnoj interoperabilnosti na Web-u.¹³² Ima dvije glavne funkcije:¹³³

(1) Predstavlja ontologiju gornje razine s oko 28.000 referentnih pojmova izvučenih iz Cyc baze znanja te podijeljenih u temeljne i geo module (za slučajeve kada se ne koriste GeoNames pridruživanja), koji su organizirani u više od 30 uglavnom razdvojenih (disjunktnih, engl. disjoint) nadvrsta (engl. SuperType, klase koje su uglavnom disjunktne jedna s drugom)¹³⁴. Stvorena je kako bi osigurala opća pridruživanja u svrhu povezivanja različitih ontologija ili shema. UMBEL referentna struktura je široko primjenjiva kao usmjeravajući čvorovi za bilo koju domenu znanja, koji su jedinstveno strukturirani i međusobno povezani. UMBEL referentna pojmovna ontologija je u biti sadržajni graf (engl. content graph) predmetnih čvorova povezanih jednih s drugim putem širi-od (engl. broader-than, nadređeni) i uži-od (engl. narrower-than, podređeni) odnosa. S druge strane, UMBEL referentni pojmovni model može biti povezan na vanjske klase i individue (instance i imenovane entitete) putem skupa relacija, ekvivalenata ili svojstava za poravnavanje. Cilj je pružiti jedinstveni okvir širokih predmeta i tema (UMBEL reference concepts), odnosno povezujuće čvorove prikladne za pridruživanje relevantnih web dostupnih sadržaja sa specifičnim ciljem promicanja interoperabilnosti i razumijevanja jedinstvene referentne strukture i njegovih povezanih izvora. Skup UMBEL referentnih pojmova započinje s Cyc bazom znanja. Međutim, budući su njegov opseg i sofisticiranost daleko nadmašili ono što je prilagodljivo za laku referentnu strukturu, provodi se pročišćavanje Cyc-a do značajnog stupnja. Svi UMBEL referentni pojmovi i njihovi odnosi izvedeni su iz OpenCyc ontologije, što znači da je UMBEL čisti podskup od OpenCyc-a. Dakle, svrha UMBEL-a je osigurati

¹³⁰ OpenCyc. Dostupno na URL: <http://www.opencyc.org/> (27.11.2015.).

¹³¹ Structured Dynamics. Dostupno na URL: <http://structureddynamics.com/> (27.11.2015.).

¹³² Upper Mapping and Binding Exchange Layer (UMBEL). Dostupno na URL: <http://www.umbel.org/> (27.11.2015.).

¹³³ Bergman, M.K. UMBEL: A Subject Concepts Reference Layer for the Web. Dostupno na URL: <http://www.slideshare.net/mkbergman/umbel-a-subject-concepts-reference-layer-for-the-web> (26.10.2015.).

¹³⁴ UMBEL Specification. Dostupno na URL: http://techwiki.umbel.org/index.php/UMBEL_Specification (26.10.2015.).

fiksni skup pojmova (engl. *UMBEL Reference Concepts*) prema kojima ta aproksimativna poravnanja mogu biti usmjerena. Ovaj skup fiksnih referentnih točaka nije konačan, niti su namijenjeni za oblikovanje svijeta u cijeloj njegovoj složenosti. Umjesto toga, cilj je pružiti niz fiksnih referenci prema kojima se može usmjeravati i pregledavati sastavni sadržaj. Praktična zamisao je tematika (odnosno očemnost, engl. *aboutness*).¹³⁵ Može li se opisati sastavna informacija pomoću onoga *o čemu je* tako da je moguće okupljati slične relevantne sadržaje? Ova referentna struktura može osigurati početne čvorove za pridruživanje dostupnih relevantnih web sadržaja te pružiti snažan skup konstelacija za vođeno otkrivanje povezanih skupova podataka i sadržaja. UMBEL rječnik je dobra osnova za stvaranje domenskih ontologija koje se također mogu ponašati kao referentne ontologije (engl. *reference ontologies*) unutar svojih vlastitih domena¹³⁶.

(2) Rječnik za pomaganje ontologijskih pridruživanja. Sami UMBEL ne rješava problem interoperabilnosti, zamišljen je kao koristan skup prvih koraka na putu prema tome. Njegova svrha je osigurati opći rječnik (UMBEL rječnik) klasa i svojstava za opisivanje domenskih ontologija sa specifičnim ciljem promicanja interoperabilnost s vanjskim bazama podataka i domenama. UMBEL navodi neke predikate (povezujući glagole ili svojstva) za povezivanje različitih informacijskih izvora zajedno. Dakle, *UMBEL* rječnik je stvoren za prepoznavanje izvora informacija različitog konteksta i strukture. Rječnik značenja je potreban kako bi se mogli izraziti potencijalni odnosi između dva izvora informacija poštujući njihove razlike u strukturi i opsegu.

UMBEL je klasna struktura korištena u Linked Data, zajedno s OpenCyc, YAGO, i DBpedia ontologijama. Budući je strogi podskup OpenCyc pojmova i odnosa, pridruživanja na UMBEL mogu iskoristiti alate za zaključivanje (engl. *reasoning engines*) unutar Cyc-a. Prva UMBEL produkcija i stabilno izdanje bila je verzija 1.00 objavljena u veljači 2011¹³⁷, pridružena je na 60% Wikipedije te osigurava poveznice na više od 2 miliona stranica

¹³⁵ Prijevod termina *aboutness* = *tematika* također je preuzet iz hrvatskog prijevoda modela FRASAD (Uvjeti za funkcionalnost predmetnih autoriziranih podataka : konceptualni model / urednice Marcia Lei Zeng, Maja Žumer i Athena Salaba ; IFLA-ina Radna skupina za Uvjete za funkcionalnost predmetnih preglednih zapisa (FRASAR) ; odobrio Stalni odbor IFLA-ine Sekcije za klasifikaciju i indeksiranje u lipnju 2010. ; [s engleskog preveo Predrag Perožić]).

Važno je napomenuti da je u stručnoj literaturi na hrvatskom jeziku termin *aboutness* poznat i pod nazivom očemnost. Prijevod očemnost upotrijebljen je u prijevodu knjige Intelektualne osnove organizacije informacija / Elaine Svenonius ; prevela Mirna Willer. Lokve, 2005. Str. 46. Prijevod je predložila Marija Laszlo, a u prijevodu spomenute knjige upotrijebila ga je prevoditeljica Mirna Willer.

¹³⁶ UMBEL Specification, loc. cit.

¹³⁷ UMBEL 1.00. Dostupno na URL: <http://umbel.org/resources/news/finally-umbel-v-100/> (26.10.2015.).

Wikipedia-e (engleska verzija). Uključujući OpenCyc, UMBEL ima oko 48,000 formalnih pridruživanja na DBpedia¹³⁸-u, PROTON¹³⁹, GeoNames¹⁴⁰ i schema.org¹⁴¹ te osigurava veze na više od 2 milijuna Wikipedijinih stranica engleske verzije.

3.2.4 Cyc

Cyc projekt se započeo razvijati 1986 godine. Sastoji se od temeljne ontologije i nekoliko domenski specifičnih ontologija. Pokrenut je 1984. godine unutar Microelectronics and Computer Technology Corporation (MCC) te se 1994. godine odvojio u nezavisnu kompaniju Cycorp koja 2002. objavljuje prvu verziju (0.7.0) OpenCyc,¹⁴² koja je javno dostupnog koda (oko 40 posto) ontologije. U veljači 2005., objavljena je verzija 0.9.0 (oko šest puta veća od 0.7.0). Postoji i neznatno šira verzija za istraživače, nazvana ResearchCyc. Cyc ontologija ima svoj vlastiti, svrhovito građeni program zaključivanja, koji se pojavljuje u paketu zajedno s OpenCyc i ResearchCyc. Cilj Cyc projekta formalne ontologije Lenat identificira kao „uhvatiti zdravi razum, odnosno sve što jedan šestogodišnjakinja zna, a što mu omogućuje da razumije prirodni jezik i počne nezavisno učiti.“^{143, 144} Primarni cilj Cyc projekta je izgraditi veliku bazu znanja koja će sadržati formalizirano znanje prikladno za različita zaključivanja i rješavanje problemskih zadataka u različitim domenama. Cyc projekt se temelji na preko 600.000 kategorija jer je svrha Cyc ontologije omogućiti upotrebu znanja preko više domena, više od 2.2 milijuna tvrdnji (činjenice i pravila) opisujući više od 250,000 termina, uključujući gotovo 15,000 predikata.¹⁴⁵ Cyc baza znanja se tradicionalno dijeli na ontologije gornje, srednje i niže razine (engl. upper, middle, lower). Ove razlike su namijenjene dohvaćanju razine općenitosti informacije sadržane u njima. Ontologija gornje razine je ograničena na široke i apstraktne pojmove te je kao takva je najmanja, ali je najšire upućivano područje Cyc ontologije. Namijenjena je za pojmove svjetovnog sadržaja, matematike te vrste odnosa. Ontologija srednje razine je namijenjena za dohvaćanje sloja

¹³⁸ DBpedia. Dostupno na URL: <http://wiki.dbpedia.org/> (26.10.2015.).

¹³⁹ PROTON (PROTo ONtology). Dostupno na URL: <http://ontotext.com/proton-ontology/> (26.10.2015.).

¹⁴⁰ GeoNames Ontology. Dostupno na URL: <http://www.geonames.org/ontology/documentation.html> (26.10.2015.).

¹⁴¹ Schema.org. Dostupno na URL: <http://schema.org/> (26.10.2015.).

¹⁴² OpenCyc. Dostupno na URL: www.opencyc.org (26.10.2015.).

¹⁴³ Lenat, D. B. Cyc : a large-scale investment in knowledge infrastructure. // Communications of the ACM, 38, 11(1995), 32-59.

¹⁴⁴ Lenat, D. B.; R. V. Guha. Building large knowledge-based systems : representation and inference in the Cyc project. Reading (MA) : Addison Wesley, 1990.

¹⁴⁵ Sowa, J. The challenge of knowledge soup. Dostupno na: <http://www.jfsowa.com/pubs/challenge.pdf> (27.09.2015.).

apstrakcije koji se često koristi, npr. geospacijalni odnosi, široko znanje ljudske interakcije, svakodnevne stavke i događaji. Ontologija niže razine sadrži domenski specifično znanje, kao što je područje proučavanja poput kemije ili informacije o određenoj osobi. Ovaj sloj čini najveći segment znanja u Cyc bazi znanja, ali se najmanje široko primjenjuje. Ontološki inženjeri su izvorno zamislili da Cyc projekt krene od početne ručno kodirane baze znanja prema sustavu koji automatizirano usvaja znanje, tako da nekvalificirane osobe mogu čitati i postavljati pitanja¹⁴⁶ Drugi veliki cilj projekta je izgradnja sučelja na prirodnom jeziku jer su nazivi koji predstavljaju pojmove u Cyc ontologiji, odnosno riječi na engleskom ili bilo kojem drugom jeziku međusobno odvojeni, što zahtijeva pridruživanje. Stoga, Cyc projekt indeksira nazive i pokušava opisati njihovo značenja na strojno razumljivi način u CycL jeziku. Cycorp je ugradio druge ontologije i baze podataka u Cyc, kao što su Federal Information Processing Standards (FIPS), Central Intelligence Agency (CIA), WorldFactbook¹⁴⁷ i WordNet. Nažalost, konstruktori semantičkog weba izvan Cycorpa koristili su do sada vrlo malo ovu ontologiju¹⁴⁸ jer je njen sustav kategorija izrazito složen te zahtijeva filozofsku uvježbanost za razumijevanje sustava, a dugo je vremena bio i oskudno obrazložen. Provodljivost zaključaka također zadaje dosta brige u pogledu krajnje izražajnosti CycL jezika. Ramachandran, Regan i Goolsbey¹⁴⁹ su istražili koliko taj problem može biti ublažen tako da se na dijelove baze znanja primijeni logika prvog reda. Budući Cyc obuhvaća integrirani skup aplikacija, koji su kao takvi nepogodni za razdvajanje, čini ga „nedistribuiranom” tehnologijom koja je u suprotnosti vizije semantičkog weba.¹⁵⁰

Podskup Cyc ontologije je besplatno izdan pod nazivom OpenCyc te predstavlja opću ontologiju i bazu znanja otvorenog pristupa. Dakle, OpenCyc je slobodno dostupan dio nekog općeg znanja u Cyc sustav s namjerom uspostavljanja općeg rječnika za automatsko zaključivanje. Uključuje i besplatno dostupni izvršni server znanja (engl. Knowledge Server) sa strojem za zaključivanje i druge alate za pristupanje, korištenje i proširivanje sadržaja baze

¹⁴⁶ Searching for common sense : populating Cyc from the web. / Matuszek, C. et al. // Proceedings of the twentieth national conference on artificial intelligence 20(2005)

Dostupno na: <http://www.cyc.com/doc/white~paperdAA4IO51MatuszekC.pdf>

¹⁴⁷ Reed, Stephen L; Lenat, Douglas B. Mapping ontologies into Cyc. // AAAI 2002 Conference workshop on ontologies for the semantic web (2002). Dostupno na:

http://www.cyc.com/doc/white_papers/mapping-ontologes-into-cyc_v31.pdf (30.10.2015.).

¹⁴⁸ Integrating descriptions of knowledge management learning activities into large ontological structures : a case study / Sicilia, Miguel-Ángel...[et al.] // Data & Knowledge Engineering 57, 2(2006), 111-121.

¹⁴⁹ Ramachandran, Deepak; Reagan, Pace; Goolsbey, Keith. First-orderized ResearchCyc : expressivity and efficiency in a common-sense ontology. // AAAI Workshop on Contexts and Ontologies: Theory, Practice and Applications Dostupno na: <http://www.cyc.com/doc/white-papers/fofication.pdf> (30.10.2015.).

¹⁵⁰ usp. Legg, Catherine, op. cit. (bilj. 110), str. 452-445.

znanja. U trenutnoj verziji OpenCyc uključuje ontologiju od 47,000 termina, koji su definirani i razrađeni pomoću 306,000 tvrdnji. OpenCyc pruža osnovne pojmove koji se mogu odmah koristiti i jednostavno proširiti. U srpnju 2009 je izdana njegova verzija 2.0, koja je više ili manje cjelovita verzija dostupna za nekomercijalnu upotrebu pod imenom ResearchCyc. Znanje u OpenCyc-u je mali podskup znanja u ResearchCyc bazi znanja, koja je verzija Cyc-a usmjerena na istraživačku zajednicu. Osim taksonomske informacije sadržane u OpenCyc, ResearchCyc uključuje znatno više semantičko znanje (tj. dodatnih činjenica) o pojmovima u svojoj bazi znanja, veliki rječnik, parsiranje engleskog i alate za stvaranje, Java temeljeno sučelje za uređivanje i pretraživanje znanja te sustav za integraciju podataka temeljen na ontologiji.

3.2.5 BabelNet

BabelNet¹⁵¹ (engl. Building a Very Large Multilingual Semantic Network) projekt je velike višejezične semantičke mreže, koja automatski stvara resurs pomoću metodologije integriranja leksičkog i enciklopedijskog znanja iz WordNet-a i Wikipedi-je, kao i integriranjem leksičkih praznina jezično siromašnih resursa uz pomoć strojnog prevođenja u svrhu obogaćivanja resursa leksičkim informacijama. Takvi višejezični leksički pojmovi nazivaju se *Babel sinonimski grozdovi* (engl. *Babel synsets*). Rezultat je „enciklopedijski rječnik“, koji osigurava pojmove i imenuje entitete leksikalizirane na mnogim jezicima te je povezan s velikom količinom semantičkih odnosa. Pojmovi i odnosi u BabelNet-u se pobiru (*harvestiraju*) iz najvećeg dostupnog semantičkog leksikona engleskog jezika WordNet-a i opće suradničke enciklopedije Wikipedije. Prikupljanje se odvija u fazama: (1) iz WordNet-a sva dostupna značenja riječi (kao pojmovi) te svi leksički i semantički pokazivači između sinonimskih grozdova (kao odnosi); (2) iz Wikipedije sve enciklopedijske stavke (odnosno Wikistranice kao pojmovi) i semantički nespecificirani odnosi iz hiperlinkiranog teksta. WordNet i Wikipedia se mogu preklopiti u smislu pojmova i odnosa te se pridruživanjem između stranice Wikipedije i WordNeta dobiva presjek ovih dvaju resursa znanja kao

¹⁵¹ BabelNet je višejezična leksikalizirana semantička mreža i ontologija razvijena na Linguistic Computing Laboratory na Department of Computer Science na Sapienza University od Rome. BabelNet je dobio sredstva iz European Research Council (ERC) u ugovoru br. 259234. Koncept i dizajn je pod pravima Roberto Navigli. BabelNet i njegov API su licencirani pod Creative Commons Attribution-Non Commercial-Share Alike 3.0 License. // BabelNet. Dostupno na URL: <http://www.babelnet.org> (3.06.2016).

jedinstveni resurs, čime se izbjegavaju dvostruki pojmovi i omogućuje se međusobno nadopunjavanje njihovih pojmovnih sadržaja. Višejezičnost se omogućava prikupljanjem leksičkih realizacija dostupnih pojmova na različitim jezicima koristeći ljudski generirane prijevode navedene u Wikipediji (tzv. međujezične poveznice) i sustave za strojno prevođenje, a zatim se višejezični Babel sinonimski grozdovi povezuju uspostavljanjem semantičkih odnosa između njih.

3.2.6 CIDOC

CIDOC (engl. Conceptual Reference Model, CRM) je proširiva ontologija za koncepte i informacije u kulturnoj baštini i muzejskoj dokumentaciji. To je međunarodni standard (ISO 21127:2006) za kontroliranu razmjenu informacija o kulturnoj baštini.¹⁵² Arhivi, knjižnice, muzeji i ostale kulturne institucije potiču upotrebu CIDOC CRM-a u svrhu poboljšanja dostupnosti srodnih informacija. CIDOC CRM je proizašao iz CIDOC Documentation Standards Group u International Committee for Documentation of the International Council of Museums.¹⁵³ Rad je usmjeren na razvoj entitet-odnos modela za muzejske informacije do 1994. godine, međutim u 1996. godini pristup je pomaknut na metodologiju objektno orijentiranog modeliranja (engl. object-oriented modeling), što je rezultiralo "*CIDOC Conceptual Reference Model (CRM)*" u 1999. godini. Proces standardizacije CIDOC CRM započeo je 2000. godine te je dovršen 2006. godine s njegovim prihvaćanjem kao ISO 21127 standarda. Svrha CIDOC CRM je osigurati referentni model i informacijski standard koji bi muzeji i ostale institucije kulturne baštine mogle koristiti za opis svojih zbirki u svrhu poboljšanja razmjene informacija. CIDOC Conceptual Reference Model (CRM) osigurava definicije i formalnu strukturu za opisivanje pojmova i odnosa upotrijebljenih u dokumentaciji kulturne baštine, kako bi se omogućilo zajedničko razumijevanje informacija kulturne baštine i osigurao zajednički proširivi semantički okvir koji može pridružiti bilo koju informaciju koja se odnosi na kulturnu baštinu. Namjera je da to bude zajednički jezik za stručnjake područja kulturne baštine u svrhu formuliranja zahtjeva informacijskim sustavima i vodiča dobrog konceptualnog modeliranja. Na taj se način osigurava „semantička poveznica“

¹⁵² ISO 21127:2006 (Information and documentation — A reference ontology for the interchange of cultural heritage information) Dostupno na URL: http://www.iso.org/iso/catalogue_detail?csnumber=34424 (21.08.2015.).

¹⁵³ CIDOC Conceptual Reference Model (CRM). Dostupno na URL: <http://www.cidoc-crm.org/> (30.10.2015.).

potrebna za posredovanje između različitih izvora informacija kulturne baštine objavljenih od strane muzeja i arhiva¹⁵⁴.

Usvajanjem formalne semantike CIDOC CRM modela uspostavljeni su preduvjeti za strojnu interoperabilnost i integraciju. Stoga je CIDOC CRM model dobro postavljen kao važan informacijski standard i referentni model za inicijativu semantičkog weba te je dobar vodič općenito za modeliranje podataka ili baza znanja. Mnoge institucije kulturne baštine istražuju ili grade aplikacije koje upotrebljavaju CIDOC CRM¹⁵⁵ model, koji tehnički vodi prema softverskim aplikacijama koje intenzivno koriste XML i RDF.

Sljedeći uspješnu standardizaciju CIDOC CRM-a, FRBRoo¹⁵⁶ inicijativa ga je 2006. godine započela usklađivati s FRBR (engl. Functional Requirements for Bibliographic Records). Cilj ove inicijative je osigurati formalnu ontologiju namijenjenu za dohvaćanje i predstavljanje temeljne semantike bibliografskih informacija te olakšati integraciju, posredovanje i razmjenu bibliografskih i muzejskih informacija.^{157,158} CIDOC objektno-orijentiran „Conceptual Reference Model“ (CRM) je domenska ontologija, koja također uključuje svoju vlastitu verziju ontologije gornje razine (engl. upper ontology). Temeljne klase pokrivaju vremenski raspon, događaje, materijalne i nematerijalne stvari.¹⁵⁹ U Hrvatskoj su o primjeni CIDOC-CRM konceptualnog referentnog modela dostupne informacije i izlaganja o sljedećim projektima. Zbirke na portalu Hrvatska kulturna baština¹⁶⁰ su opisane u različitim standardima, između ostalog i CIDOC-CRM.¹⁶¹ Na 2. skupu muzejskih dokumentarista Hrvatske održanom u Zadru 2015. godine predstavljena su sljedeća izlaganja. Mikica Maštrović je predstavila *Steve* istraživački projekt u službi korisnika gdje je izlagala o CIDOC-CRM konceptualnom referentnom modelu koji je i nastao u svrhu što boljeg zadovoljenja korisnika i njihovih potreba. Prema tom modelu, dokumentacija zbirki uključuje detaljan opis pojedinačnog entiteta, skupine entiteta, kao i muzejsku zbirku u cijelosti. *Steve*

¹⁵⁴ Ibid.

¹⁵⁵ CIDOC CRM Applications. Dostupno na URL: http://www.cidoc-crm.org/uses_applications.html (30.10.2015.).

¹⁵⁶ FRBRoo Introduction. Dostupno na URL: http://www.cidoc-crm.org/frbr_inro.html (30.10.2015.).

¹⁵⁷ Ibid.

¹⁵⁸ Doerr, Martin; LeBoeuf, Patrick. Modelling Intellectual Processes: The FRBR-CRM Harmonization // Digital Libraries: Research and Development 4877, 2007, 114-123. Dostupno na URL: http://www.cidoc-crm.org/docs/doer_le_boeuf.pdf (30.10.2015.).

¹⁵⁹ CIDOC Conceptual Reference Model. Dostupno na URL: http://www.cidoc-crm.org/cidoc_core_graphical_representation/graphical_representation.html (30.10.2015.).

¹⁶⁰ Hrvatska kulturna baština. Dostupno na URL: <http://www.kultura.hr> (6.06.2016.).

¹⁶¹ Sažetak 3. sjednice Vijeća projekta „Hrvatska kulturna baština“, održane 09. Siječnja 2013. godine. / Ministarstvo kulture Republike Hrvatske. Dostupno na URL: <http://www.min-kulture.hr/default.aspx?id=9946> (6.06.2016.).

istraživački projekt prvi put je predstavljen široj javnosti 2004. godine., a u njega su bili uključeni stručnjaci različitih muzeja i brojni volonteri. Projekt je nastao suradnjom muzealaca i šire društvene zajednice. Da bi se uspješno realizirao, za njega su izrađeni posebni softverski alati, a provodio se u suradnji sa stručnjacima Sveučilišta Maryland. Istraživački projekt nazvan je T3 (Text, Tags, Trust). Njime se željelo ukazati na važnost korisnika muzeja pri sadržajnom opisivanju muzejske građe. Rezultati do kojih je došao istraživački tim Steve muzeja zajedno sa stručnjacima Sveučilišta u Torontu potvrdili su da je uključivanje šire zajednice u sadržajno označivanje u umjetničkim muzejima dovelo do poboljšanja pristupa online zbirkama. Steve projekt koristi folksonomiju koja podrazumijeva interaktivnu dvosmjernu komunikaciju između korisnika i računala, te korisnika međusobno, pri čemu korisnik od pasivnog postaje aktivni sudionik i jednim od stvaratelja muzejske dokumentacije. Da bi se sve planirano moglo i ostvariti, trebalo je osigurati arhitekturu T3 sustava koja je sastavljena od niza dijelova koji čine jedinstvenu cjelinu.¹⁶²

Održala se radionica *Muzejski standardi u Hrvatskoj (promišljanje, stvaranje, izrada)* osmišljena s namjerom predstavljanja standarda te stavljanjem posebnog naglaska na standarde koji postoje u muzejskom okruženju Amerike i Europe, među kojima je posebno predstavljen i CIDOC-CRM.¹⁶³ Predstavljena je moderna web platforma INDIGO za online prikaz, upravljanje i objavu digitalnih zbirki, vođenje primarne i sekundarne muzejske dokumentacije od tvrtke ArhivPro. Navedena platforma je namijenjena muzejima, galerijama i dokumentacijskim institucijama, a primjerena je za razvoj kooperativne mreže u kojoj institucije mogu dijeliti serverske resurse i razvijati vlastite ili zajedničke tezauruse i tematske zbirke. Usklađena je s međunarodni standardima te sadrži dijelove CIDOC-CRM modela.¹⁶⁴

¹⁶² Maštrović, Mikica. Steve istraživački projekt u službi korisnika ili kako nam druga iskustva mogu pomoći u razumijevanju onoga što radimo, str. 14. // Dokumentacija i korisnici, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.

¹⁶³ Dejanović, Antonija; Svrtan, Zoran. Muzejski standardi u Hrvatskoj (promišljanje, stvaranje, izrada), str. 20 // Dokumentacija i korisnici, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.

¹⁶⁴ Crnković, Kristijan. INDIGO – platforma za upravljanje i objavu digitalnih zbirki, vođenje primarne i sekundarne muzejske dokumentacije, str. 31. // Dokumentacija i korisnici, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.

3.2.7 WordNet

U prethodnim projektima je već često spominjan WordNet,¹⁶⁵ nezaobilazni projekt semantičkog weba. Skupina psiholingvisti i lingvisti je 1985. godine na Sveučilištu Princeton (George A. Miller, Katherine Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross), započela izradu WordNet leksičke baze podataka engleskog jezika te je s vremenom projekt prerastao u računalno podržan prikaz temeljnog semantičkog rječnika engleskog jezika. WordNet odražava jezik cjelokupne jezične zajednice te je građen na jezičnim resursima (korpus i rječnik). WordNet razlikuje doslovno pojavljivanje riječi u tekstovima i njihov stvarni smisao¹⁶⁶ te predstavlja računalni resurs na semantičkoj razini koji daje podatke o leksiku pojedinoga jezika. WordNet trenutno sadrži približno 95 600 različitih oblika riječi (51 500 jednostavnih riječi i 44 100 kolokacija¹⁶⁷) organiziranih u 70 100 značenja riječi, odnosno skupove sinonima.¹⁶⁸ Svrha WordNet-a je stvoriti kombinaciju rječnika i tezaurusa koji je intuitivnije upotrebljiv te podržati automatsku analizu teksta i primjenu umjetne inteligencije. Početna zamisao je bila pružiti pomoć za konceptualno, a ne samo abecedno, pretraživanje rječnika. Stoga je početna ideja bila pronaći uzorke semantičkih odnosa među leksemima engleskoga jezika, bez namjere da se izgradi cjeloviti leksikon.¹⁶⁹ Najočitiija razlika između WordNet i standardnog rječnika je ta što WordNet obuhvaća samo četiri najveće skupine vrsta riječi (semantički pune riječi): imenice, glagole, pridjeve i priloge.¹⁷⁰ Uz pojedinačne lekseme, u WordNetu se mogu još naći kolokacije, jer su nerijetko jedini mogući izraz pojedinog koncepta te samo najčestotniji idiomi¹⁷¹, jer su oni samo jedna od mogućih leksikalizacija nekog koncepta. Imenice, glagoli, pridjevi i priloge su grupirani za predstavljanje jedinstvenih leksičkih pojmova. Imenice i glagoli su organizirani u hijerarhije,

¹⁶⁵ WordNet. Dostupno na URL: <http://www.wordnet.princeton.edu> (3.06.3016).

¹⁶⁶ Miller, G. A. Introduction to Wordnet: An on-line lexical database. // *International Journal of Lexicography* 3, 4 (1990), 29.

¹⁶⁷ Kolokacija, *lingv.* obvezatna ili uobičajena veza riječi koja nije određena gramatikom [*podnijeti molbu ali uložiti žalbu*]. // Hrvatski jezični portal. Dostupno na URL: <http://hjp.novi-liber.hr/index.php?show=search> (14.07.2015.).

¹⁶⁸ Miller, G. A., loc cit.

¹⁶⁹ Miller, G. A.: *Nouns in WordNet*. // *Wordnet: An electronic lexical database*. / ed. by Christiane Fellbaum. Cambridge, Mass : MIT Press, 1998. Str. XVII.

¹⁷⁰ Fellbaum, Christiane. *A Semantic Network of English: The Mother of All WordNets*. // *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. / ed. by Piek Vossen. Dordrecht: Boston: London: Kluwer Academic Publishers, 1998. Str. 138.

¹⁷¹ Idiom, *lingv.* *lingv. a.* svaki od entiteta jezika koji se ne želi označiti ni kao jezik, ni kao dijalekt, ni kao govor *b.* poseban i prepoznatljiv govor nekog kraja ili pojedinog govornika *c.* riječ ili izraz svojstven nekom narječju, govoru *d.* osebujna riječ ili izraz, ob. teško prevodiv ili neprevodiv na drugi jezik [*šilo za ognjilo*]. // Hrvatski jezični portal. Dostupno na URL: <http://hjp.novi-liber.hr/index.php?show=search> (14.07.2015.).

definirani s hiperonimima ili *is-a* odnosima. Mreža imenica je daleko dublja nego kod ostalih vrsta riječi. Glagoli imaju poprilično razgranatiju strukturu, a pridjevi su organizirani u mnoge različite skupine (klastere). Prilozi su definirani u terminima pridjeva od kojih su izvučeni te stoga nasljeđuju njihovu strukturu.

Međutim, značajka WordNet-a je njegov pokušaj organizacije leksičkih informacija u obliku značenja riječi, a ne oblicima riječi. U tom smislu, WordNet više nalikuje tezaurusu nego rječniku. Problem s abecednim tezaurusom je višak zapisa (stavki, engl. entry): ako su riječi *W_x* i *W_y* sinonimi, par treba dva puta upisati, jednom po abecedi pod *W_x* i drugi put ponovo po abecedi pod *W_y*. Problem s tematskim tezaurusom je jer to zahtijeva dva pretraživanja koja korisniku oduzimaju vrijeme (tj. udvostručenje njegovo vrijeme pretraživanja), a tu vrstu mehaničkog zadatka računalo može obavljati brzo i učinkovito. Skupina srodnih riječi je međusobno povezana putem konceptualno semantičkih i leksičkih odnosa, što rezultira stvaranjem mreže povezanih riječi i pojmova. WordNet grupira riječi u svoje osnovne gradbene jedinice tj. skupove sinonima nazvane *sinonimski grozdovi*¹⁷² (engl. synsets), osiguravajući kratke općenite definicije značenja te bilježeći različite semantičke veze između sinonimnih skupova. Dakle, skup riječi koje dijele jedan smisao naziva se *sinonimski grozd*. Značenje sinonimskih grozdova je razjašnjeno objašnjenima (glosama) koje sadrže definicije i/ili primjere rečenica s primjerima uporabe. Svaki sinonimski grozd identificira jedan smisao (tj. semantički koncept), odnosno riječi različitog značenja su u različitim sinonimskim grozdovima. Dakle, jedan sinonimski grozd je zapravo koncept kojemu su pridružene njegove moguće leksikalizacije – *sinonimski leksemi*¹⁷³. Sinonimija se promatra kao zamjenjivost u istome kontekstu te stoga svi leksemi pojedinog sinonimskog grozda pripadaju istoj vrsti riječi. Riječi s višestrukim značenjem (višeznačne riječi) pripadaju više sinonimskih grozdova. Sinonimski grozd sadrži grupu sinonimnih riječi ili kolokacija (niz riječi koje idu zajedno kako bi se stvorilo određeno značenje). No, osim što sadrži listu sinonimskih leksema riječi, svaki sinonimski grozd istovremeno sadržava i skup semantičkih odnosa kojima je povezan s drugim grozdovima.¹⁷⁴ Ovi sinonimski grozdovi ne objašnjavaju

¹⁷²Termin *sinonimski grozd* je uveden kao ekvivalent engleskome *synset*. U hrvatskoj literaturi se koristio termin *sinskup* (Bekavac, B.; Tadić, M.; Šojat, K.: *Zašto nam treba Hrvatski WordNet?* Dostupno na URL: <http://www.hnk.ffzg.hr/txts/ksbbmt4HDPL2004.pdf> (12.01.2016.).

¹⁷³*Sinonimski leksem* odgovara engleskome terminu *literal*, koji se isto tako upotrebljavao u hrvatskoj literaturi o WordNetu (Bekavac, B.; Tadić, M.; Šojat, K., loc cit.). Termin *sinonimski leksem* predstavlja svaku pojedinu lemu unutar sinonimskoga grozda.

¹⁷⁴Gomes, P.: Software Design Retrieval Using Bayesian Networks and WordNet. // *Advances in Case-Based Reasoning. 7th European Conference, ECCBR. Proceedings.* / ed. by P. Funk, P. González Calero. Berlin: Springer, 2004. Str. 187.

što su pojmovi, oni samo naznačuju da pojmovi postoje. Pretpostavlja se da su ljudi koji znaju engleski već naučili pojmove te se očekuje kako će ih prepoznati s liste riječi navedenih u sinonimskim grozdovima. Dakle, leksički matrica može biti predstavljena za teorijske svrhe pridruživanjem između pisane riječi i sinonimskih grozdova. Budući je engleski bogat sinonimima, sinonimski grozdovi su često dovoljni za razlikovne svrhe. Ponekad nije dostupan primjeren sinonim te u tom slučaju polisem može riješiti taj problem s kratkim objašnjenjem, kako bi se razlikovao smisao od drugih te se to može smatrati sinonimski grozd s jednim članom. Objasnjenje nije međutim namijenjeno za korištenje prilikom izgradnje novog leksičkog koncepta nekome tko nije već upoznat s njim te se razlikuje od sinonima po tome jer se ne koristi kako bi se pristupilo informacijama pohranjenim u mentalnom leksikonu. Svoju svrhu ispunjava jer korisniku WordNet-a omogućava (za kojeg se pretpostavlja da zna engleski) razlikovanje ovog smisla od ostalih s kojima može biti zbunjen. Budući se 'riječ' obično koristi kako bi se uputilo na izraz i povezani pojam, leksička semantika prepoznaje kako je 'riječ' dogovorena veza između leksičkog pojma i izraza što igra sintaktičku ulogu. U cilju smanjenja nejasnoća, oblik riječi će uputiti na stvarne iskaze, a značenje riječi će uputiti na leksički pojam čiji se oblik može upotrijebiti za izražavanje. Stoga se može kazati kako je ishodište leksičke semantike pridruživanje između oblika i značenja.¹⁷⁵ Pridruživanja između oblika i značenja riječi su mnoga jer neki oblici imaju nekoliko različitih značenja, a nekoliko značenja može biti izraženo s nekoliko različitih oblika. Polisemi i sinonimi se mogu promatrati kao komplementarni aspekti tog pridruživanja, što znači da su polisemi i sinonimi problemi koji nastaju u tijeku dobivanja pristupa informacijama u mentalnom leksikonu: slušatelj ili čitatelj koji prepoznaje oblik mora se nositi sa svojim polisemima; govornik ili pisac koji se nada da će izraziti smisao mora odlučiti između sinonima.

Korisnici jezika mogu započeti s značenjem i tražiti odgovarajuće oblike koji ga izražavaju, ili mogu početi s oblikom i dohvaćati odgovarajuća značenja. U svom početku, WordNet se bavio isključivo obrascima semantičkih odnosa između leksikaliziranih pojmova. Važni leksički odnosi su morfološki odnosi između oblika riječi, koje nije bilo plan uključivati u WordNet. Međutim kako je rad napredovao, postalo je sve jasnije kako se leksički odnosi ne mogu ignorirati te će se WordNet u praktičnoj primjeni morati nositi s flektivnom morfologijom tj. ukoliko je netko npr. stavio kursor računala na riječ *trees* i klikne

¹⁷⁵ Miller, G. A. Dictionaries in the Mind. *Language and Cognitive Processes* 1, 3(1986), 176.

kao znak zadavanja zahtjeva za informacijom, WordNet ne bi trebao odgovoriti da riječ nije u bazi podataka nego program treba maknuti pluralni sufiks te potražiti riječ *tree* koja je svakako u bazi podataka. Stoga, WordNet razlikuje semantičke i leksičke odnose te iako je naglasak i dalje na semantičkim odnosima između značenja, također su uključeni i odnosi između riječi. Semantički i leksički odnosi između leksema i sinonimskih grozdova se razlikuju za pojedine vrste riječi. Osim već spomenute sinonimije kao najvažnijega odnosa, WordNet kodira i hiperonimija i hiponimija (eng. hypernymy / hyponymy, tj. odnos između nad-pojma i pod-pojma), holonimija i meronimija (eng. holonymy/meronymy, tj. odnos između dijela i cjeline), antonimija (posebno važne za imenice)¹⁷⁶, polisemija (za pridjeve)¹⁷⁷, troponimija, implikacija, kauzalnost (za glagole).¹⁷⁸

Sinonimija je leksički odnos između oblika riječi te se zbog toga što joj je dodijeljena središnja uloga u WordNet-u obilježava u vitičastim zagradama '{' i '}', a drugi leksički odnosi se obilježavaju u uglatim zagradama "[" i "]". WordNet je organiziran semantičkim odnosima. Otkada su semantički odnosi relacije između značenja, a otkada značenja mogu biti predstavljena sinonimskim grozdovima, prirodno je razmišljati o semantičkim odnosima kao pokazivačima između sinonimskih grozdova. Karakteristika je semantičkih odnosa da su oni izmjenjivi tj. ako postoji semantički odnos R između značenja $\{x, x\phi, \dots\}$ i značenja $\{y, y\phi, \dots\}$, onda također postoji i relacija $R\phi$ između $\{y, y\phi, \dots\}$ i $\{x, x\phi, \dots\}$. Nazivi semantičkih odnosa će služiti dvostrukoj ulozi: ako se odnos između značenja $\{x, x\phi, \dots\}$ i $\{y, y\phi, \dots\}$ zove R, tada se R također koristiti za označavanje odnosa između pojedinih oblika riječi koji pripadaju tim sinonimskim grozdovima. Dakle, očito je da je najvažniji odnos za WordNet sličnost značenja, budući je sposobnost prosudbe tog odnosa između oblika riječi preduvjet za predstavljanje značenja u leksičkoj matrici. Ukoliko su koncepti predstavljeni sa sinonimskim grozdovima te ukoliko sinonimi moraju biti međusobno zamjenjivi, onda riječi u različitim sintaktičkim kategorijama ne mogu biti sinonimi (tj. ne mogu formirati sinonimski grozd) jer oni se ne mogu međusobno zamijeniti. Pogodno je za pretpostaviti da je odnos simetričan: ako je x sličan y, onda je y jednako sličan x.

Antonimija je leksički odnos između oblika riječi, a ne semantički odnos između značenja riječi. Na primjer, značenja $\{rise, ascend\}$ i $\{fall, descend\}$ mogu biti konceptualno

¹⁷⁶ Bekavac, B.; Tadić, M.; Šojat, K., op. cit. (bilj. 166), str. 4.

¹⁷⁷ Christiane. A Semantic Network of English: The Mother of All WordNets. // EuroWordNet: A Multilingual Database with Lexical Semantic Networks. / ed. by Piek Vossen. Dordrecht: Boston: London: Kluwer Academic Publishers, 1998. Str. 140.

¹⁷⁸ Usp. Nikolić-Hoyt, A. Konceptualna leksikografija: prema tezaursu hrvatskoga jezika. Zagreb: Hrvatska sveučilišna naklada, 2004. Str. 146-147.

suprotni, ali oni nisu antonimi¹⁷⁹ Antonimija pruža središnje organizacijsko načelo za pridjeve i priloge u WordNet-u.

Za razliku od sinonimije i antonimije, koje su leksički odnosi između oblika riječi, WordNet posjeduje između sinonimskih grozdova sljedeće odnose.

Hiponim¹⁸⁰/hiperonim¹⁸¹ (engl. hyponymy / hypernymy, tj. pod-pojma i nad-pojma) je semantički odnos značenja riječi. Primjerice, {*avor*} je hiponim od {*drvo*}, a {*drvo*} je hiponim od {*biljka*}. Mnogo pažnje je posvećeno hiponim/hiperonim odnosu (nazivani podređenost / nadređenost, podskup / nadskup, ISA odnos). Pojam predstavljen sinonimnim grozdom {*x, x', . . .*} će biti hiponim pojma predstavljenog sinonimnim grozdom {*y, y', . . .*} ukoliko izvorni govornici engleskog prihvate rečenice oblika kao što je *An x is a (kind of) y*. Odnos može biti predstavljen uključujući u {*x, x', . . .*} pokazivač na njegov hiperonim (nadređeni) i uključujući u {*y, y', . . .*} pokazivače na njegove hiponime.¹⁸²

Hiponim je tranzitivan i asimetričan,¹⁸³ budući je obično jedan hiperonim (nadređen), što stvara hijerarhijsku semantičku strukturu u kojoj će hiponim biti ispod svog hiperonima (nadređenog). Takav hijerarhijski prikaz se koristi u izgradnji sustava za pretraživanje informacija, odnosno u tzv. naslijeđenim sustavima (engl. inheritance systems)¹⁸⁴ jer hiponim nasljeđuje sva obilježja od višeg generičkog koncepta i dodaje barem jedno obilježje koje ga razlikuje od hiperonima (nadređenih) i od njegovih hiponima. Primjerice, *maple* nasljeđuje obilježja od svog nadređenog, *tree*, ali se razlikuje od drugih stabala po tvrdoći svog drva, obliku svog lišća, korištenju svojih sokova za sirup, itd. Ovo ustaljeno pravilo pruža središnje organizacijsko načelo za imenice u WordNet.

Meronim (engl. meronym) je dio-cjelina (ili *hasA*) semantički odnos, poznat leksičkoj semantici kao meronim/holonim. Pojam je predstavljen sa sinonimskim grozdom {*x, x', . . .*} je

¹⁷⁹antonim (anti- + grč. *ὄνυμα*: ime, po uzoru na sinonim), riječ suprotstavljena po značenju drugoj riječi iz iste vrste, protuznačnica (*dan – noć, doći – otići*). Lexička pojava značenjske suprotstavljenosti naziva se *antonimijom*. Objašnjenje preuzeto iz: antonim // Hrvatska enciklopedija, mrežno izdanje / glavni urednik Slaven Ravlić. Zagreb: Leksikografski zavod Miroslav Krleža, 2013-2015. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=3159> (12.01.2016.).

¹⁸⁰ hiponim, *lingv.* leksem koji se sadržajno može obuhvatiti u riječ širega značenja; podređenica, *opr. hiperonim*. // Hrvatski jezični portal. Dostupno na URL <http://hjp.novi-liber.hr/index.php?show=search> (12.01.2016.).

¹⁸¹ hiperonim, *lingv.* leksem koji obuhvaća leksičke sadržaje drugih leksema (hiponima); nadređenica, *opr. hiponim*. Hrvatski jezični portal. Dostupno na URL <http://hjp.novi-liber.hr/index.php?show=search> (12.01.2016.).

¹⁸² Miller, G. A., *op. cit.* (bilj. 166), str. 238.

¹⁸³ Lyons, John. *Semantics*. Cambridge ; New York: Cambridge University Press, 1977. Str. 315.

¹⁸⁴ Touretzky, David S. *The Mathematics of Inheritance Systems*. London : Pitman ; Los Altos, Calif. : Morgan Kaufmann Publishers, 1986. Str. 87.

meronim pojma predstavljenog sa sinonimskim grozdom {y, y', . . . } ukoliko izvorni govornik engleskog prihvaća rečeničnu konstrukciju oblika kao što je *A y has an x (as a part)* ili *An x is a part of y*.¹⁸⁵ Meronimski odnos je tranzitivan i asimetričan¹⁸⁶ i može se upotrijebiti za stvaranje dijela hijerarhije (s nekim uvjetima, budući meronim može imati više holonima. Ovi i drugi slični odnosi služe za organiziranje mentalnog leksikona, a predstavljaju povezivanja tako da tvore složenu mrežu, a važan dio poznavanja značenja riječi je znati gdje je riječ smještena u mreži.

Ovi odnosi, posebice hiponimija i hiperonimija, omogućuju stvaranje leksičkih hijerarhija. Dubina grananja taksonomija pojedinih leksema je vidljiva unutar njihovih ontologija¹⁸⁷, koje predstavljaju leksičku hijerarhiju unutar koje se nalazi pojedini koncept i ukazuju na njegove odnose s drugim konceptima u sustavu. Iz ontologija su očiti svi hiperonimi pojedinoga leksema, koji dosežu čak od 10 do 12 razina kod imenica da bi se došlo do vršnoga polazišta ili vršne ontologije^{188,189} s kojim je povezan svaki temeljni koncept. Vršna ontologija je najopćenitiji leksem u hijerarhiji, odnosno hiperonim koji nije hiponim nijednoga leksema. Podudarnost vršnih ontologija WordNeta za sve jezike preduvjet je za njihovo sravnjivanje, što je veoma važan čimbenik izgradnje ovoga resursa, posebice osmišljavanjem EuroWordNet-a¹⁹⁰, MultiWordNet-a¹⁹¹, Multilingual Central Repository¹⁹² i sličnih mreža. Vršne ontologije omogućavaju očuvanje zajedničke strukture, a time i kompatibilnost svih WordNet-ova. Na najnižim se razinama WordNet-ovi mogu razlikovati s obzirom na jezične i kulturološke posebitosti. Dakle, stvaranjem prinstonskoga WordNeta je stvorena velika leksička baza, odnosno leksičkosemantička mreža koja je zaintrigirala lingviste s drugih jezičnih područja.¹⁹³ Tako su se počele razvijati semantičke mreže za

¹⁸⁵ Miller, G. A., op. cit. (bilj. 166), str. 241.

¹⁸⁶ Cruse, D. A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986. Str. 115.

¹⁸⁷ Za dodatne primjere ontologija za hrvatski jezik: Bekavac, B.; Tadić, M.; Šojat, K., op. cit. (bilj. 172), str. 10.

¹⁸⁸ Miller, G. A., op. cit. (bilj. 166), str. 29.

¹⁸⁹ Nikolić-Hoyt, A. op. cit. (bilj. 178), str. 142: Za ostale je vrste riječi ontologija nešto niža, ali je svima zajedničko da se negdje na sredini ontologije nalazi koncept osnovne razine (*basic level concept*), odnosno koncept koji ima najviše prototipnih obilježja.

¹⁹⁰ Vossen, Piek. EuroWordNet: A Multilingual Database with Lexical Semantic Networks. // *Computational Linguistics* 25, 4(1998), 427-430. Dostupno na URL: <https://www.aclweb.org/anthology/J/J99/J99-4008.pdf> (3.12.2015.).

¹⁹¹ Pianta, Emanuele; Bentivogli, Luisa; Girardi, Christian. MultiWordNet: developing an aligned multilingual database. // *Proceedings of the First International Conference on Global WordNet*, 2002. Str. 21-25. Dostupno na URL: <http://multiwordnet.fbk.eu/paper/MWN-India-published.pdf> (15.06.2015.).

¹⁹² The MEANING Multilingual Central Repository. / J. Atserias...[et al.] // *In Proceedings of the Second International Global WordNet Conference (GWC'04)*. Dostupno na URL: <http://www.fi.muni.cz/gwc2004/proc/106.pdf> (27.8.2015.).

¹⁹³ Fellbaum, Christiane, op. cit. (bilj. 170), str. 137.

mnoge druge jezike, pri čemu je osnovni model bio originalni prinstonski WordNet. Kompatibilnost WordNetâ je bila nužna zbog pretpostavljenoga kasnijeg spajanja u jednu zajedničku mrežu¹⁹⁴, no, budući je njegova izvorna struktura bila prilagođena engleskome jeziku često se nailazilo na poteškoće pri njegovoj izgradnji za pojedine jezike (slavenski jezici, a među njima i hrvatski) proizašle ne samo iz jezične nego i iz kulturološke raznolikosti. Naime, neki temeljni koncepti iz prinstonskoga WordNeta nisu leksikalizirani ili su leksikalizirani ponešto drugačije u ostalim jezicima. Stoga je osnovni model izvornog prinstonski WordNet ponešto prilagođen svakome pojedinome jeziku.

U većini slučajeva pridružene su jedino su kategorije gornje razine WordNet-a. Primjerice, DBpedia je s 3.4 milijuna pojmova¹⁹⁵ povezana na WordNet; SUMO ontologija je proizvela pridruživanja između svih WordNet sinonimskih grozdova (uključujući imenice, glagole, pridjeve i priloge) i SUMO klasa; OpenCyc koristi Cyc-WordNet Linking Tool kako bi se dozvolilo korisnicima uspostavljanje linkova između WorldNet sinonimskih grozdova i Cyc-a. Isti alat omogućuje korisnicima pregledavanje WordNet-a i postojećih Cyc-WordNet linkova.¹⁹⁶ Hrvatski WordNet (CroWN) računalni je leksikon hrvatskog jezika koji se izrađuje prema postavkama izvornog WordNet-a sa Sveučilišta Princeton. Hrvatski WordNet izrađuje se u okviru projekta *Leksička semantička u izradi Hrvatskog WordNeta*¹⁹⁷, kao dijela programa *Računalnolingvistički modeli i jezične tehnologije za hrvatski jezik*¹⁹⁸. Izrada Hrvatskoga WordNeta može se podijeliti u dvije faze. Prva od njih obuhvaćala je prijevod i prilagodbu tzv. skupa temeljnih pojmova. Skup temeljnih pojmova jest skup sinonima preuzet kao dio izvornog prinstonskog WordNeta koji je poslužio kao zajednički okvir u projektima EuroWordNet I i II¹⁹⁹ i BalkaNet²⁰⁰. Ti su projekti obuhvaćali izradu Wordneta za razne jezike, ali sa zajedničkom, paralelnom strukturom. Njihova zajednička okosnica omogućuje višejezično pretraživanje sinonimskih skupova kao pojmovnih i značenjskih ekvivalenata. Skup temeljnih pojmova (dalje u tekstu BCS3) podijeljen je na tri podskupa — BCS1 korišten u EuroWordNetu I, BCS2 u EuroWordNetu II te BCS3 korišten u BalkaNetu. Iako Hrvatski

¹⁹⁴ Danas postoje mreže poput EuroWordNeta i BalkaNeta.

¹⁹⁵ What is WordNet? Dostupno na URL: <https://wordnet.princeton.edu/wordnet/> (30.10.2015.).

¹⁹⁶ OpenCyc. Dostupno na URL: <http://www.opencyc.org/doc/toc> (30.10.2015.).

¹⁹⁷ Leksička semantička u izradi Hrvatskog WordNeta. Dostupno na URL: <http://hnk.ffzg.hr/rmjt/p3html> (3.06.2016.).

¹⁹⁸ Računalnolingvistički modeli i jezične tehnologije za hrvatski jezik. Dostupno na URL: <http://rmjt.ffzg.hr> (3.06.2016.).

¹⁹⁹ EUROWORDNET: A multilingual database with lexical semantic networks / ed. by Piek Vossen. Dordrecht, Boston, London : Kluwer Academic Publishers, 1998.

²⁰⁰ Special Issue on BalkaNet. / ed. by Dan Tufiş. // Romanian Journal on Science and Technology of Information, 7 (3—4): 9—44, 2004.

WordNet ne nastaje u okviru nekog višejezičnog projekta, navedeni je skup temeljnih pojmova preveden i iskorišten kao osnovna struktura u prvoj fazi njegove izrade. Odluka je motivirana zadržavanjem određenog stupnja kompatibilnosti između njega i wordneta iz EuroWord Neta i BalkaNeta, odnosno mogućnošću višejezične pretraživosti i višestruke iskoristivosti. Druga faza projekta obuhvaća proširenje²⁰¹ Hrvatskog WordNeta, odnosno daljnju izradu sinonimskih skupova i razradu međusobnih semantičkih odnosa koji bi trebali odražavati konceptualizaciju te leksikalizaciju specifičnu za hrvatski jezik.²⁰² CroWN je povezan s inačicom 3.0. princetonskog WordNeta. CroWN 1.0. ima 10 031 sinonimski skup s ukupno 31 367 leksičkih jedinica, a CroWN 2.0. ima 23 122 sinonimska skupa s ukupno 47 906 leksičkih jedinica. CroWN 2.0. se trenutno može pretraživati na mrežnoj stranici Open Multilingual WordNet.²⁰³

3.2.8 Schema.org

Svakako se neophodno dotaknuti Schema.org, važnog projekta za razvoj semantičkog weba. Trenutno postoji mnogo standarda i shema za označavanje različitih vrsta informacija na web stranicama. Kao rezultat toga, webmasterima se teško odlučiti koji standard koristiti kao najrelevantniji i najpodržavaniji za obilježavanje. Stvaranje schema-e podržane od strane svih glavnih tražilica čini webmasterima lakšim označavanje, a tražilicama čini lakše stvaranje bogatih pretraživačkih obilježja za korisnike.²⁰⁴ Schema.org je rezultat suradnje između tražilica Bing, Google, Yahoo! i Yandex²⁰⁵ te se one na nju oslanjaju kako bi se poboljšao prikaz rezultata pretraživanja čineći ljudima lakše pronalaženje odgovarajuće web stranice te kako bi se webmasterima olakšalo osiguranje boljeg direktnog usmjerenja korisnika na njihove stranice i podatke. Ove velike tražilice će koristiti označene podatke (engl. marked-up data) s web stranica. Schema.org je njihov zajednički napor poboljšanja

²⁰¹ proširivanje (expand approach): sinonimski se skupovi iz princetonskog WordNeta prevode na određeni jezik te se preuzimaju tamo postojeće strukture leksičkih hijerarhija i međusobni semantički odnosi između njihovih članova (npr. za nizozemski i talijanski); pristup počiva na pretpostavci o bliskosti pojmovnog i leksičkog sustava (američkog) engleskog na temelju koje bi se preuzete leksičke hijerarhije i prevedeni sinonimski skupovi uz male ili nikakve preinake mogli iskoristiti za izradu vlastitog wordneta. // EUROWORDNET: A multilingual database with lexical semantic networks / ed. by Piek Vossen. Dordrecht, Boston, London : Kluwer Academic Publishers, 1998, str. 1-17.

²⁰² usp. Šojat, Krešimir. Struktura glagolskog dijela hrvatskog WordNeta // Filologija 59(2012), 154—156.

²⁰³ Hrvatski WordNet (CroWN). Dostupno na URL: <http://crown.ffzg.hr> (3.06.2016).

²⁰⁴ Shema.org. Dostupno na URL: <http://schema.org/> (01.04.2015).

²⁰⁵ About Shema.org. Dostupno na URL: <http://schema.org/docs/about.html> (01.04.2015).

weba stvaranjem strukturirane sheme za obilježavanje podataka (engl. structured data markup schema) podržane od strane glavnih tražilica. Stoga, Schema.org site osigurava zbirku schema koje webmasteri mogu koristiti za označavanje HTML stranica na načine prepoznate od strane glavnih tražilica (engl. search providers), kao i za strukturiranu interoperabilnost podataka (npr. u JSON). Mnoge stranice su generirane iz strukturiranih podataka često pohranjenih u bazama podataka. Mnoge aplikacije, posebice tražilice, mogu imati velike koristi od direktnog pristupa ovim strukturiranim podacima. Označavanje stranice omogućuje tražilicama razumijevanje informacija na web stranicama i osiguravanje bogatijih rezultata pretraživanja u svrhu olakšavanja korisnicima pronalaženja relevantnih informacija na webu. Zajednički rječnik za označavanje olakšava webmasterima donošenje odluke o shemi označavanja i dobivanje maksimalne koristi. Stvaranje sheme je nastalo kako bi se osigurala zajednička podrška webmasterima, tražilicama i korisnicima:

- webmasteri s jednog mjesta uče o označavanju umjesto zajedničkog kombiniranja schema iz različitih izvora, svaka sa svojim vlastitim pravilima i konvencijama;
- schema.org se koncentrira na definiranje vrsta stavki i svojstava koje su najvrjednije za tražilice, što znači da će tražilice dobiti strukturirane informacije koje najviše trebaju kako bi poboljšale pretraživanje;
- korisnici će u konačnici imati bolje pretraživačke rezultate na webu.

Schema.org radi na napretku koji će zadržati razvoj u narednih mnogo godina. Očekuje se razvoj iz dva glavna izvora: (1) kako se prepoznaju nove vrste strukturiranih podataka koje se mogu koristiti za poboljšanje rezultata pretraživanja, ugrađivat će se u schema.org, (2) snažno se potiču shema programeri na razvijanje i dijeljenje svoje shema. Schema.org je mrežno mjesto gdje se dokumentiraju sheme koje podupiru tri glavne tražilice. Označene stranice pomažu tražilicama razumjeti informacije na web stranicama i osigurati bogatije rezultate pretraživanja.

Cilj Schema.org je učiniti lakšim indeksiranje web stranica kako bi se poboljšala izgradnja modernih pretraživačkih servisa. Schema.org metapodaci se također mogu upotrijebiti za druge aplikacije, npr. u elektroničkim knjigama ili kao samostalni zapisi metapodataka. Cilj inicijative je pomoći tražilicama da interpretiraju informacije na web stranicama kako bi se mogle iskoristiti za poboljšanje prikazivanja rezultata pretraživanja, olakšavajući ljudima pronalaženje potrebnih informacija. Kako bi to učinili, objavljiivači sadržaja umeću strojno čitljive informacije u HTML web stranica pomažući tražilicama razumijevanje značenja teksta na ovim web stranicama. Ove informacije pomažu identificirati

naslov, ime autora, poveznicu na izdavača itd. Drugim riječima, to omogućuje čovjeku čitljivom opisu izvora njihovo stvaranje i predstavljanje kao strojno čitljivih metapodataka ili ono što Google naziva strukturiranim podacima. Schema.org ima dvije komponente: ontologiju tj. rječnik za imenovanje vrsta i karakteristika izvora, njihovih međusobnih odnosa i ograničenja o tome kako opisati te osobine i odnose. Izraz tih informacija su strojno čitljivi formati kao što su microdata, RDFa Lite i JSON-LD. Tako su npr. Schema.org metapodaci kodirani u HTML stranicama kao microdata dodajući attribute HTML elementima na web stranici.

Bilo koja tražilica može upotrebljavati Schema.org kako bi poboljšala svoje korisničko sučelje, primjerice: razlikovanje između različitih stvari s istim imenom, dopuštajući najrelevantnijim informacijama da se prikažu kao istaknutije na stranici s rezultatima pretraživanja, omogućujući filtriranje rezultata po svojstvima, kao što su cijena, dobavljač ili datuma objave, pružajući linkove na više rezultata o istom predmetu ili od istog izdavača²⁰⁶.

3.2.9 PROTON

PROTON (PROTo ONtology) je ontologija gornje razine koju je razvio Ontotext Lab²⁰⁷ u sklopu SEKT²⁰⁸ projekta te služi kao modelirajuća osnova za brojne zadatke u različitim domenama. Važan praktični pristup stvaranja ontologije je upotreba pozadine ili već postojećih znanja u obliku osnovne ontologije gornje razine. PROTON je ontologija osnovne gornje razine koja sadrži oko 300 klasa i 100 svojstava, osiguravajući pokrivenost općih pojmova potrebnih za široki raspon zadataka, uključujući semantičko označavanje, indeksiranje i pretraživanje dokumenata. Principi dizajniranja ontologije su: (1) domenski neovisna; (2) lagane logičke definicije ; (3) poravnata s popularnim standardima; (4) dobra pokrivenost imenovanih entiteta i konkretnih domena (npr. ljudi, organizacije, lokacije, brojevi, datumi, adrese).

²⁰⁶Barker, Phil; Campbell, Lorna M. What is schema.org? A Cetus Briefing Paper for LRMI Dostupno na URL: <http://publications.cetus.ac.uk/wp-content/uploads/2014/06/schemaBriefing.pdf> (1.04.2015.)

²⁰⁷ OntoText. Dostupno na URL: <http://ontotext.com/> (1.04.2015.)

²⁰⁸ SEKT. Semantically – Enabled Knowledge Technologies. Dostupno na URL: <http://www.sekt-project.com/project> (1.04.2015.)

PROTON je razvijen od KIMO ontologije²⁰⁹, koja je stvorena i upotrijebljena u okviru KIM²¹⁰ platforme za semantičko označavanje, indeksiranje i pretraživanje. PROTON ontologija je originalno kodirana u OWL Lite (nasuprot RDFS za KIMO) i organiziran u tri razine uključujući četiri modula: System, Top, Upper i KM (engl. Knowledge Management) koji je neovisan (KIMO- srodni pojmovi su premješteni na određene odvojene module); kao što je prikazano na Slika 7.

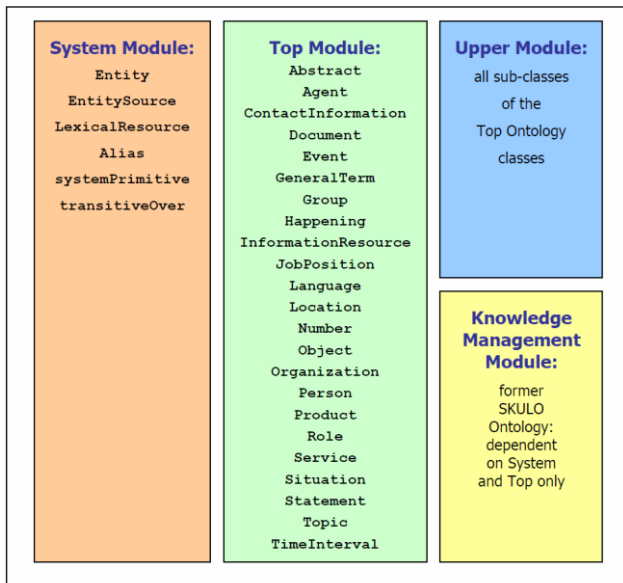
Sistemska modul (engl. System modul) je osnovni sloj, sadrži nekoliko jednostavnih meta-razina (5 klasa i 5 svojstava). Ovaj modul se može smatrati primijenjena razina ontologije koja definira nekoliko pojmova tehničke prirode bitnih za djelovanje bilo kojeg ontologijski temeljenog software-a, kao što je semantičko označavanje i pristupanje znanju alatima. System modul se referira putem "*protons:*" prefiksa. Uključuje klasu *protons:Entity*, koja je najviša („master“) klasa za bilo koju vrstu objekta ili stvari stvarnog svijeta, a može biti od interesa u nekoj disciplini ili diskursu. U system ontologiji definirano je da entiteti (npr. instance od *protons:Entity*) mogu imati višestruka imena (instance od *protons:Alias*), a te informacije o njima mogu biti ekstrahirane iz određenog *protons:EntitySource*, itd.

Vršni modul (engl. Top modul) je najveća i najopćenitija konceptualna razina, koja se sastoji od oko 20 klasa. Osigurava dobru ravnotežu iskoristivosti, neovisnosti domene, te lakoće razumijevanja i upotrebe. Gornji sloj je obično najbolja razina za uspostavljanje izravnavanja s drugim ontologijama i shemama. Općenito, Top modul se referira putem "*protont:*" prefiksa. Top ontologijski modul počinje s nekoliko osnovnih filozofskih obrazloženih razlika između tipova entiteta, kao što su: *protont:Object* - postojeći entiteti kao agenti, mjesta, prevozna sredstva; *protont:Happening* – događaji i situacije; *protont:Abstract* - apstrakcije koje nisu niti objekti ni događaji. Oni su nadalje specijalizirani u realnom svijetu, u biti stvarne vrste entiteta općenite važnosti, kao npr. sastanci, vojni sukobi, poslovne pozicije, komercijalne, vladine i druge organizacije, ljudi, i različite lokacije. Također obuhvaća brojeve, vrijeme, novac i druge određene vrijednosti. Prikazane vrste entiteta imaju svoje karakteristične atribute i odnose definirane za njih (npr. *protont:subRegionOf* svojstvo

²⁰⁹ Ontotext: KIM platform. Dostupno na URL: <http://ontotext.com/semantic-solutions/dynamic-semantic-publishing-platform/> (1.04.2015.).

²¹⁰ Prikaz KIM-a se može pronaći u radu: Semantic Annotation, Indexing, and Retrieval. / Kiryakov, Atanas...[et al.]. // Journal of Web Semantics, Vol 2, No 1 (2004). Dostupno na URL: <http://www.websemanticsjournal.org/index.php/ps/article/view/53> (10.03.2015.). Web stranica KIM platforme je: Ontotext. Dostupno na URL: <http://www.ontotext.com/kim> (10.03.2015.).

za *protont:Location*, *protont:hasPosition* za *protont:Person*, *protont:locatedIn* za *protont:Organization*, *protont:hasMember* za *protont:Group*, itd.).



Slika 7: PROTON (PROTo ONtology) moduli (preuzeto iz: (Terziev, I.; Kiryakov, A.; Manov; D., 2005.))

Gornja razina proširuje PROTON u treći sloj, gdje se mogu upotrijebiti jedna od dvije neovisne ontologije, koje definiraju specifičnije klase: PROTON Upper modul ili PROTON KM (Knowledge Management) modul. Takvi primjeri su: *protonu:Mountain*, kao specifična vrsta od *protont:Location*; *protonu:ResourceCollection* kao podklasa od *protont:InformationResource*; *protonkm:User* kao podklasa od *protont:Agent*. Imajući ovu ontologiju kao osnovu, lako se mogu dodati specifična proširenja za pojedinu domenu i u cjelini ju je jednostavno oblikovati prema specifičnim zahtjevima pojedine aplikacije, ontologije, odnosno KM alata koji će ga koristiti.

Gornji modul (engl. Upper modul) čini preko 200 općih klasa entiteta, koji se često pojavljuju u višestrukim domenama (npr. različite vrste organizacija, sveobuhvatan raspon lokacija itd.). Općenito, Upper module PROTON-a se referira putem "*protonu:*" prefiksa. KM (engl. Knowledge Management) modul sadrži 38 klasa neznatno specijaliziranih entiteta koji su specifični za tipične zadatke upravljanja znanjem i aplikacijama. KM modul je

zapravo bivša SKULO ontologija²¹¹, dalje razvijena i integrirana u PROTON. Općeniti, PROTON KM modul se referira putem "*protonkm*:" prefiksa.

Na kraju, možda je dobro ovdje spomenuti slojeve koji nisu više dio PROTON-a. Dio KIMO ontologije (prethodnik PROTON-a) bio je specifičan za KIM platformu, dakle sve ove informacije su uzete iz PROTON²¹² ontologije, te su organizirane u dva odvojena, KIM-specifična modula: KIM System Ontology i KIM Lexical Ontology. KIMO ontologija se dalje ne razvija. Najnovije verzije KIM –a (poslije 1.12) koriste PROTON, zajedno s nekoliko dopunskih KIM-specifičnih modula, kao što slijedi:

- KIMSO (engl. KIM System Ontology) – sadrži nekoliko meta – ili sustav – razina upotrebljenih od KIM-a.
- KIMLO (KIM Lexical Ontology) – sadrži neke leksičke izvore povezanih pojmova, koristeći za predstavljanje leksičkih potreba od strane KIM informacijskog ekstrakcijskog podsustava.

Objekt KIMSO i KIMLO uvoze System Modul od PROTON-a.

Kao što je istraženo u Maynard, D. et al.²¹³, unatoč razlici distribucija vrste, postoji nekoliko vrsta općih entiteta koji se pojavljuju u svim korpusa, npr. *Person*, *Location*, *Organization*, *Money (Amount)*, *Date* itd. Pravilno predstavljanje i pozicioniranje tih osnovnih vrsta bio je jedan od ciljeva i podloga stvaranja PROTON-a, a to se postiže najvećim dijelom na razini PROTON Top modularnog sloja.

Princip PROTON-a je osigurati minimalnu i dovoljnu ontologiju, pogodnu za semantičko označavanje, kao i konceptualni temelj za općenitije KM aplikacije. Njegovi prethodnici su KIMO, izvori gornje razine poput OpenCyc, Wordnet 2.0, DOLCE, EuroWordnet Top i ostali.²¹⁴ Jedan od razvojnih ciljeva PROTON-a je usklađenost s Dublin

²¹¹ Davies, J.; Boncheva, K.; Manov, D. *D5.0.1 Ontology Engineering in SEKT (informal project report)*, 2004. Dostupno na URL:

[http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20\(BULO\)%20Guidance.pdf](http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20(BULO)%20Guidance.pdf) (10.03.2015.).

²¹² PROTON. Dostupno na URL: <http://ontotext.com/proton-ontology/> (10.03.2015.).

²¹³ Multi-Source Entity recognition – an Information Extraction System for Diverse Text Types / Maynard, D. ... [et al.] Technical report CS—02—03, University of Sheffield, Department of Computer Science, 2003. Dostupno na URL: <http://gate.ac.uk/gate/doc/papers.html> (18.09.2015.).

²¹⁴ Terziev, Ivan; Kiryakov, Atanas; Manov; Dimitar. D1.8.1 Base upper-level ontology (BULO) Guidance // SEKT:Semantically Enabled Knowledge Technologies. Dostupno na URL: [http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20\(BULO\)%20Guidance.pdf](http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20(BULO)%20Guidance.pdf) (18.09.2015.).

Core, ACE označujućim vrstama²¹⁵ i ADL Feature Type Thesaurus.²¹⁶ Formalno pridruživanje odgovarajućih klasa i instanci je jednostavno i na temelju formalnih bilješke u PROTON glosama (objašnjenjima, tumačenjima, komentarima), koji ukazuju na odgovarajuća pridruživanja. Razvojna filozofija PROTON-a je njegova kompatibilnost sa svim popularnim standardima i ontologijama, poput FOAF-a.²¹⁷

3.3 Enciklopedijski ontologijski projekti

3.3.1 DBpedia

DBpedia projekt vrši ekstrakcije strukturiranih informacija iz Wikipedije s ciljem da ih se učini dostupnim na webu. Rezultat je DBpedia baza znanja, koja opisuje veliki broj entiteta (osobe, mjesta, glazbene radove, filmove, organizacije) te sadrži RDF triplete ekstrahirane iz različitih višejezičnih izdanja Wikipedije. Pokriva mnoge domene te se ažurira usporedno s Wikipedijom, višejezična je i dostupna na Webu. Opisuje 3.5 milijuna entiteta (od čega su 312 000 osoba, 413 000 mjesta, 94 000 glazbenih albuma, 140 000 organizacija, 49 000 filmova, 5 081 000 linkova na vanjske web stranice, 9 393 000 RDF linkova na druge web podatke i 841 000 linkova na fotografije)²¹⁸.

Za svaki entitet DBpedia definira jedinstveni identifikator upotrebom jedinstvenog URI-a. Upotrijebljeni su engleski nazivi članaka za stvaranje identifikatora. Izvorima se dodjeljuju URI-i u skladu sa uzorkom *http://dbpedia.org/page/Name*, gdje je *Name* preuzet iz URL-a Wikipedijinog članka, koji ima oblik *http://en.wikipedia.org/wiki/Name*. Ovo ima određene prednosti jer DBpedijini URI-i pokrivaju široki opseg enciklopedijskih tema te je njihova opsežna tekstualna definicija entiteta dostupna na poznatoj web lokaciji

²¹⁵ ACE (engl. Automatic Content Extraction) je jedan od najutjecajnijih programa informacijskog ekstrahiranja. Vidjeti Information Technology Laboratory (ITL), dostupno na URL: <http://www.itl.nist.gov/iad/894.01/tests/ace/>. Skup vrsta entiteta je definiran s ACE razvojnim planom. Vidjeti The ACE 2003 Evaluation Plan. Dostupno na URL: ftp://jaguar.ncsl.nist.gov/ace/doc/ace_evalplan-2003.v1.pdf (18.09.2015.).

²¹⁶ Alexandria Digital Library Feature Type Thesaurus. Dostupno na URL: <http://legacy.alexandria.ucsb.edu/~lhill/FeatureTypes/ver070302/> (18.09.2015.).

²¹⁷ *Friend of a Friend* (FOAF) je projekt stvaranja Weba sa strojno čitljivim stranicama (engl. machine-readable homepage) opisujući ljude, poveznice između njih, njihove radove i poslove. Vidjeti *Friend of a Friend* (FOAF). Dostupno na URL: <http://www.foaf-project.org/> (18.09.2015.).

²¹⁸ DBpedia. Dostupno na: <http://dbpedia.org/About> (18.09.2015.).

(Wikipedijinoj stranici). DBpedijini entiteti su klasificirani unutar četiri klasifikacijske sheme u svrhu ispunjavanja različitih zahtjeva primjene (Wikipedijine kategorije u SKOS prikazu, YAGO, UMBEL i DBpedia ontologija.). Projekt je načinio sljedeće doprinose razvoju weba podataka (engl. Web of Data):

- Razvijen je informacijski ekstrakcijski okvir koji konvertira sadržaj Wikipedije u bogatu višedomensku bazu znanja. Mapiranjem (pridruživanjem, preslikavanjem) Wikipedijinih predložaka infookvira²¹⁹ s ontologijom povećava se kvaliteta podataka.
- Definiran je web identifikator za svaki DBpedia entitet, što je pomoglo prevladati problem nedostatka identifikatora entiteta i postaviti osnovu za međusobno povezivanje izvora podataka na webu.
- Objavljuju se RDF linkovi koji ukazuju od DBpedije na druge izvore podataka i podržavaju nakladnike u postavljanju RDF linkova na njihove izvora podataka prema DBpediji, što je rezultiralo da DBpedia postane središte međusobnog povezivanja rastućeg *weba podataka* (engl. *Web of data*)..

DBpedia projekt također omogućuje sljedeće usluge korisnicima: postavljanje upita Wikipediji kao bazi podataka²²⁰, DBpedia kao izvor podataka za web aplikacije^{221,222}, tagiranje web sadržaja²²³ i upotreba DBpedia ontologije kao rječnika i semantičke okosnice integracije povezanih podataka²²⁴.

DBpedia ontologija je višedomenska ontologija (geografske informacije, osobe, ustanove, filmovi, muzika, knjige...). Pregled domena koje ontologija obuhvaća može se detaljnije dobiti uvidom u samu taksonomijsku strukturu DBpedia ontologije²²⁵. Izrađena je ručno na temelju najčešće korištenih predložaka infookvira unutar Wikipedije. Trenutno ima

²¹⁹ Infookviri su standardizirane stranice koje sadrže osnovne informacije o entitetu opisanom u članku.

Primjerice, infookvir za zemlje sadrži izvorno ime zemlje, njezin glavni grad i veličinu.

²²⁰ O tome podrobnije u: Faceted Wikipedia Search / Hahn, Rasmus...[et al.]. Dostupno na URL:

<http://www4.wiwiss.fu-berlin.de/bizer/pub/hahn-et-al-faceted-wikipedia-search-BIS2010.pdf> (18.09.2015.).

²²¹ O tome podrobnije u: Becker, Christian; Christian Bizer. DBpedia Mobile: A Location-Enabled Linked Data Browser. Dostupno na URL: <http://www4.wiwiss.fu-berlin.de/bizer/pub/BeckerBizer-DBpediaMobile-Ldow2008.pdf> (18.09.2015.).

²²² O tome podrobnije u: Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections / Kobilarov, Georgi...[et al.]. Dostupno na URL:

<http://www.georgikobilarov.com/publications/2009/eswc2009-bbc-dbpeda.pdf> (18.09.2015.).

²²³ O tome podrobnije u: Preliminary Results in Tag Disambiguation using DBpedia / Garcia, A. ...[et al.]. Dostupno na URL: <http://eprints.ecs.soton.ac.uk/17792/1/CKCaR09-final.pdf> (18.09.2015.).

²²⁴ Bizer, Christian ...et al. DBpedia - A crystallization point for the Web of Data // Web Semantics : Science, Services and Agents on the World Wide Web 7, 3(2009), 154-165. Dostupno na URL: http://jens-lehmann.org/files/2009/dbpedia_jws.pdf (3.06.2016.).

²²⁵ Taksonomijski pregled ontologije je dostupan na URL:

<http://www4.wiwiss.fu-berlin.de/dbpedia/dev/ontology.htm> (18.09.2015.).

više od 259 klasa u hijerarhijskoj formi opisanoj s 1.200 različitih svojstava. Mapiranjem se rješavaju slabosti Wikipedijinog sustava infookvira: različiti infookviri pridružuju se istoj klasi (trenutno je 1.055 Wikipedijinih predložaka pridruženo na 259 ontologijskih klasa), upotrebljavaju se različiti nazivi svojstava za isto svojstvo (trenutno je 3 690 infookvirnih svojstava mapirano na 1.200 ontologijskih svojstva) i ne postojanje jasno definiranih vrsta podataka (eng. datatypes) za vrijednosti svojstava²²⁶.

DBpedijin ekstrakcijski okvir trenutno sadrži 11 ekstraktora koji obrađuju sljedeće vrste Wikipedijinog sadržaja²²⁷ te su nakon ekstrakcije dostupni na webu (npr. entitet Split:

http://dbpedia.org/page/Split_%28city%29):

- *Labels*. Svi Wikipedijini članci imaju naslove, koji su upotrijebljeni kao *rdfs:label* za odgovarajući DBpedia izvor.
- *Abstracts*. Ekstrahira se kratki sažetak (prvi paragraf predstavljen upotrebom *rdfs:comment*) i dugi sažetak (tekst prije tablice sadržaja, uglavnom 500 riječi, upotrebom svojstva *dbpedia:abstract*) iz svakog članka.
- *Interlanguage links*. Ekstrahiraju se linkovi koji povezuju članke o istoj temi u različitim jezičnim izdanjima Wikipedije i upotrebljavaju se za dodjeljivanje *labels* and *abstracts* na različitim jezicima DBpedia izvoru.
- *Images*. Linkovi ukazuju na izvor Wikimedia Commons opise slika koji se izvlači i predstavljen je upotrebom the *foaf:depiction* svojstva.
- *Redirects*. U svrhu identificiranja sinonima, Wikipedijini članci osiguravaju preusmjerenja na odgovarajući članak. Ova preusmjerenja se ekstrahiraju i koriste za upućivanje između DBpedia izvora.
- *Disambiguation*. Wikipedijine razdvojbene stranice objašnjavaju različita značenja homonima. Ekstrahiraju se i predstavljaju razdvojbene linkovi upotrebom svojstva *dbpedia:disambiguates*.
- *External links*. Članci sadrže reference na vanjske web stranice koje se predstavljaju upotrebom DBpedia svojstva *dbpedia:reference*.
- *Page links*. Ekstrahiraju se svi linkovi između Wikipedijinih članaka i predstavljaju se upotrebom svojstva *dbpedia:wikilink property*.

²²⁶ DBpedia ontology. Dostupno na URL: <http://wiki.dbpedia.org/Ontology?v=zj4> (18.09.2015.).

²²⁷ Bizer, Christian et al., loc. cit.

- *Homepages*. Ovim ekstraktorom dobivaju se linkovi na homepage stranice entiteta kao što su kompanije i organizacije tražeći termine kao što su homepage ili website unutar linkova članka (predstavljani upotrebom *foaf:homepage*).
- *Categories*. Članci Wikipedije su razvrstani u kategorije, koje su predstavljene upotrebom SKOS rječnika²²⁸. Kategorije postaju *skos:concepts*; a kategorijske relacije su predstavljene upotrebom *skos:broader*.
- *Geo-coordinates*. Geo-ekstraktor izražava koordinate upotrebom Basic Geo (WGS84 lat/long) rječnika²²⁹ i GeoRSS Simple kodiranjem W3C Geospatial rječnika²³⁰.

Najvrjednija vrsta ekstrahiranih informacija za DBpediu su zasigurno Wikipedijini infookviri unutar članaka. Wikipedijin sustav predložaka infookvira se razvija tijekom vremena bez centralne koordinacije. Infookviri predstavljaju najbogatije semantičke strukture jer sadrže najvažnije činjenice o članku. No, pojavljuje se sljedeće nedosljednosti: različite zajednice upotrebljavaju različite predloške da bi opisale iste stvari (npr. *infobox_city_japan*, *infobox_swiss_town* and *infobox_town_de*), različiti predlošci koriste različite nazive za ista svojstva (npr. *birthplace* i *placeofbirth*), mnogi Wikipedijini urednici ne slijede strogo preporuke dane na stranici koja opisuje predložak te su vrijednosti svojstava izražene upotrebom širokog dosega različitih formata i mjernih jedinica. Postupak mapiranja provodi se upravo zbog prevladavanja problema sinonimnih naziva svojstava i višestrukih predložaka te ekstrakcije informacija u formi RDF tripleta. Postupkom mapiranja pridruženo je 36 predložaka infookvira hrvatskog izdanja Wikipedije na 32 klase DBpedia ontologije. Primjer višestruko pridruženih infookvira na jednu klasu izvršen je u postupku kojim su npr. infookvir austrijski muzej ili galerija, infookvir hrvatski muzej ili galerija i infookvir londonski muzej ili galerija pridruženi ontologijskoj klasi *Museum*. Također je pridruženo 188 infookvirnih svojstva na 173 ontologijska svojstva DBpedije. Uočljivo je pridruživanje svojstava s istim značenjem iz različitih infookvira na jedno ontologijsko svojstvo (npr. infookvirna svojstva *URL*, *web*, *webstranice* koje navode web adresu unutar različitih predložaka infookvira pridružena su na jedno ontologijsko svojstvo *foaf:homepage*). Odabrani su predlošci iz

²²⁸ SKOS: Simple Knowledge Organization System. Dostupno na URL: <http://www.w3.org/2004/02/skos/> (18.09.2015.).

²²⁹ Basic Geo Vocabulary. Dostupno na URL: <http://www.w3.org/2003/01/geo/> (18.09.2015.).

²³⁰ W3C Geospatial Vocabulary. Dostupno na URL: <http://www.w3.org/2005/Incubator/geo/XGR-geo/> (18.09.2015.).

domene književnosti, filma, glazbenika, knjiga, medija, kulture itd. Mapiranje se vrši upotrebom *DBpedia Mapping Language*²³¹.

U svrhu ostvarivanja ekstrakcije informacija iz infookvira hrvatskog izdanja Wikipedije bilo je potrebno uspostaviti osnovni preduvjet tj. otvoriti wiki namespaces *Mappings_hr*²³², što je DBpedijin tim prema uspostavljenom zahtjevu i omogućio. Time je Hrvatska postala četvrta zemlja uključena u postupak mapiranja infookvira u DBpedia projektu (prethodila su joj mapiranja za englesko, njemačko i grčko izdanje Wikipedije)²³³. Svaki infookvir u osnovi sadrži popis parova *svojstvo: vrijednosti* (Slika 8). DBpedia projekt ekstrahira upravo ove infookvirne parove *svojstvo: vrijednost* iz predložaka i konvertira ih u RDF triplete te se u tom slučaju npr. iz članka o Splitu dobije određeni broj RDF tripleta koji opisuju entitet Split (Slika 9). Generiraju se URI identifikatori: za subjekt (<http://dbpedia.org/resource/> kao prefiks) i predikat tripleta te objekt koji je URI ili odgovarajuća vrijednost iz infookvira.



Država	 Hrvatska
Županija	 Splitsko-dalmatinska
Površina	
- ukupna	79,33 km ²
- gradskog naselja	63 km ²
Stanovništvo (2001. (samo admin. granice grada))	
- ukupno	188.694 stan.
- ukupna gustoća	2995 stan./km ²
- gradsko naselje	188.694 stan.
- Metropolitansko područje - aglomeracija	408.251 stan.
Gradonačelnik	Željko Kerum
Gradska naselja	8 naselja
Dan grada	7. svibnja
Zaštitnik	Sv. Dujam
Poštanski broj	21 000
Pozivni broj	+385 021
Autooznaka	ST
Službena stranica	www.split.hr 

Slika 8: Infookvir iz Wikipedijinog članka Split (preuzeto iz: (Sarić, I.; Essert, M., 2011.))

²³¹ DBpedia Mappings. Dostupno na URL: http://mappings.dbpedia.org/index.php/Main_Page#Tutorials (18.09.2015.).

²³² Mapping hr namespace. Dostupno na URL:

<http://mappings.dbpedia.org/index.php?title=Special%3AAllPages&from=&to=&namespace=284> (18.09.2015.).

²³³ Detaljnije o postupku uključivanja Hrvatske u DBpedia projekt može se pročitati u Sarić, Ivana; Essert, Mario. Uključivanje hrvatskog Web sadržaja u DBpediju // 14. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastructure : zbornik radova / uredili Sanjica Faletar Tanacković i Damir Hasenay. Zagreb : Hrvatsko knjižničarsko društvo, 2011. Str. 165-179.

Subject	Predicate	Object	Triple Provenance
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/2000/01/rdf-schema#label	Split	http://hr.wikipedia.org/wiki/ISplit#
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Thing	http://hr.wikipedia.org/wiki/ISplit#absolute-line=1
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Place	http://hr.wikipedia.org/wiki/ISplit#absolute-line=1
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/PopulatedPlace	http://hr.wikipedia.org/wiki/ISplit#absolute-line=1
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Settlement	http://hr.wikipedia.org/wiki/ISplit#absolute-line=1
http://dbpedia.org/resource/ISplit_%28city%29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/City	http://hr.wikipedia.org/wiki/ISplit#absolute-line=1
http://dbpedia.org/resource/ISplit_%28city%29	http://xmlns.com/foaf/0.1/name	Split	http://hr.wikipedia.org/wiki/ISplit#absolute-line=2
http://dbpedia.org/resource/ISplit_%28city%29	http://dbpedia.org/ontology/areaTotal	7.933E7	http://hr.wikipedia.org/wiki/ISplit#absolute-line=9
http://dbpedia.org/resource/ISplit_%28city%29	http://dbpedia.org/ontology/areaTotal	6.3E7	http://hr.wikipedia.org/wiki/ISplit#absolute-line=11
http://dbpedia.org/resource/ISplit_%28city%29	http://dbpedia.org/ontology/populationTotal	188694	http://hr.wikipedia.org/wiki/ISplit#absolute-line=16
http://dbpedia.org/resource/ISplit_%28city%29	http://dbpedia.org/ontology/postalCode	21 000	http://hr.wikipedia.org/wiki/ISplit#absolute-line=34
http://dbpedia.org/resource/ISplit_%28city%29	http://xmlns.com/foaf/0.1/homepage	http://www.split.hr	http://hr.wikipedia.org/wiki/ISplit#absolute-line=37
http://dbpedia.org/resource/ISplit_%28city%29	http://dbpedia.org/ontology/leaderName	http://dbpedia.org/resource/%C5%BDeljko_Kerum	http://hr.wikipedia.org/wiki/ISplit#absolute-line=24

Slika 9: RDF tripleti ekstrahirani iz Infookvira Split (preuzeto iz: (Sarić, I.; Essert, M., 2011.))

Pridruživanjem određenog infookvirnog predloška (npr. Grad) na određenu ontologijsku klasu (npr. City) ekstrahirat će se RDF tripleti svih infookvirnih parova *svojstvo:vrijednost* postojećih članaka za sve gradove koji u Wikipediji imaju članak s ovim infookvirnim predloškom. Rezultat načinjenog postupka mapiranja su ekstrahirani RDF tripleti koji opisuju trenutni broj prisutnih entiteta u hrvatskom izdanju Wikipedije: npr. 2000 glazbenika, 500 glazbenih instrumenata, 10 muzeja, 2000 filmskih umjetnika, 56 glazbenih žanrova, 19 umjetničkih kipova, 215 likovnih umjetnika, 267 papa, 163 skladatelja, 19 svećenika, 7 web stranica, 13 astronauta, 2585 muzičkih albuma, 171 biografija, 46 filozofa, 196 knjiga, 58 slika, 89 televizijskih epizoda, 118 znanstvenika, 1159 filma, 9 glazbenih festivala, 132 hrvatska grada, 117 jezika, 443 književnika, 232 monarha, 653 singla, 30 svetaca, 141 vojne osobe, 45 časopisa²³⁴. DBpedia baza znanja i web podataka oko DBpedia položili su temelj za širok raspon primjena te su upravo na tim temeljima izrasle aplikacije u kojima je DBpedia upotrebljena kao međusobno povezano središte za pregledavanje i istraživanje weba podataka (DBpedia Mobile), aplikacije koje upotrebljavaju DBpediju za odgovaranje na složena pitanja (DBpedia Query Builder, Relationship Finder) i aplikacije

²³⁴ Navedeni broj Wikipedijinih entiteta opisanih RDF tripletima je ustanovljen 17.12.2015. Stalnim rastom Wikipedijinih članaka automatski će se povećavati i navedeni broj entiteta opisanih RDF tripletima koji u člancima sadrže mapirane infookvirne predloške i svojstva.

koje upotrebljavaju identifikatore DBpedijskih entiteta za označavanje sadržaja weba (Muddy Boots, Open Calais, Faviki, LODr).²³⁵

3.3.2 Freebase

Freebase²³⁶ je suradnički izgrađena i uređena baza znanja općih strukturiranih podataka kojoj pridonose korisnici prikupljajući podatke iz različitih izvora, kao što su Wikipedia i MusicBranz²³⁷. Namijenjena je širokoj publici te predstavlja sustav baze podataka stvoren kao otvoreni web repozitorij svjetskog znanja, a njezino stvaranje je motivirano širokom upotrebom informacijskih zajednica, kao što su Semantic Web i Wikipedia.²³⁸ Sadrži objekte gotovo svih vrsta (filmovi, knjige, slavne osobe, lokacije, kompanije itd) te broji 47 560 817 tema i 2 903 361 537 činjenica.²³⁹ Osiguran je skupa alata za pojednostavljeno uređivanje, objavljivanje i pretraživanje podataka. Podaci se mogu pretraživati u RDF-u ili JSON-u te je osiguran lako upotrebljivi MQL upitni jezik (engl. Metaweb Query Language) u svrhu stvaranja suradničkih podatkovno usmjerenih web aplikacija. Navedeno omogućuje suradničko stvaranje, strukturiranje i upravljanje podacima u Freebase-u. Freebase sustav je načinjen tako da se može jednostavno i lako pridruživati na ontologijske strukture drugih sustava, sastojeći se od sljedećih vrsta objekata: *tema*, *literal*, *vrsta*, *svojstvo* i *shema*.

Tema (engl. topic) je objekt koji predstavlja entitet. Tema može biti specifična i konkretna (npr. Bill Clinton, The City of Vancouver, Canada) ili apstraktan koncept (npr. the number PI, Zoroastrianism). Svakoj temi je dan točno jedan globalni jedinstveni identifikator (engl. globally unique identifier, GUID) koji upućuje samo na tu temu. Tema može imati mnogo imena i upotrebljavati se u mnogo različitih konteksta, ali svaka tema treba predstavljati jedan i samo jedan pojam ili entitet u stvarnom svijetu.

Literal u Freebase je jednostavan string, numerička vrijednost, boolean, ili timestamp. Sve ostale vrste podataka su izgrađene upotrebom jasne strukture.

Vrsta (engl. type) je objekt koji se koristi za semantičko grupiranje tema. Tema povezana s vrstom se smatra instancom te vrste. Primjeri vrsta su Filmski glumac, Osoba,

²³⁵ Bizer, Christian ... et al., op. cit. (bilj. 224).

²³⁶ Freebase API će potpuno biti ugašen 31. 08.2016., a do tada se mogu skinuti 22 GB informacija u RDF-u. // Freebase API. Dostupno na URL: <https://developers.google.com/freebase/> (3.06.2016.):

²³⁷ MusicBranz. Dostupno na URL: <https://musicbrainz.org/> (12.01.2016.).

²³⁸ A Platform for Scalable, Collaborative, Structured Information Integration / Kurt Bollacker... [et al.] Dostupno na URL: <http://aaai.org/Papers/Workshops/2007/WS-07-14/WS07-14-004.pdf> (18.09.2015.).

²³⁹ Freebase. Dostupno na URL: <https://www.freebase.com/> (17.06.2015.).

Zračna luka, Programski jezik itd. Teme mogu imati višestruke vrste koje im se tijekom vremena mogu dodati ili ukloniti. Poput tema, vrsta predstavlja točno jedno značenje, čak i ako ima više imena i upotreba. Ne postoji zamisao hijerarhije vrsta u Freebase-u, nego su vrste stvorene i "izmiješane" prema potrebi dobivajući različita svojstva. Na primjer, "Filmski glumac" ne nasljeđuje svojstva od "Osoba" jer neki glumci ne moraju biti čovjek (npr. psi, roboti), ali imaju ista svojstva (npr. filmovi, filmske nagrade) kao ljudski glumci.

Svojstvo može biti zamišljeno kao atribut teme u Freebase-u. Svojstva mogu biti literali ili odnosi na drugu temu.

Shema: svaka vrsta ima zbirku od nijednog ili više svojstava, poznatu kao shema te vrste. Ukoliko je tema instanca vrste, svojstva u shemi te vrste su odgovarajuća za primjenu na tu temu.

Vrste i svojstva može stvoriti bilo koji korisnik kako bi dodao strukturirane podatke na teme u Freebase-u. Kako svi podaci u Freebase-u imaju sitno zrnate atribucije, jednostavno se mogu filtrirati rezultati upita prema verziji stvarnosti na koju se želi obratiti pozornost. Očekuje se kako će svaki korisnik raditi samo koliko je potrebno za njegove vlastite potrebe, ali će se kolektivno tijekom vremena rezultat povećati i poboljšati strukturu podataka.

3.3.3 YAGO

Yago²⁴⁰ je velika semantička baza znanja proizašla automatskom ekstrakcijom činjenica iz Wikipedije²⁴¹, WordNet²⁴²-a i GeoNames²⁴³. Nastao je kao potreba stvaranja veće ontologije koristeći trenutno postojeće ontologije s glavnim ciljem ujedinjenja Wikipedije i WordNet-a te iskoristivosti bogatih struktura i informacije osiguravajući vjerodostojnost činjenica. Sadrži više od 10 milijuna entiteta (film, osobe, organizacije, gradovi, itd) i poznaje 120 milijuna činjenice o njima (u kojim filmovima su glumile pojedine osobe, koji grad je smješten u kojoj državi).²⁴⁴ Yago je velika ontologija s visokom pokrivenosti područja, preciznosti i kvalitete. Uključuje taksonomijsku IS-A hijerarhiju i semantičke odnose između entiteta. Oslanja se na oko 100 ručno definiranih odnosa. Yago ekstrahira činjenice iz

²⁴⁰ YAGO: A High-Quality Knowledge Base. Dostupno na URL: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/> (3.06.2016.).

²⁴¹ Wikipedia : The Free Encyclopedia. Dostupno na URL: http://en.wikipedia.org/wiki/Main_Page (17.06.2015.).

²⁴² Wordnet : A lexical database for English. Dostupno na URL: <http://wordnet.princeton.edu/> (17.06.2015.).

²⁴³ GeoNames. Dostupno na URL: <http://www.geonames.org/> (17.06.2015.).

²⁴⁴ D5: Databases and Information Systems: FaQ. Dostupno na URL: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/faq/> (17.06.2015.).

kategorija i infookvira Wikipedije u kombinaciji s taksonomskim odnosima iz WordNet-a. Osnovu Yago ontologije čini Wikipedia, jedna od najopsežnijih dostupnih enciklopedija. Infookviri se puno lakše parsiraju i iskorištavaju od teksta napisanog prirodnim jezikom²⁴⁵. Kategorijske stranice su popisi članaka koji pripadaju određenoj kategoriji (npr. Elvis je u kategoriji američki rock pjevači). Ovi popisi daju kandidate za entitete (npr. Elvis), kandidate za pojmove (npr. IsA (Elvis, rockSinger))²⁴⁶ i kandidate za odnose (npr. nationality(Elvis, American)).

Wikipedijske kategorije su raspoređene u hijerarhiju, ali ta hijerarhija je jedva korisna za ontologijske svrhe. Primjer, Elvis je u nadkategoriji Grammy Awards, ali Elvis je dobitnik Grammy Award nagrade, a ne Grammy Award. Nasuprot tome, WordNet osigurava čistu i pomno sastavljenu hijerarhiju pojmova. Očito, Wikipedijini pojmovi nemaju duplikate u WordNet-u, stoga je Yago projekt prvi pristup koji predstavlja tehnike povezivanja i ujedinjenja ta dva izvora s visokom preciznošću. YAGO ontologija kroz ujedinjenje ova dva izvora ostvaruje dobit iz velike količine individua i činjenica izvedenih iz Wikipedije te iskorištavanje čiste taksonomije pojmova iz WordNet-a. Yago je poseban na više načina²⁴⁷:

1. Točnost Yago-a je ručno vrednovana, što je osiguralo potvrdu točnosti od 95%. ;
2. Yago kombinira čistu taksonomiju WordNet-a s bogatstvom Wikipedijinog kategorijskog sustava, dodjeljujući entitete u više od 350.000 klasa;
3. Yago je ontologija koja dodjeljuje vremensku i prostornu dimenziju mnogim svojim činjenicama i entitetima;
4. Uz taksonomiju, Yago ima tematska područja (domene) kao što su glazba ili znanost iz WordNet domena.²⁴⁸

YAGO se temelji na podatkovnom modelu entiteta i binarnih odnosa. No, uz pomoć konkretizacije (odnosno uvođenjem identifikatora za odnos) izražavaju se odnosi između činjenica (npr. koje činjenice su pronađene na kojoj Web stranici), n- ti odnosi (npr. Elvis je osvojio Grammy Award 1967. godine) i opća svojstva odnosa (npr. tranzitivnost). YAGO su

²⁴⁵ Dbpedia: A nucleus for a web of open data / Sören Auer...[et al.] // The Semantic Web - ISWC 2007 + ASWC 2007, 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference, volume 4825, 2007, 715-728. Dostupno na URL: <http://iswc2007.semanticweb.org/papers/715.pdf> (17.06.2015.).

²⁴⁶ Kinzler, Daniel. Wikisense - mining the wiki // In Proceedings of Wikimania 2005. The First International Wikimedia Conference, 2005.

²⁴⁷ YAGO, op. cit. (bilj. 240.).

²⁴⁸ WordNet Domains. Dostupno na URL: <http://wndomains.fbk.eu/> (17.06.2015.).

prvo predstavili F. M. Suchanek, G. Kasneci, i G. Weikum²⁴⁹ te je iz članka vidljivo znakovito proširenje dosadašnjeg rada iskorištavanja infookvira, uvođenje tehnika kvalitetne kontrole i definiranja novog upitnog jezika. YAGO je zamišljen kao detaljan i izražajan model podataka jer je sposoban izraziti entitete, činjenice, odnose između činjenica i svojstva odnosa. To je razlog zbog kojeg je YAGO model izražen kroz blaga proširenja RDFS-a.²⁵⁰ Svi objekti (npr. gradovi, ljudi, događaji, čak i URL-ovi) su predstavljeni kao entiteti u Yago modelu. Entiteti su apstraktni ontološki objekti, koji su u idealnom slučaju jezično neovisni. Jezik koristi riječi kako bi uputio na ove entitete. U Yago modelu, riječi su također entiteti, što čini mogućim izraziti kako određena riječ upućuje na određeni entitet (npr. "Einstein" znači AlbertEinstein), što omogućava bavljenje sinonimijom i dvosmislenošću. Tako, sljedeća izreka navodi kako se "Einstein" može odnositi i na muzikologa Alfred Einstein-a, tj. "Einstein" znači AlfredEinstein. Koriste se navodnici kako bi se razlikovale riječi od drugih entiteta. Entiteti su grupirani u klase, npr. klasa Riječ obuhvaća sve riječi. Svaki entitet je instanca najmanje jedne klase, što se izražava s *type* odnosom (npr. AlbertEinstein type physicist). Klase su također entiteti. Klase su uređeni u taksonomijsku hijerarhiju, koja je izražena sa *subClassOf* odnosom (npr. physicist subClassOf scientist).

Dva entiteta mogu stajati u odnosu. Brojevi, datumi, stringovi i ostali literali su također predstavljeni kao entiteti, što znači da oni mogu stajati u odnosu s drugim entitetima. U YAGO modelu odnosi su također entiteti. Ovo čini mogućim predstaviti svojstva odnosa (poput tranzitivnosti ili odnos podrazumijevanja- engl. subsumption) unutar modela. Primjer *subclassOf type transitiveRelation* ukazuje da je subClassOf odnos tranzitivan čineći ga instancom klase transitiveRelation.

Trojka entitet – odnos – entitet se naziva činjenicom. Entiteti se nazivaju činjenični argumenti. Svakoj činjenici je dan identifikator činjenice. Kao RDFS, YAGO model također razmatra činjenične identifikatore kao entitete, što omogućuje predstavljanje određenih činjenica na određenom URL. Primjerice, pretpostavimo kako prethodna činjenica (Albert Einstein, bornInYear, 1879) ima činjenični identifikator #1, tada bi izreka #1 foundIn [http : //www.wikipedia.org/Einstein](http://www.wikipedia.org/Einstein) trebala kazati kako je ova činjenica pronađena u Wikipediji. YAGO ontologija osigurava da je svaki činjenični identifikator pridružen na

²⁴⁹ Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard. Yago: A Core of Semantic Knowledge Unifying Wordnet and Wikipedia. Dostupno na URL: . <http://www2007.org/papers/paper391.pdf> (17.06.2015.).

²⁵⁰ Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard. Yago: A Large Ontology from Wikipedia and WordNet. // Web Semantics 6, 3(2008), 203-217. Dostupno na URL: . <http://dx.doi.org/10.1016/j.websem.2008.06.001> (17.06.2015.).

točno jednu činjenicu. Neke činjenice zahtijevaju više od dva argumenta (primjerice činjenica Einstein won the Nobel Prize in 1921).

Yago model nudi jednostavno rješenje za ovaj problem, a ono se temelji na pretpostavci da se za svaki n-ti odnos, primarni par njegovih argumenata može biti identificiran. Primjerice, za gornju won-prize-in-year-relation, par osoba i nagrada se može smatrati primarnim parom, koji se može predstaviti kao binarna činjenica s činjeničnim identifikatorom:

#1 : AlbertEinstein hasWonPrize NobelPrize

Svi ostali argumenti se mogu predstaviti kao odnosi koji su sadržani između primarnog para i drugih argumenata: #2 : #1 time 1921²⁵¹.

3.3.4 YAGO2

YAGO2 je novo izdanje YAGO baze znanja s naglaskom na vremensko i prostorno znanje. Razvijen je kao prošireni pristup za ekstrakciju činjenica integrirajući vremenske i prostorne semantičke informacije Wikipedije, WordNet-a, i GeoNames-a. Sadrži gotovo 10 milijuna entiteta i događaja, 80 milijuna činjenica te 76 milijuna ključnih riječ koje predstavljaju opće svjetsko znanje.²⁵² Poboľšan prikaz podataka uvodi vrijeme i mjesto kako bi se bogatstvo prostorno-vremenske informacije u YAGO-u moglo istražiti grafički ili putem posebnog prostorno-vremenski svjesnog upitnog jezika.

U tu svrhu prošireno je klasično subjekt – predikat – objekt predstavljanje znanja s tri dimenzije tj. zemljopisnom, vremenskom i ključnom riječju kako bi se osigurao prikladan načina pregledavanja i postavljanja upita novih podataka. Korisnici mogu vizualizirati događaje i entitete na karti i sinkroniziranoj vremenskoj traci s mogućnošću zumiranja i pomicanje vremenske i geografske dimenzije. Uvedena su dva opća entitet-vrijeme odnosa: *startsExistingOnDate* i *endsExistingOnDate* za definiranje početne i završne vremenske točke entiteta. Odnosi određene vrste *wasBornOnDate*, *diedOnDate* itd. čine pododnose prethodnih dvaju općih odnosa, tako da svi vremenski ovisni entiteti imaju opći raspon postojanosti.

²⁵¹ Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard, loc. cit. (bilj. 249).

²⁵² YAGO2. Dostupno na URL: <http://meta-guide.com/data-processing/knowledge-synthesis/yago2> (04.01.2016.).

Stvoreni su vremenski rasponi (engl. time span) za sve entitete gdje je moguće izvući takvu vrstu informacije iz Wikipedije.

Činjenice, također, mogu imati vremensku dimenziju. YAGO2 ekstraktori mogu pronaći činjenični vremenski događaj iz Wikipedijinih okvira, kao što je npr. supružnici koji se često spominju s datumom sklapanja braka i razvoda. Ekstraktori konkretiziraju vremenski ovisne činjenice dajući im identifikator (kao što je npr. #42) te zatim stvaraju dvije meta-činjenice #42 occursSince date i #42 occursUntil date, što definira početnu i završnu vremensku točku trajanja nekog događaja.

U nekim slučajevima, entiteti koji se pojavljuju u činjenicama mogu ukazivati na njezin vremenski događaj. Na primjer, za činjenicu BobDylan wasBornIn Duluth čini se prirodnim koristiti Dylan-ov datum rođenja kao vrijeme događaja. U tu svrhu, stvoreni su odnosi koji ukazuju na stvaranje svog subjekta (kao što je wasBornIn), odnosno instance nove klase *subjectStartRelation*. Isto tako, načinjeni su svi odnosi koji ukazuju na stvaranje svog objekta (kao što je created), odnosno instance nove klase *objectStartRelation*. Uvedeno je implikacijsko pravilo od početaka vremenske postojanosti entiteta do vremenskog događaja činjenice, kojim YAGO2 okvir dodjeljuje početak i / ili kraj vremenskog razdoblja (engl. time spans) na sve entitete, činjenice i događaje, ukoliko isti imaju poznatu početnu ili završnu točku. Ukoliko nema takvih vremenskih točaka, možda se mogu zaključiti iz baze znanja. Vremenske točke su obično označene s raščlanjivanjem na dane, ali ponekad i s grubljim raščlanjivanjem kao što su godine.

Svi fizički objekti imaju lokaciju u prostoru te se zato YAGO2 bavi entitetima koji posjeduju trajnu prostornu površinu na Zemlji (npr. zemlje, gradovi, planine, rijeke). Takvi entiteti u izvornom Yago (i WordNet-u) nemaju zajedničku nad-klasu, stoga YAGO2 uvodi novu klasu *yagoGeoEntity* koja zajedno grupira sve geo entitete, tj. sve entitete sa stalnim fizičkim mjestom na Zemlji. Položaj geo-entiteta se može opisati zemljopisnim koordinatama. Uvedena je posebna vrsta podataka za pohranu geografskih koordinata, *yagoGeoCoordinates*, te je svaki geo entitet povezan na svoje koordinate s *hasGeoCoordinates* odnosom.²⁵³ YAGO2 harvestira (pobire) geo entitete iz dva izvora. Prvi je Wikipedia, koja sadrži veliki broj mjesta, regija, planina, rijeka, jezera itd., koji dolaze s povezanim geografskim koordinatama. Ekstrakcijski okvir dohvaća koordinate za 176.474 geo entiteta. Međutim, nisu svi geo entiteti u Wikipediji označeni sa zemljopisnim koordinatama, a tu su i mnogi geo

²⁵³ Dodjeljivanje koordinata slijedi pravila iz Wikipedia. Vidjeti WikiProject Geographical coordinates. Dostupno na URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Geographical_coordinates (17.06.2015.).

entiteti koji uopće nisu obuhvaćeni Wikipedijom. Stoga se iskorištava još bogatiji izvor geografskih podataka GeoNames, koji opisuje više od 7 milijuna mjesta.²⁵⁴ GeoNames svrstava mjesta u površnu kategorijsku strukturu, a svakom mjestu se dodjeljuje samo jedna klasa, npr. Berlin je "glavni grad političkog entiteta". Nadalje, GeoNames sadrži *locatedIn* hijerarhiju, npr. Berlin se nalazi u Njemačkoj, koja je u Europi. Ove *locatedIn* činjenice oblikuju predodžbu jednog entiteta sadržanome unutar drugog, što nije moguće samo sa zemljopisnim koordinatama. GeoNames također osigurava alternative nazive za svaku lokaciju.

Svi ti podaci vrijedan su dodatak YAGO-u, stoga je načinjen trud njihovog integriranja u potpunosti. Združili su se pojedinačni zemljopisni entiteti koji postoje i u Wikipediji i u GeoNames kako bi se izbjegla dupliranja, uzimajući u obzir ne samo tekstualnu sličnost između imena entiteta, nego i zemljopisne koordinate gdje su takve informacije dostupne. Osim toga, GeoNames dodjeljuje klasu svakoj lokaciji. Kako bi se izbjeglo dupliciranje klasa, potrebno ih je združiti na postojeće YAGO klase, a prethodnim radom su već poravnate GeoNames klase s WordNet-om u tzv. GeoWordNet.²⁵⁵ Proces združivanja obogaćuje YAGO2 s gotovo 7 milijuna zemljopisnih entiteta i gotovo 50 milijuna novih činjenica iz GeoNames-a, posebno dodajući zemljopisne koordinate koje se ne mogu izvući iz Wikipedije, što uzvraća dostupnošću većeg broja entiteta prostornim upitima. Prostorna dimenzija se rješava tako da se položaj dodjeljuje i entitetima i činjenicama, gdje je to ontološki logično i može se zaključiti iz podataka. Položaj činjenica i entiteta je zadan s geo entitetom, koji može biti ekstrahiran iz Wikipedije, npr., lokacija od Woodstock je White Lake, NY, koja je instanca od *yagoGeoEntity*.

U bazi znanja postoje prostorni podaci za događaje koji zauzimaju mjesto na određenoj lokaciji, npr. bitke i sportska natjecanja, skupine ili organizacije koje imaju sjedište, sjedište tvrtke ili sveučilišnog kampusa, artefakti s fizičkim mjestom (npr. Mona Lisa u Louvre).

Neke činjenice također imaju prostornu dimenziju. Na primjer, činjenica da je Leonard Cohen rođen 1934. godine dogodila se u njegovom rodnom gradu Montrealu. Uveden je odnos *occursIn* između činjenice i geo entiteta. Primjerice, ukoliko postoji činjenica fact #1: LeonardCohen bornOnDate 1934, pisalo bi se #1 occursIn Montreal.

²⁵⁴ GeoNames, loc. cit. (bilj. 243)

²⁵⁵ GeoWordNet: A Resource for Geo-spatial Applications. / by Fausto Giunchiglia...[et al.]. // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6088 LNCS, 2010, 121-136. Dostupno na URL: http://link.springer.com/chapter/10.1007%2F978-3-642-13486-9_9 (17.06.2015.).

Ključ čistog semantičkog obrađivanja prostorne dimenzije leži u odnosima. Neke činjenice se pojavljuju na mjestu koje je označeno njihovim subjektom ili objektom. Primjerice, činjenica *Jimi Hendrix je rođen u Seattle* se dogodila u Seattle. Uvedene su dvije nove klase za opisivanje takvih odnosa, *relationLocatedByObject* i *relationLocatedBySubject*, obje podklase od *yagoRelation*. Prva klasa sjedinjuje odnose čije su lokacije dane s lokacijom njihovih objekata, npr. uključuje odnose *wasBornIn*, *diedIn*, *worksAt* i *participatedIn*. Druga klasa okuplja odnose čije su lokacije dane subjektima, npr. uključuje odnos *hasMayor*. Neki odnosi se pojavljuju jedan iza drugoga tj. jedna relacija određuje lokaciju druge. Primjerice, *wasBornOnDate* definira vrijeme odgovarajuće *wasBornIn* činjenice, koji definira lokaciju prethodnog *wasBornOnDate* odnosa. Ova situacija je izražena odnosom *timeToLocation*, koja je sadržana između dva odnosa. Prvi odnos određuje vrijeme događaja dok druga određuje lokaciju. Drugi primjeri uključuju parove *diedOnDate/diedIn* i *happenedOnDate/happenedIn*. Uvedeno je implikacijsko pravilo koje prenosi lokaciju od jednog do drugog odnosa. Ova pravila zajedno izvode lokaciju za činjenicu kadgod je to semantički smisljeno.

Za svaki entitet, YAGO2 sadrži kontekstualne informacije (tj. kontekst), koje se okupljaju ekstraktorima iz Wikipedije. Oni uključuju odnose *hasWikipediaAnchorText* (povezuje entitet sa stringom koji se pojavljuje kao tekstualni link (engl. anchor text) u članku o obrađivanom entitetu), *hasWikipediaCategory* (sadrži ime kategorije u koju Wikipedia smješta članak), *hasCitationTitle* (sadrži naslov reference na Wikipedia stranici). Svi oni su pod-svojstva odnosa *hasContext*, osiguravajući tako obilje ključnih riječi povezanih uz entitet članka. Za pojedine entitete se ekstrahira višejezični prijevod između jezičnih linkova u Wikipediji, dozvoljavajući postavljanje upita individui koristeći ne engleska imena. Višejezična klasna imena su dobivena integriranjem UWN²⁵⁶ u YAGO, koji pridružuje WordNet riječi i značenje s njihovim odgovarajućim dopunama u drugim jezicima osiguravajući tako 1.5 milijuna prijevoda i značenja dodanih za 800,000 riječi na preko 200 jezika s preciznošću od 90%. Sve u svemu, to daje višejezične nazive za većinu entiteta i klasa u YAGO te omogućuje višejezične upite. Kako bi se olakšali upiti, na svaki su triplet (subjekt (S), predikat (P), i objekt (O)) u YAGO2 dodane su još tri dimenzije (vrijeme (T), lokacija (L), i kontekst *contexT* (X)), što daje 6-torka (engl. tuple) predstavljanja, što se naziva SPOTLX. Dakle, SPOTLX se dobiva pridruživanjem događajima vremenskog raspona

²⁵⁶ Melo, Gerard De; Weikum, Gerhard. Towards a Universal Wordnet by Learning from Combined Evidence. // In CIKM, 2009, 0-9. Dostupno na URL: <http://dl.acm.org/citation.cfm?doid=1645953.1646020> (27.03.2015.)

činjenica, lokacija činjenica i konteksta obuhvaćenih entiteta. Za vizualizaciju je upotrijebljen TimeMap²⁵⁷.

Kombinirajući nove dimenzije u YAGO2 se mogu riješiti različite vrste upita. Sve navedeno čini YAGO2 vrijednim pojmovnikom (engl. gazetteer) za zemljopisne, vremenske i semantičke podatke općenito. Ovakvo prostorno-vremensko znanje je ključan oslonac za mnoge aplikacije uključujući povezivanje entiteta preko neovisnih izvora (npr., u the Linked Data cloud²⁵⁸) i semantičko pretraživanje.^{259, 260}

3.3.5 YAGO3

YAGO3 je proširenje YAGO baze znanja koja kombinira podatke iz višejezičnih Wikipedija s engleskim WordNet-om kako bi se izgradila jedna suvisla baza znanja. Iskoristile su se kategorije, infookviri, Wikidata²⁶¹, kao i značenje infookvirnih atributa na više jezika. Metoda je pokrenuta na deset različitih jezika te je postignuta preciznost od 95% - 100% u pridruživanju atributa, čime je YAGO povećao broj novih entiteta i činjenica. Ulazni podaci su popis Wikipedijinih članaka. Cilj je stvoriti bazu podataka od ovih višejezičnih Wikipedija, što uključuje tri izazova: (1) odrediti skup entiteta, (2) ekstrahirati činjenice o tim entitetima iz višejezičnih Wikipedija, (3) stvoriti taksonomiju. Ključna prednost je mogućnost ostavljanja postojećeg engleskog YAGO-a kao referentne baze znanja i glavnog taksonomijskog oslonca. Dok u YAGO-u svaki Wikipedijski članak postaje entitet, kod YAGO3 treba paziti kako se entiteti ne bi duplirali jer isti entitet može biti opisan u različitim jezičnim Wikipedijama. U ovu svrhu je upotrijebljen Wikidata, koji upravlja svojim vlastitim repozitorijem entitetskih i kategorijskih identifikatora te ih pridružuje enciklopedijskim

²⁵⁷ Timemap. Dostupno na URL: <http://timemap.googlecode.com> (27.03.2015.)

²⁵⁸ Bizer, Christian; Heath, T.; Berners-Lee, T. Linked Data - the story so far. International Journal on Semantic Web and Information Systems 5, 3(2009), 1–22. Dostupno na URL: <http://www.igi-global.com/article/linked-data-story-far/37496> (27.03.2015.)

²⁵⁹ YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. / Johannes Hoffart...[et al.]. Dostupno na URL: <http://suchanek.name/work/publications/www2011demo.pdf> (27.03.2015.)

²⁶⁰ YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia. / Johannes Hoffart...[et al.]. Dostupno na URL: <http://resources.mpi-inf.mpg.de/yago-naga/yago/publications/aij.pdf> (27.03.2015.)

²⁶¹ Projekt Wikidata ima cilj biti strukturirana verzija Wikipedije. To je baza znanja koju ispunjavaju činjenicama volonteri. Wikidata osigurava središnje identifikatore za entitete te ih povezuje s člancima u višejezičnim Wikipedijama. Na primjer, Elvis Presley ima identifikator Q303, te ima poveznice člancima Wikipedija na 147 jezika. Wikidata pruža slične podatke za predloške infookvira i kategorijskih imena. Više o Wikidata projektu se može proučiti na: Wikidata. Dostupno na URL: https://www.wikidata.org/wiki/Wikidata:Main_Page (12.01.2016.).

člancima u Wikipedijama na različitim jezicima. Wikidata ekstraktor uzima Wikidata podatke te proizvodi tematski rječnik. Ove teme sadrže činjenice koje pridružuju svako strano kategorijsko i entitetsko ime na svoj engleski par, kao što je prikazano u primjeru:

```
"de/Amerikanische S"anger" hasTranslation "American Singer"  
de/Elvis hasTranslation Elvis
```

Ukoliko neki strani entiteti nemaju engleskog para, na ulaznoj listi jezika na kojoj se pojavljuju entiteti izabire se prvi jezik. Tako se dobiva datoteka koja pridružuje svako strano i englesko ime članka na jedinstveno ime entiteta. Skup ovih jedinstvenih imena je skup YAGO3 entiteta. Ekstrahirane su činjenice iz infookvira. Prvo su upotrijebljeni postojeći ekstraktori za engleski YAGO, a zatim ekstraktori za višejezični YAGO. Ekstrakcija iz višejezičnih Wikipedija protječe slično kao i za englesku verziju, npr. njemački tematski infookvirni atribut sadrži neprerađene parove koji se pojavljuju u članku u sljedećem obliku:

```
de/Elvis de/heirat "[Priscilla Presley], 1967 in [Las Vegas (Stadt)]"
```

Ova činjenica utvrđuje da se Elvis oženio s Priscilla 1967 u Las Vegas. Kao i za englesku Wikipediju, terminski ekstraktor ekstrahira sve moguće termine iz atributnih vrijednosti:

```
de/Elvis de/heirat de/Priscilla_Presley  
de/Elvis de/heirat "1967-##-##"^^xsd:date  
de/Elvis de/heirat de/Las_Vegas_(Stadt)
```

Nakon ekstrakcije termina, prevoditelj entiteta koristi rječnik kako bi entitete preveo na njihova jedinstvena imena:

```
Elvis de/heirat Priscilla_Presley  
Elvis de/heirat "1967-##-##"^^xsd:date  
Elvis de/heirat Las_Vegas
```

Ove činjenice se zovu strane atributne činjenice. Dok su entiteti stranih atributnih činjenica pridruženi na njihova jezično neovisna imena, atributi trebaju biti mapirani na YAGO shemu. U navedenom se primjeru želi zaključiti kako njemački infookvirni atribut

de/heirat odgovara YAGO odnosu *marriedTo*. Ova pridruživanja stranih atributa na YAGO odnose izvodi tzv. *atributni združivač* (*engl. Attribute Matcher*). U navedenom primjeru će se dobiti:

```
de/heirat hasTranslation marriedTo
```

Ova pridruživanja će upotrijebiti atributni pridruživač kako bi se od atributnih činjenica stvorile YAGO činjenice. U navedenom primjeru se dobiva:

```
Elvis marriedTo Priscilla_Presley
Elvis marriedTo "1967-##-##"^^xsd:date
Elvis marriedTo Las_Vegas
```

Tako dobivene činjenice će proći istu provjeru i filtriranje kao ostale YAGO činjenice te će provjera vrste ostaviti:

```
Elvis marriedTo Priscilla_Presley
```

U prethodno navedenom primjeru se dobila činjenica koja je svejedno bila u engleskom YAGO. Međutim, infookvirne atributne činjenice ostalih stranih Wikipedija će dovesti do novih YAGO činjenica koje nisu bile u engleskoj Wikipediji, što dovodi do novih čvorova u YAGO bazi znanja.

Taksonomije se za YAGO jednojezični slučaj engleskog ekstraktora stvara tako da uglavnom dolazi iz Wikipedijinih kategorija. Postupak se pokreće nizom ekstraktora tako da svaki obavi jednu određenu transformacija podataka. Kategorijski ekstraktor ekstrahira članove kategorije, stvarajući temu koja sadrži činjenice, npr.:

```
Elvis inCategory "Rock Music"
Elvis inCategory "American Singers"
```

Naknadni ekstraktor će filtrirati kategorije koje ne odgovaraju klasnim imenima (kao što je *Rock Music*²⁶²) i stvoriti temu koja sadrži izjave kao što su:

```
Elvis type American_Singer
```

²⁶² Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard, loc. cit. (bilj. 249).

Sljedeći ekstraktor će raščlaniti imensku frazu kako bi pronašao da je ova klasa najbližnja podklasi WordNet-ove klase Singer, tako stvarajući:

American_Singer subclassOf Singer

Drugi ekstraktor transformira cijelu WordNet taksonomiju u triplete, što rezultira sljedećim:

Singer subclassOf Person

Person subclassOf LivingBeing

Ovim načinom je entitet Elvis Presley povezan putem podklase American Singer u WordNet taksonomiju.

U engleskoj Wikipediji kategorije članaka su identificirane ključnom riječju *Category*, a drugi jezici upotrebljavaju druge ključne riječi za čije je pronalaženje upotrijebljen Wikidata. Primjerice u Wikidata je *Category: American singers* pridružen na njemački ekvivalent *Kategorie: Amerikanische Sanger*. Navedeno otkriva da se kategorije u njemačkoj Wikipediji uvode s ključnom riječju *Kategorie*. Svi ovi prijevodi su ekstrahirani iz Wikidata. Zatim se može preinačiti postojeći YAGO kategorijski ekstraktor kako bi se ekstrahirali kategorijski članovi iz stranih Wikipedija, primjerice za njemačku Wikipediju kategorijski ekstraktor ekstrahira:

de/Elvis inCategory "de/Amerikanische Sanger"²⁶³

3.3.6 InPhO

Indiana Philosophy Ontology (InPhO)²⁶⁴ je projekt pokrenut na Indiana Sveučilištu, koji upotrebljava kombinaciju automatskih metoda i povratne informacije stručnjaka za stvaranje dinamične računalne ontologije iz filozofske discipline. InPho projektom se razvija

²⁶³ Navedeni primjeri su preuzeti iz članka: Mahdisoltani, Farzaneh; Biega, Joanna; Suchanek, Fabian M. YAGO3: A Knowledge Base from Multilingual Wikipedias. Dostupno na URL: http://people.mpi-inf.mpg.de/~jbiega/papers/yago3_cidr2015.pdf (17.12.2015.).

²⁶⁴ InPhO Project. Dostupno na URL: <http://inpho.cogs.indiana.edu/> (27.03.2015.).

ontologija za dinamično referentno djelo online otvorenog pristupa Stanford Encyclopedia of Philosophy (SEP)²⁶⁵. Ontologija je prikladna i u drugim digitalnim filozofskim primjenama²⁶⁶. InPhO ontologija podrazumijeva ontologiju u značenju termina informacijskih znanosti, odnosno „računalnu ontologiju“ tj. formalno-kodiranu specifikaciju pojmova relevantnih za predmetnu domenu (uključujući njihova svojstva i relacije između njih) i hijerarhijsku klasifikaciju ovih pojmova u kategorije i podkategorije (Noy and McGuinness²⁶⁷; Gruber^{268,269,270}).

Svrha takve ontologije je pomoći ljudima i automatskim agentima u razumijevanju sadržaja domene (posebno u pogledu svojstava, odnosa i nasljednih pododnosa) te omogućavanje interoperabilnosti s drugim projektima. SEP enciklopedija je prvi pokušaj akademske zajednice organiziranja vlastite profesionalne predmetne građe kolaborativnim pisanjem, objavljivanjem i upravljanjem dinamično ažuriranim referentnim djelom u cijelosti na webu. Članke dostavljaju više od 1.300 volontera autora i asinkrono ih ažuriraju kako bi odražavale najnovije spoznaje u znanosti. Sve što autori napišu pregledavaju volonteri pojedinih predmeta prije objave na webu, što SEP pristup čini različitim od Wikipedije. Rezultat organizacije ovakvog radnog procesa je visoka pozicija SEP-a na Google pretraživačkoj listi (GoogleTM search lists) za filozofske pojmove i filozofe čineći ga

²⁶⁵ Stanford Encyclopedia of Philosophy. Dostupno na URL: <http://plato.stanford.edu/> (27.03.2015.).

²⁶⁶ Primjerice, InPhO je integriran s Noesis philosophy search engine (Noesis. Dostupno na URL: <http://noesis.evansville.edu/> [15.05.2015.]) te samo InPhO sučelje već osigurava ontologijski vođeno pretraživanje u SEP, Noesis, i Google Scholar.

²⁶⁷ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36)

²⁶⁸ Gruber, Thomas R. A Translation Approach to Portable Ontology Specifications. // Knowledge Acquisition 5, 2 (1993), 199–220.

²⁶⁹ Gruber, Thomas; Liu, Ling; Özsu, M. Tamer. Ontology // In Encyclopedia of Database Systems / ed. by Ling Liu and M. Tamer Özsu. New York: Springer-Verlag, 2009. Dostupno na URL:

<http://www.dainf.ct.utfpr.edu.br/~tacla/Ontologias/2013/a01-020-Ontology-definitionTomGruber.pdf> (19.09.2015.).

²⁷⁰ Točnije rečeno, ontologija se shvaća kao usmjereni aciklički graf u kojem čvorovi predstavljaju pojmove te su veze između pojmova predstavljene taksonomskim *isa* odnosom (npr. u grafu gdje je *Wine* → *Red Wine* → *Brunello di Montalcino*, sve što je (*isa*) instanca od *Brunello di Montalcino* je (*isa*) instanca od *Red Wine*, i sve što je (*isa*) instanca od *Red Wine* je (*isa*) instanca od *Wine*. Baza znanja je ontologija ispunjena s individuama; matematički, baze znanja sadrže drugu vrstu poveznice označene "instancā od" taksonomskog odnosa, te novu vrstu čvor označavanja individua (uz ograničenje da nijedna individua ne može imati nikakvu „djecu“). Tako se ontologija može ispuniti s individuom *Campogiovanni Brunello*, koji je instanca pojma *Brunello di Montalcino*. Ontologija će sadržavati izjave brojnih netaksonomijskih odnosa, koji mogu biti sadržani između individua i konstanti (svojstva) ili između dviju (ili više) individua (odnosa). Instance tih netaksonomskih odnosa se mogu kodirati u procesu ispunjavanja ontologije. Konačno, računalna ontologija može sadržavati niz pravila zaključivanja i aksioma koji se mogu koristiti za zaključivanje o objektima u domeni (npr zaključiti prisutnost određenih odnosa na temelju drugih, provoditi dosljednost i zadanost svojstva itd.).

najvidljivijim i najpopularnijim online referentnim djelom iz filozofske discipline, što je i evidentirano s prosječno skoro 600,000 preuzimanja članaka tjedno²⁷¹.

Glavni cilj SEP-a je zadržati enciklopediju dostupnom bez naknade za znanstvenike i javnost. Međutim, inovativna narav organizacije posla sa sobom donosi niz novih poteškoća koje nisu bile prisutne u tradicionalnom enciklopedijskom radu, npr. veoma je nepraktično imati redakciju za ručno upravljanje unakrsnim uputnicama, tablicama sadržaja, pretraživanje ključnih riječi i druge metasadržaje zbog asinkronih davanja na uvid i revizija članaka. Tu je i pritisak uredničkog tereta stavljenog na volontere suradnike, koji ne mogu stalno pratiti svaku promjenu sadržaja SEP-a i ažuriranje samih metapodataka.

Navedeni izazovi sa željom očuvanja slobodnog otvorenog pristupa enciklopediji, stvaraju snažnu potrebu razvoja automatskih i poluautomatskih alata za upravljanje informacijama koji se mogu integrirati u tijek uredničkog rada. Razvoj ontologije za područje filozofije je središte uspjeha u ovom projektu. Umjesto traženja potpuno automatiziranog rješenja za SEP-ove potrebe metapodataka, nastoji se iskoristiti SEP-ov najvrjedniji izvor, a to je redovni pristup stručnjacima domene filozofije. Pristup dinamičkoj InPho ontologiji počinje s malom količinom početne ručno izgrađene ontologije, nakon čega se koriste razne automatizirane metode za strukturiranje oblika povratne informacije i njezino razvijanje u valjane podatke, popunjavanje ontologije te polu-automatsko proširivanje ontologijske taksonomijske strukture.

Najbogatiji izvori informacija za projekt su dvije bibliografije upravljane SEP-ovim urednicima. David Chalmers je stvorio taksonomiju bibliografije za specijalizaciju filozofije uma²⁷². Dodatno, tu je i bibliografija filozofije znanosti održavana na Sveučilištu Pittsburgh (stvorili su je Rob Clifton, John Earman, i John Norton)²⁷³. Dopuštenjem autora, obje bibliografije su odgovarajućim preoblikovanjem uključene u ontologiju. Wikipedia stalno dodaje sve više klasifikacijske strukture za svoju enciklopediju, kao i različite odlike poput tablice sadržaja (engl. "table of contents") sadržeći proširivu klasifikaciju filozofskih misli i njihovo pozicioniranje. Međutim, Wikipedija zbog svoje otvorenosti ipak ostavlja mogućnost netočnih informacija, zlonamjerno upravljanje kontraverznim entitetima i vandalizam, što je

²⁷¹ About the Stanford Encyclopedia of Philosophy. Dostupno na URL: <http://plato.stanford.edu/about.html> (01.04.2015).

²⁷² MindPapers. Dostupno na URL: <http://consc.net/biblio.html> (15.06.2015.). David Chalmers i David Bourget razvijaju proširenje ove taksonomije kako bi pokrili sva područja filozofije, nazvanu Philpapers. Dostupno na URL: <http://consc.net/taxonomy.html> (15.06.2015.).

²⁷³ PhilSci Archive. Dostupno na URL: <http://philsci-archive.pitt.edu/view/subjects/> (15.06.2015.).

čini neprikladnom za svrhu akademskih referentnih izvora²⁷⁴. InPho pristup tretira metapodatake izvedene iz Wikipedije kao vrijedan početni neprovjereni izvor, koji provjerava stručnjak domene prije nego se ugradi u InPho predstavljanje. U razvoju InPho ontologije vodi se principima koja u svom radu navode Noy and McGuinness.²⁷⁵ InPho ontologija se sastoji od šest osnovnih kategorija: *Filozof*, *Dokument*, *Organizacija*, *Nacionalnost*, *Profesija* i *Misao*. Kategorije *Filozof*, *Profesija* i *Nacionalnost* imaju malo klasifikacijske strukture i sadrže samo instance popisanih filozofa, nacionalnosti i zanimanja dobivenih iz različitih izvora ili dopuštajući autorima (odnosno urednicima) ručno unošenje osoba koje nedostaju.

Pojmovi gornje razine (engl. top-level) su pridruženi u standardne ontologije gdje je to moguće u svrhu povećanja interoperabilnosti podataka. Tako je npr. *Filozof* mapiran u W3C klasu *FOAF:person* (proširen s dodatnim biografskim svojstvima)²⁷⁶. Hijerarhijska struktura i svojstva od *Dokument* su usvojena od AKT Reference Ontology²⁷⁷, a *Organization* je uzeta (uz odgovarajuće preinake) iz Protégé ontology library²⁷⁸.

Za početni nacrt ontologije izravno je prikupljena taksonomijska shema od urednika SEP-a te je tako početno građenje kategorija oblikovano traženjem poddisciplinske taksonomije od strane stručnjaka područja, kojima se daje više slobode za osiguranje najbolje konceptualizacije ciljane domene. Prednost ovog pristupa je osiguranje visoko semantički relevantnih informacija, koji će najvjerojatnije dovesti do objektivnog, sveobuhvatnog i jasnog predstavljanja predmeta²⁷⁹.

Drugi važan aspekt prilikom stvaranja ontologije tiče se podvlačenja razlike između kategorije i individua. Na primjer, treba li *connectionism* biti razmatran kao instanca kategorije *philosophy of artificial intelligence* ili kao podkategorija sa svojim vlastitim instancama? Hoće li *connectionism* biti kategorija ili individua ovisit će o količini obrađivanja koje dobiva u trenutnoj verziji SEP-a.

²⁷⁴ Više o tome se može pročitati u: 1) Jaschik, Scott. A stand against *Wikipedia*, *Inside Higher Ed* (January), 2007. Dostupno na URL: <http://www.insidehighered.com/news/2007/01/26/wiki> (15.06.2015.). i 2) Giles, Jim. Internet encyclopaedias go head to head: special report // *Nature* 438, 15(2005), 900–901. Dostupno na URL: <http://www2.stat.unibo.it/mazzocchi/macroeconomia/documenti/Nature%202006%20wikipedia.pdf> (15.06.2015.).

²⁷⁵ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

²⁷⁶ FOAF Vocabulary Specification. Dostupno na URL: http://xmlns.com/foaf/spec/#term_Person (15.06.2015.).

²⁷⁷ Aktors. Dostupno na URL: <http://www.aktors.org/publications/ontology/> (15.06.2015.).

²⁷⁸ Protégé ontology library. Dostupno na URL: http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library (15.06.2015.).

²⁷⁹ Sanger, Larry. Who says we know?: On the new politics of knowledge. *The Edge*, 208. Retrieved March 22, 2008. Dostupno na URL: http://www.edge.org/3rd_culture/sanger07/sanger07_index.html (15.06.2015.).

InPho ontologija je podložna reverzibilnoj korekciji, stoga se stavke obrađivane kao individue u jednoj iteraciji mogu obrađivati kao kategorije (klase) u sljedećoj. Iako *connectionism* može početno biti određen kao individua, kako enciklopedija dodaje više entiteta u kognitivne znanosti, može postati prikladno da ga se započne obrađivati kao kategoriju (“*ideas about connectionism*”) s vlastitim instancama kao što su *parallel processing*, *backpropagation*, *distributed representation*, itd.

Neki pojmovi i individue ne mogu prirodno stajati u jednom jedinstvenom položaju u ontologiji te su višestruko klasificirani od strane stručnjaka i (ili) automatiziranim metodama. Ovo pitanje se pokušalo riješiti ograničavanjem pojmova na jednom pojavljivanju po poddisciplini dodavanjem “semantičkih unakrsnih linkova” na ontologiji. Unakrsni linkovi koji dohvaćaju srodnost *Misli* se smatraju međusobno uzajamno relevantnim po povratnim informacijama ili automatiziranim metodama koje su ručno klasificirane od stručnjaka u udaljenim područjima ontologije. Semantičke informacije koje slijede kao zaključci su bitne za uredničke potrebe poput unakrsnog upućivanja, kao i za čitatelje koji će možda poželjeti slijediti vodstvo u srodne teme. Uzimajući u obzir mogućnost pretraživosti pojmova putem semantičkih veza također se dozvoljava pročišćavanje ontologije i klasificiranje individua u najpovoljniju lokaciju tijekom popunjavanja ontologije²⁸⁰.

Misao koju je stručnjak kategorizirao na nisku razinu specifičnosti te koja ne odražava njezin odgovarajući termin za veću prisutnost u korpusu, očuvati će svoj visoki stupanj povezanosti imajući velik broj semantičkih međuveza.

Poluautomatsko popunjavanje InPhO taksonomijskih i netaksonomijskih odnosa odvija se kroz zahtijevanje povratne informacije stručnjaka kao dio SEP-ovog dokumentiranog procesa davanja na uvid. Kada autori SEP-a predaju dokument za objavljivanje traži se njihova povratna informacija u trostupanjskom procesu. Na dokumentima autora se izvode statističke mjere kako bi se obuhvatili potencijalno važni termini i imena koji se pojavljuju u dokumentu. Autorima se predstavlja popis takvih termina s upitom je li se oni u njihovom dokumentu uistinu obrađuju ili se pojavljuju u njemu samo usput (“Je li vaš članak o *connectionism*?”).

Autori procjenjuju srodnost (skala s pet točaka, od 'potpuno nepovezani' da 'vrlo povezani') i općenitost ('više općeniti', 'manje općeniti', 'općeniti', i 'neusporedivi') tako statistički visoko rangiranih termina s drugim terminima, procjenjujući ih kao kandidate

²⁸⁰ Niepert, Mathias; Buckner, Cameron, Allen; Colin. Answer set programming on expert feedback to populate and extend dynamic ontologies. // *Proceedings of 21st FLAIRS Conference*, 2008, 500-505. Dostupno na URL: <http://www.aaai.org/Papers/FLAIRS/2008/FLAIRS08-116.pdf> (16.06.2015.).

hiponima (engl. hyponymy, od manje općenitosti) ili hipernima (engl. hypernoms, od veće općenitosti) kako bi se dobili ciljani pojmovi. Dobivena povratna informacija se upotrebljava za popunjavanje ontologije razvrstavanjem pojavljivanih termina dokumenta u njezinu pojmovnu hijerarhiju, što se ostvaruje predstavljanjem povratne informacije stručnjaka kao niza činjenica računom predikata prvog reda (engl. first-order predicate calculus)²⁸¹ i upotrebom tehnika zaključivanja (engl. inference techniques) s ciljem pretpostavljanja klasifikacije uvedene povratnom informacijom.²⁸² Trenutno se ta metoda koristi samo za klasificiranje instanci klase *Misao* prema postojećoj ontologijskoj strukturi.²⁸³

U završnoj fazi se od stručnjaka traži evaluacija netaksonomijskih odnosa pretpostavljenih automatiziranim metodama.²⁸⁴ Uvedeni su netaksonomski odnosi iz Wikipedije i statističke metode kao što su skriveni Markovljev model i uvjetna slučajna polja (engl. conditional random fields), kako bi se pretpostavili odnosi pribavljeni između termina pojavljivanih u dokumentu. Autori i urednici potvrđuju ili negiraju dobivene hipoteze o netaksonomskim odnosima između individua u bazi znanja, a njihove povratne informacije će se koristiti za izgradnju skupa za testiranje daljnjeg povećanja statističkih hipoteza. Ova strategija iskorištavanja povratne informacije pokazuje nastojanja postizanja većeg stupnja pouzdanosti od čistog pojavljivanja (engl. co-occurrence) i nenadziranog wiki stil pristupa.

²⁸¹ Primjerice, ukoliko autor ukaže kako je *connectionism* visoko srodan i više specifične od *philosophy of mind*, ove dvije činjenice se predstavljaju kao visoko_srodne(*connectionism, philosophy of mind*)” i “više_specifičan(*connectionism, philosophy of mind*)”.

²⁸² Pristup ovom zadatku, kao i njegove prednosti su detaljnije opisane u Niepert, Mathias; Buckner, Cameron, Allen; Colin, loc. cit.

²⁸³ Pogledati Niepert, Mathias; Buckner, Cameron; Allen; Colin, loc. cit. za prijedloge o tome kako bi se to moglo ostvariti.

²⁸⁴ Niepert, Mathias; Buckner, Cameron; Allen; Colin. A dynamic ontology for a dynamic reference work. // *Proceedings of JCDL 2007*, 2007, 288–297. Dostupno na URL: <https://inpho.cogs.indiana.edu/papers/2007NiepertBucknerAllen.pdf> (16.06.2015.).

3.4 Zaključak

Poglavlje *Dosadašnji ontologijski projekti* je dalo prikaz razvoja značajnih ontologija na semantičkom webu, kao i prikaz razvoja ontologija temeljen isključivo na enciklopedijskom znanju. Iznesene činjenice o enciklopedijskim ontologijskim projektima ukazuju na to da su navedene ontologije razvijane na temelju strukturiranih podataka enciklopedijskog članka (Wikipedijini infookviri u DBpedia projektu), doprinosa stvaranja strukturiranih podataka od strane samih korisnika (Freebase) iz različitih izvora (Wikipedia, MuzicBranz) te kombinacija automatskih metoda i povratne informacije autora i urednika domene (InPho). Navedeno ukazuje na to kako niti jedan dosadašnji projekt nije pokušao povezati i istražiti razvoj ontologije i njezinih gradivnih elemenata sa strukturom samog enciklopedijskog članka istražujući značaj pojedinih elemenata enciklopedijskog članka za razvoj ontologije, a ne samo značaj strukturiranih podataka enciklopedijskog članka. Upravo će se s tom neistraženom vezom nestrukturiranih dijelova enciklopedijskog članka i ontologije baviti ovaj rad.

4 UVOD U ONTOLOGIJSKI MODEL KNJIŽEVNOSTI I PROTEGE PROGRAM

4.1 Uvod

Poglavlje *Dosadašnji ontologijski projekti* je dalo objašnjenje pojma semantički web i kratki prikaz standarda koji se primjenjuju u njegovom ostvarivanju. U ovom poglavlju će se ukratko prikazati koji su elementi standarda i ontologijskih jezika primijenjeni u razvoju ontologije književnosti temeljenoj na biografskim enciklopedijskim člancima Hrvatske enciklopedije.

Za sami razvoj ontologije književnosti iz enciklopedijskih biografskih članaka Hrvatske enciklopedije primijenjen je Protégé²⁸⁵ računalni program otvorenog koda (engl. open source) nastao na Stanford University s jednostavnim grafičkim korisničkim sučeljem za jednostavan i brz način izrade, preoblikovanja, provjere i vizualizacije ontologije. Izradu ontologije nadgleda i provjerava sam program, koji javlja pogreške u slučaju nedefiniranih elemenata ontologije. Program podržava implementaciju brojnih dodataka (engl. plug-in) koje korisnici sami mogu stvarati ili instalirati već postojeće. Potpuno je kompatibilan sa OWL 2 jezikom za predstavljanje ontologijskog znanja.

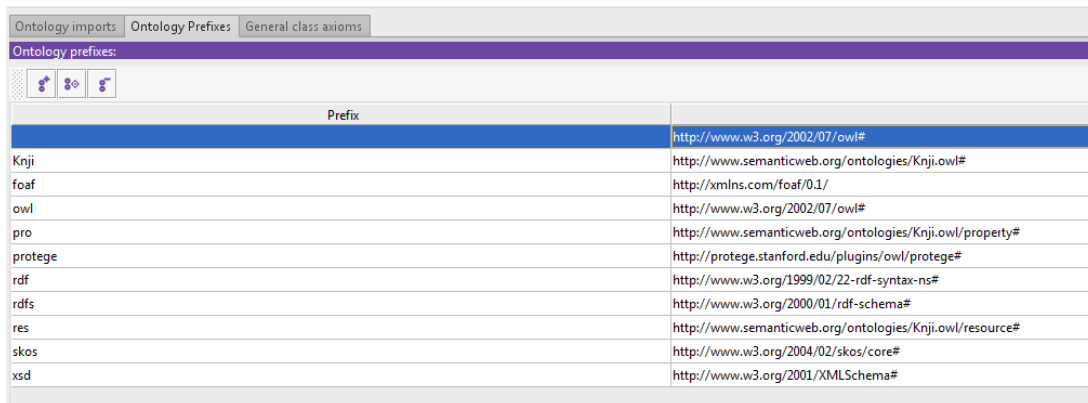
Poglavlje će ukratko kroz nekoliko primjera iz ontologije književnosti Hrvatske enciklopedije prikazati upotrijebljene gradivne elemente ontologijskih standarda i općenito postupak izrade ontologije pomoću Protégé programa-a, a poglavlje *Razvoj ontologije književnosti* će detaljno prikazati cijelu ontologiju književnosti dobivenu iz članka Hrvatske enciklopedije, izrađenu na temelju standarda semantičkog weba prikazanih u ovom poglavlju.

4.2 Definiranje domene i prefiksa

Na samom početku izgradnje ontologije potrebno je definirati njezinu domenu i prefiks. Nakon pokretanja Protégé programa, u glavnom prozoru je moguće deklarirati glavnu domenu i prefikse koji će se koristiti u ontologiji. Standardni prefiksi sadržani u OWL

²⁸⁵ Protégé je besplatno dostupni ontologijski editor otvorenog koda, razvijen je na Stanford Center for Biomedical Informatics Research (BMIR) na Stanford University School of Medicine. // Protégé. Dostupno na URL: <http://protege.stanford.edu/about.php> (3.06.2016.).

ontologiji koje Protégé automatski dodjeljuje su *rdf*, *rdfs*, *owl*, *xml* i *xsd*. Slika 10 prikazuje glavni prozor Protégé-a u kojem su dodijeljeni prefiksi i domena ontologije.



Prefix	URI
Knji	http://www.w3.org/2002/07/owl#
foaf	http://www.semanticweb.org/ontologies/Knji.owl#
owl	http://www.w3.org/2002/07/owl#
pro	http://www.semanticweb.org/ontologies/Knji.owl/property#
protege	http://protege.stanford.edu/plugins/owl/protege#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
res	http://www.semanticweb.org/ontologies/Knji.owl/resource#
skos	http://www.w3.org/2004/02/skos/core#
xsd	http://www.w3.org/2001/XMLSchema#

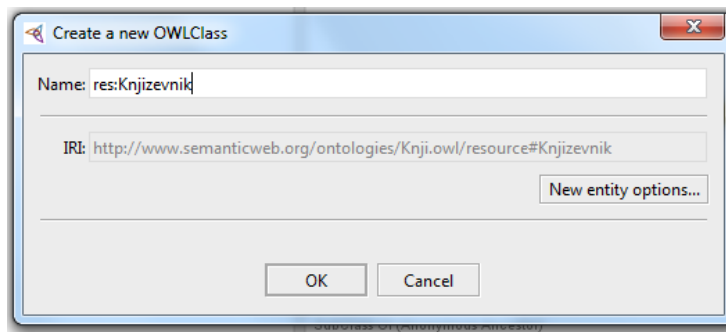
Slika 10: Glavni prozor Protege-a

4.3 Definiranje taksonomije klasa

Nakon deklariranja glavne domene i prefiksa ontologije, potrebno je izraditi taksonomiju klasa. Klase predstavljaju grupe sličnih resursa, odnosno individua te mogu sadržavati druge klase, odnosno podklase koje predstavljaju detaljnije grupe resursa. U klasama su smještene individue, odnosno krajnji resursi, a to su npr. u ontologiji književnosti konkretni hrvatski književnici. Klase se u Protege programu izrađuju u kartici *Classes*. U OWL ontologiji vrh hijerarhije predstavlja klasa *owl:Thing*, koja je ustvari osnovna (zadana, korjenska) klasa u OWL-u. Sve ostale klase koje se definiraju su ustvari podklase *owl:Thing* klase (uključujući i *rdfs:Resource* klasu). Definiranje klase se sastoji od dva dijela: davanje imena klasi (*Književnik*) te dodavanjem restrikcija (u ovom slučaju jedina restrikcija je dana pomoću *subClassOf* elementa). Na Slici 10 je uočljiv globalni *namespace* OWL rječnika <http://www.w3.org/2002/07/owl#> s uobičajenim prefiksom *owl* koji se dodjeljuje. Navedeno označava OWL jezik, što za ontologiju znači kako svi elementi sa prefiksom *owl* označavaju stvari koje se nalaze na <http://www.w3.org/2002/07/owl#> što je standardan način označavanja OWL rječnika. Ono što u RDFS-u predstavljaju *rdfs:Class* i *rdfs:Resource*, to u OWL ontologiji predstavljaju *owl:Class* i *owl:NamedIndividual*. Stoga je na sličan način preko odnosa *rdf:type* moguće definirati resurs kao klasu ili individuu, a u ovom primjeru se definira klasa *Književnik*:

```
res:Književnik    rdf:type    owl:Class .
```

Klikom na gumb *Add subclass* u Protege programu otvara se novi prozor u koji se upisuje URI koji će predstavljati novu klasu. S obzirom da su u glavnom prozoru zadani prefiksi *res*, *pro* i moguće je umjesto cijelog URI-ja pisati prefiks i ime klase odvojene dvotočkom, odnosno *qname* (npr. *res:Barac*, *Antun*). Navedeno stvaranje nove klase je prikazano na Slici 11 u Protege prozoru.



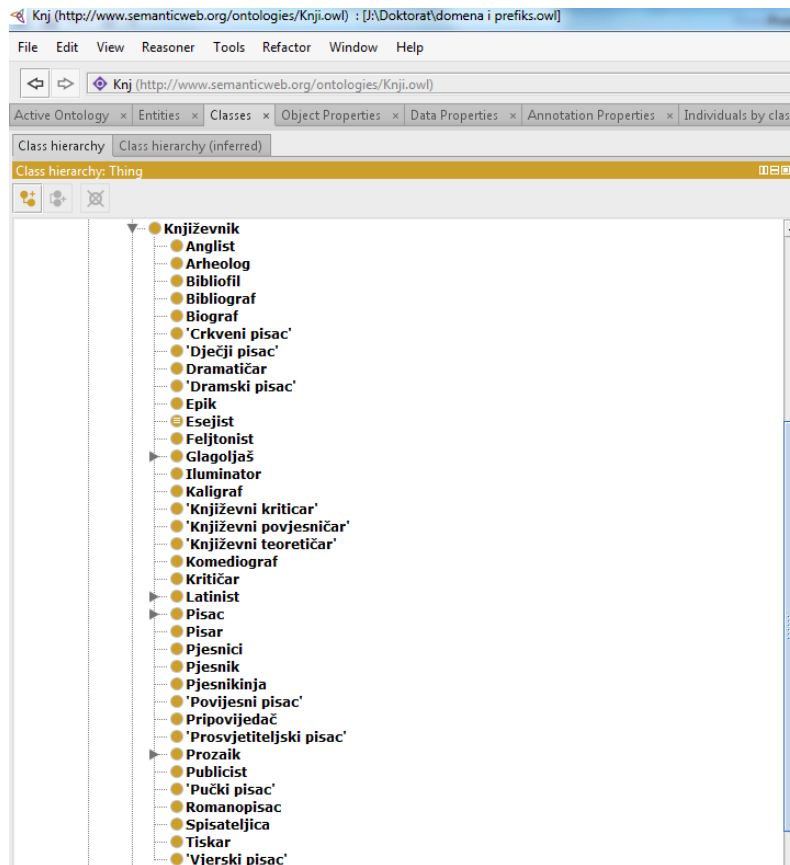
Slika 11: Stvaranje nove klase u Protégé-u

Nakon stvaranja klase *Knjizevnik*, na isti način se stvara taksonomija podklasa navedene klase. Potrebno je označiti klasu *Knjizevnik* i prethodno navedenim postupkom stvaranja novih klasa u Protege-u stvoriti podklase klase *Knjizevnik*. Ovo proširenje hijerarhije klasa se izvodi korištenjem *rdfs:subClassOf* odnosom. Globalni *namespace* RDFS rječnika je *http://www.w3.org/2000/01/rdf-schema#*, a uobičajeni prefiks koji se dodjeljuje je *rdfs*. Primjerice, moguće definirati klasu *res:Esejist* i *res:Romanopisac* kao podklase klase *res:Knjizevnik*. Ovime se zapisuje da je svaka individua koja će se naći u podklasama *res:Esejist* i *res:Romanopisac* ujedno i član klase *res:Knjizevnik*:

res:Esejist *rdfs:subClassOf* *res:Knjizevnik*.

res:Romanopisac *rdfs:subClassOf* *res:Knjizevnik*.

Rezultat opisanog postupka u Protege programu je prikazan na Slici 12, koja ilustrira podklase dodijeljene klasi *Knjizevnik*.



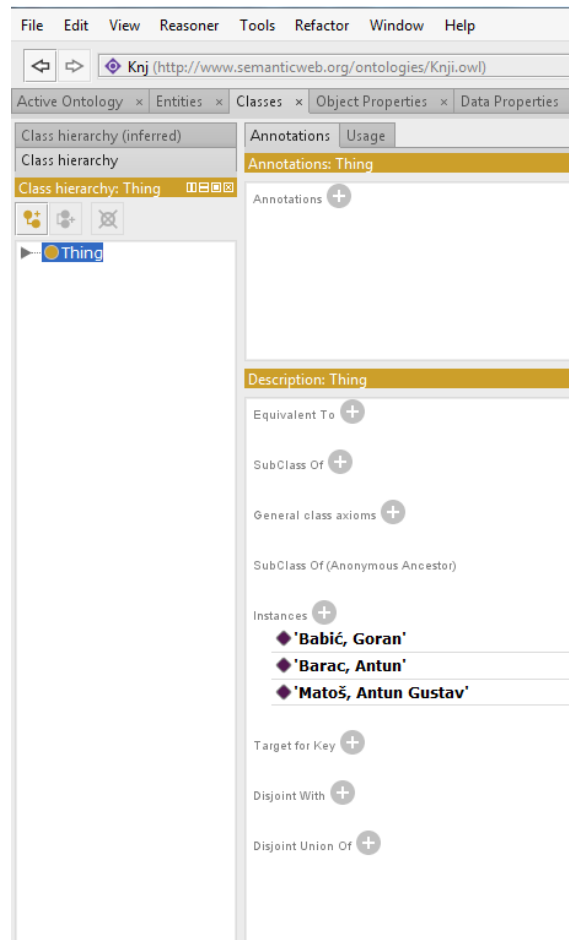
Slika 12: Prikaz dijela taksonomije klasa iz ontologije književnosti kreiranih u kartici Classes

4.4 Izgradnja individua

Individue se u OWL ontologiji definiraju s *owl:NamedIndividual*. Moguće je na isti način, kao i kod definiranja klasa, preko odnosa *rdf:type* definirati resurs kao individuu. U ovom primjeru se definira individua *Barac, Antun*:

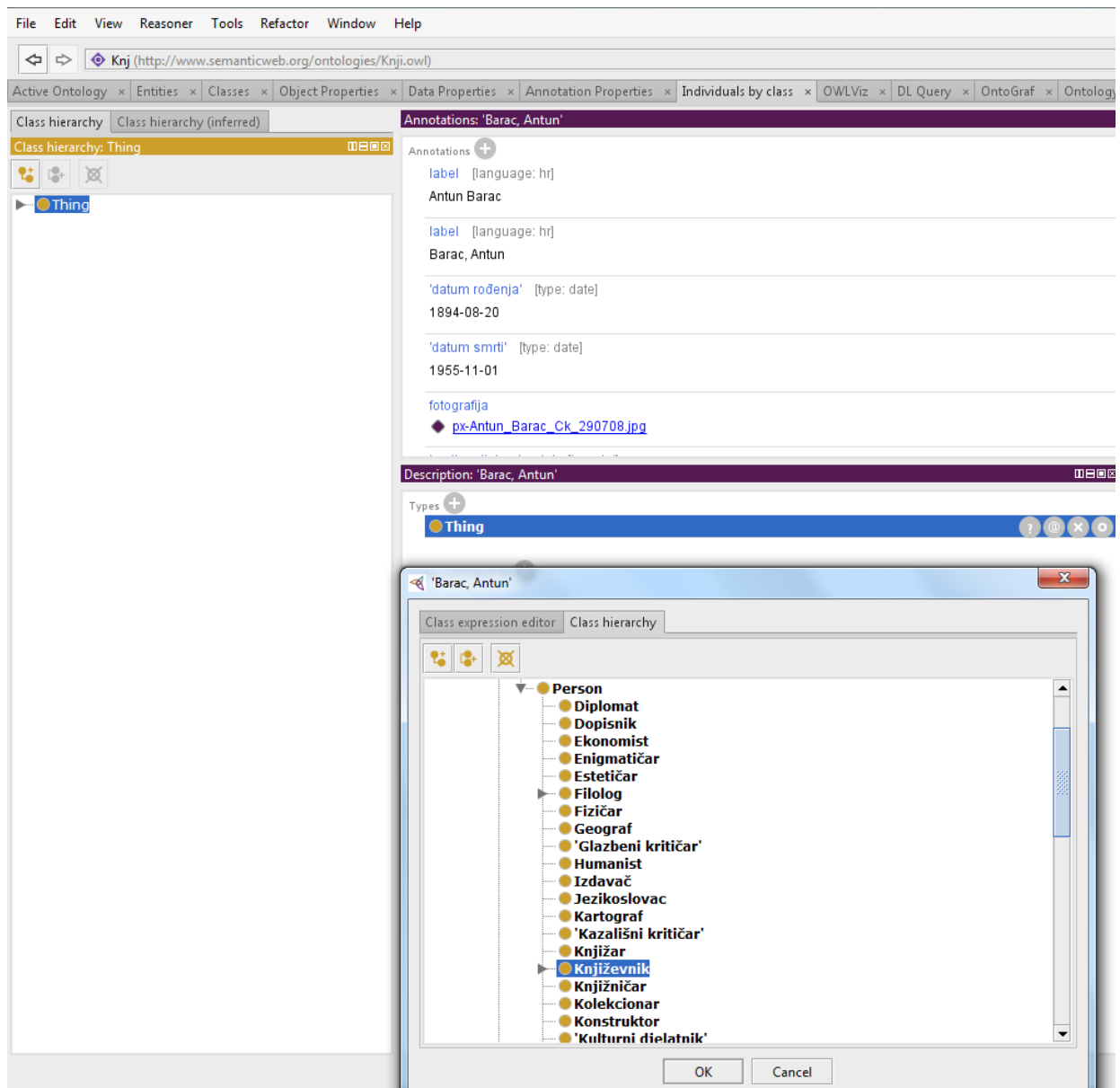
```
res: Barac, Antun    rdf:type    owl:NamedIndividual
```

U Protege programu se individue izrađuju u kartici *Individuals*. Klikom na gumb *Add individual* otvara se novi prozor u kojem se na isti način kao i za klasu upisuje URI ili *qname* pojedine individue. Slika 13 prikazuje individue (Barac, Antun; Babić, Goran; Matoš, Antun Gustav) iz ontologije književnosti stvorenih u kartici *Individuals*.



Slika 13: Prikaz stvorenih individua u Protégé-u

Na prethodnoj slici je uočljivo kako stvorene individue pripadaju klasi *Thing*. Potrebno je odrediti željenu klasu kojoj će se dodijeliti individue što se postiže odabirom gumba *Types*. Otvara se novi prozor u kojem je potrebno odabrati klasu kojoj se pridružuje individua. Slika 14 prikazuje dodavanje individue *res: Barac, Antun* u klasu *Književnici* odabirom gumba *Types* klasi *Književnici*.



Slika 14: Dodavanje individue u odgovarajuću klasu u Protégé-u

4.5 Definiranje odnosa

4.5.1 Objektni i podatkovni odnosi

Pomoću OWL klasa je moguće odrediti vrstu odnosa. Postoje dvije klase, odnosno dvije vrste odnosa, *Object property* i *Datatype property*. Ukoliko se odnos svrstava u klasu *owl:ObjectProperty* tada ona ima funkciju povezivanja klasa i/ili individua međusobno, a ako se pak svrstava u klasu *owl:DatatypeProperty* ima funkciju povezivanja klasa i/ili individua

sa doslovnim vrijednostima. Na sljedeći način se odnosi dodjeljuju u odgovarajuću klasu i pri tome definiraju:

`pro:jeObjavljujeUČasopisu` `rdf:type` `owl:ObjectProperty` .

`pro:jeGodinaRođenja` `rdf:type` `owl:DatatypeProperty` .

Odnos `pro:jeObjavljujeUČasopisu` će vezati dvije individue poput `res:Matoš`, `Antun Gustav` i `res:Vijenac`, dok će relacija `pro:jeGodinaRođenja` vezati resurse poput individue `res:Matoš`, `Antun Gustav` sa doslovnom vrijednosti `1873`.

Odnosi koje će povezivati klase, individue i doslovne vrijednosti se u Protege programu definiraju u istoimenim karticama tj. u *Object Properties* kartici i *Data Properties* kartici. Kao što je već može pretpostaviti iz prethodnog objašnjenja, u kartici *Object Properties* definiraju se odnosi koje će povezivati klase i individue međusobno, a u *Data Properties* kartici definiraju se odnosi koje će povezivati klase i individue sa doslovnim vrijednostima.

Također je moguće stvoriti hijerarhiju pododnosa, kao i kod taksonomije klasa. Vrh hijerarhije odnosa u OWL ontologiji će predstavljati odnosi `owl:topObjectProperty` i `owl:topDataProperty`. Proširenje hijerarhije odnosa se izvodi korištenjem `rdfs:subPropertyOf` odnosom. Odnosi `pro:jeBibliografskiObradio` i `pro:jeBiografOsobe` mogu pripadati grupi odnosa `pro:jeProučavanjeOsobe`:

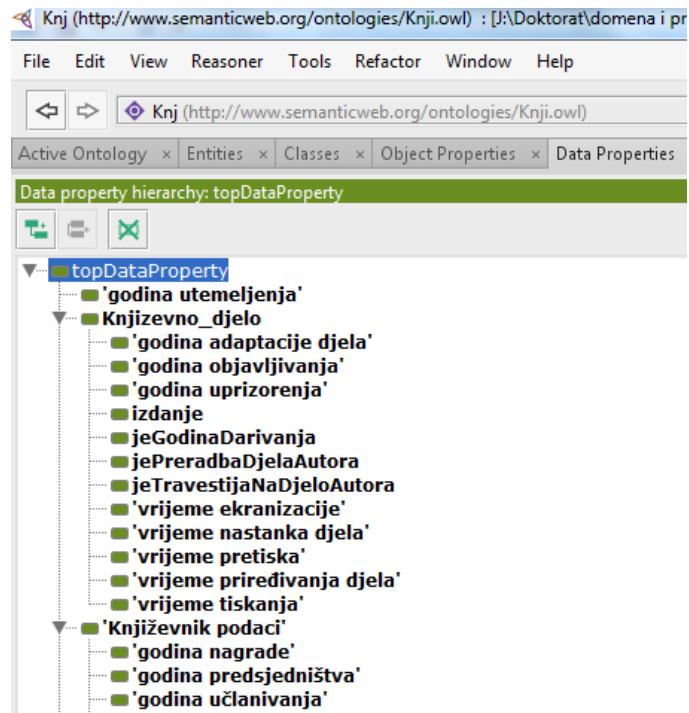
`pro:jeBibliografskiObradio` `rdfs:subPropertyOf` `pro:jeProučavanjeOsobe`.

`pro:jeBiografOsobe` `rdfs:subPropertyOf` `pro:jeProučavanjeOsobe`.

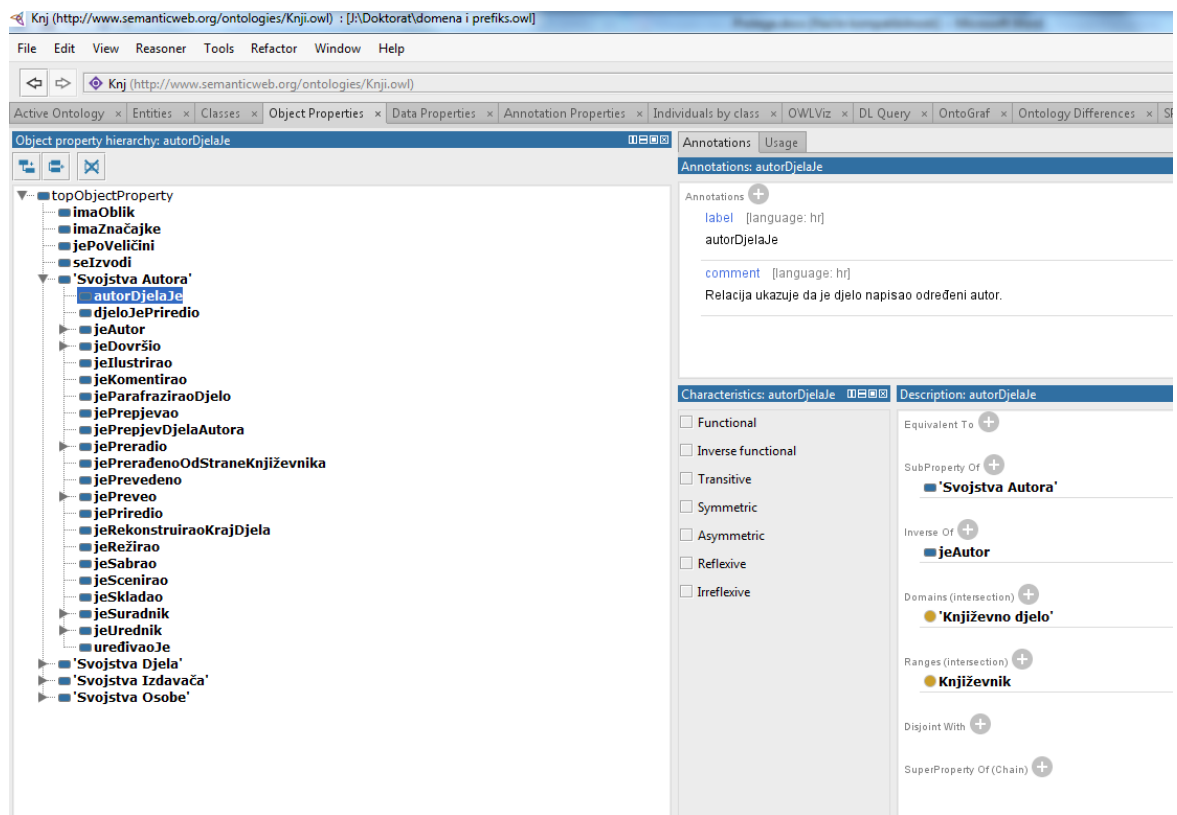
Tako svrstani odnosi definiraju da dva resursa koja su povezana odnosom

`pro:jeBibliografskiObradio` su ujedno povezana i odnosom `pro:jeProučavanjeOsobe`:

Klikom na gumb *Add sub property* u Protege programu otvara se novi prozor, u koji se upisuje URI ili *qname* novog odnosa, kao kod stvaranja klasa i individua. Slike 15 i 16 prikazuju dio hijerarhije odnosa iz ontologije književnosti stvorenih u *Data Properties* i *Object Properties* karticama.



Slika 15: Prikaz podatkovnih odnosa u Protégé-u



Slika 16: Prikaz stvorenih objektnih odnosa u Protégé-u

Dva bitna odnosa u RDFS rječnika, *rdfs:domain* i *rdfs:range*, služe za svrstavanje resursa u klase. Koristeći odnos *pro:jeAutor* definiran s dva prethodno navedena odnosa, resursi RDF trojca pod subjektom (*rdfs:domain*) i objektom (*rdfs:range*) se indirektno svrstavaju u klase.

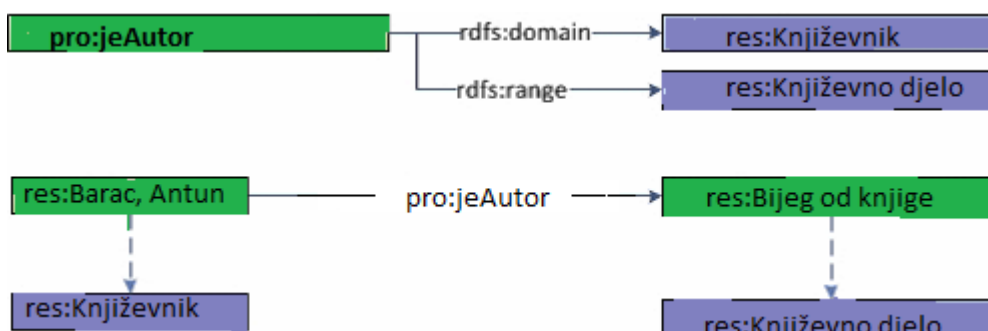
Neka postoje klase *res:Književnik* i *res:Književno djelo*, te individue *res:Barac*, *Antun* i *res:Bijeg od knjige* koje nisu dodijeljene u nijednu klasu. Odnos koji veže navedene individue je *pro:jeAutor*.

res:Barac, Antun pro:jeAutor res:Bijeg od knjige.

Tom odnosu se mogu definirati *rdfs:domain* i *rdfs:range*.

pro:jeAutor rdfs:domain res:Književnik ;
rdfs:range res:Književno djelo .

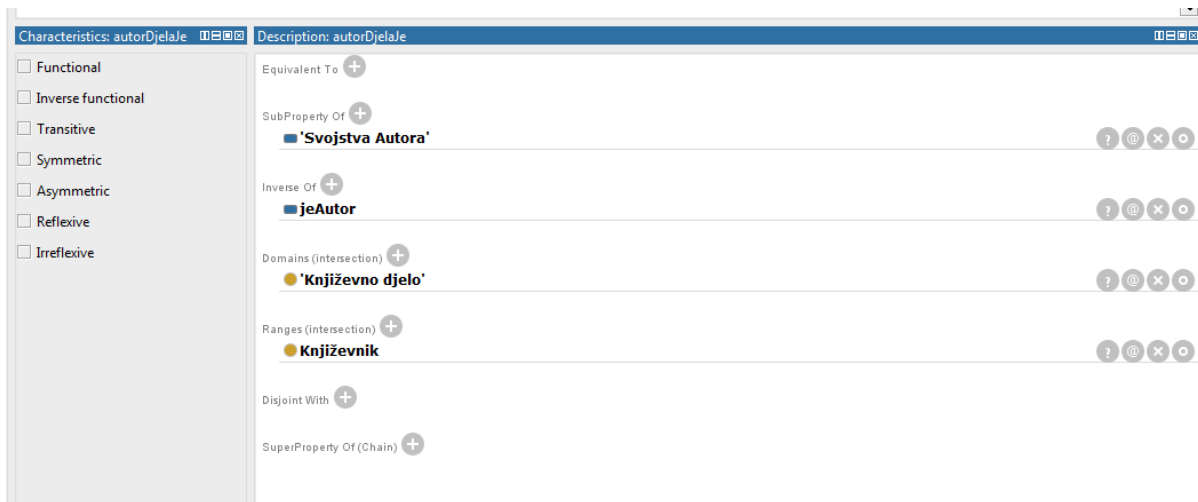
Iako individue *res:Barac, Antun* i *res:Bijeg od knjige* nigdje nisu prethodno dodijeljene na standardni način u odgovarajuće klase, samim korištenjem odnosa *rdfs:domain* i *rdfs:range* RDFS procesor automatski dodjeljuje resurs *res:Barac, Antun* u klasu *res:Književnik*, a resurs *res:Bijeg od knjige* u klasu *res:Književno djelo*. Slika 17 ilustrira takav primjer.



Slika 17: Indirektno svrstavanje resursa u klase

Osim definiranja vrsta pojedinih odnosa, OWL omogućuje i dodjeljivanje karakteristika odnosima svrstavajući ih u određene klase kako bi OWL procesor mogao izvući dodatnu semantiku i povezanost iz RDF trojaca. Tako postoji odnos *owl:inverseOf* i klase *owl:TransitiveProperty*, *owl:SymmetricProperty*, *owl:FunctionalProperty* i *owl:InverseFunctionalProperty*. U Protege-u odabirom kartice odnosi *Object Properties*

postoje mogućnosti određivanja karakteristika odnosima *Functional*, *Inverse functional*, *Transitive*, *Symmetric*, *Asymmetric*, *Reflexive* i *Irreflexive* (Slika 18), dok u kartici odnosi *Data Properties*, postoji samo *Functional*. U ovim je karakteristikama također moguće odrediti *rdfs:domain* i *rdfs:range* odnosa.



Slika 18: Definiranje dodatnih karakteristika odnosa u Protégé-u

Odnosom *owl:inverseOf* je moguće nekom odnosu dodijeliti inverzni odnos. Primjerice, neka postoje odnosi *pro:jeAutor* i *pro:autorDjelaJe* pomoću kojih se povezuje određeni književnik s njegovim napisanim književnim djelom, i obrnuto. Zbog imenovanja samih odnosa u navedenom primjeru čovjek može pretpostaviti odnos između njih, no računalo ne može. Stoga, *owl:inverseOf* čini odnos između *pro:jeAutor* i *pro:autorDjelaJe* jasnim te precizno opisuje njegovo značenje. Odnosu *pro:jeAutor* se inverzni odnos *pro:autorDjelaJe* definira na sljedeći način:

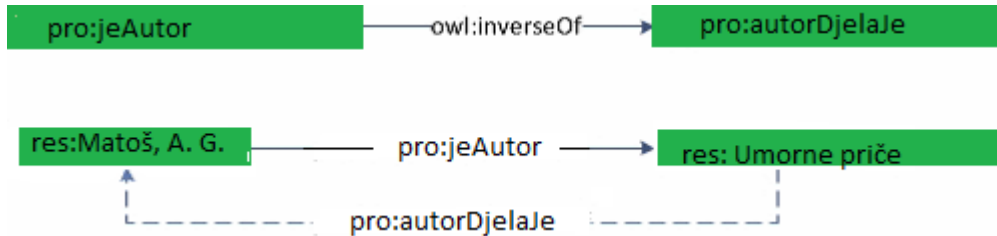
pro:jeAutor owl:inverseOf pro:autorDjelaJe.

Prethodno objašnjeno definiranje inverznog odnosa s *owl:inverseOf* kod zapisanog RDF trojca u sljedećem primjeru govori da je *res:Matoš, Antun Gustav* autor od *res:Umorne priče*, te omogućuje OWL procesoru zaključiti i da je *res:Umorne priče* djelo književnika *res:Matoš, Antun Gustav*, iako to nigdje u bazi nije direktno zapisano preko RDF trojca:

res: Matoš, Antun Gustav pro:jeAutor res: Umorne priče .

res: Umorne priče pro:autorDjelaJe res: Matoš, Antun.

Slika 19 ilustrira definiranje *owl:inverseOf* odnosa i njegovu funkciju između odnosa *pro:jeAutor* i *pro:autorDjelaJe*.



Slika 19: Definiranje inverznog odnosa

Odnos *owl:inverseOf* povezuje jedan odnos s drugim. Poseban slučaj u kojemu su ova dva odnosa zapravo jedan te isti odnos je dovoljan da OWL jezik za njega osigura posebno ime *owl:SymmetricProperty*. Suprotno od inverznog *owl:inverseOf* odnosa, postoji simetrična klasa *owl:SymmetricProperty* u kojoj je moguće definirati simetrične odnose. Za razliku od *owl:inverseOf* odnosa koji se odnosi na dva odnosa, *owl:SymmetricProperty* je zapravo samo jedno gledište istog svojstva te se u OWL-u izražava kao klasa. Izražava se da je odnos simetričan na isti način na koji se izražava članstvo u bilo kojoj klasi. Simetrični odnos je jedna od glavnih značajki OWL-a, a njime se vežući subjekt sa objektom RDF trojca ujedno povezuje objekt sa subjektom preko iste te relacije. Primjerice neka postoji odnos *pro:jeSuautor* pomoću kojih se povezuje jedan književnik s nekim drugim književnikom s kojim je zajedno pisao neki rad, tada također isti taj odnos povezuje tog drugog književnika s prvim. Simetrični odnos se definira na sljedeći način:

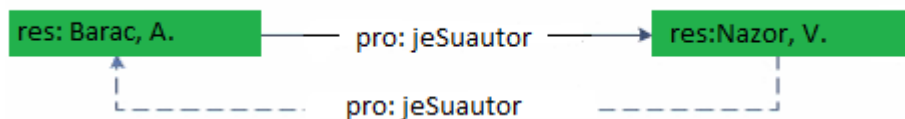
pro:jeSuautor *rdf:type* *owl:SymmetricProperty*.

Prethodno objašnjeno definiranje simetričnog odnosa s *owl:SymmetricProperty* kod zapisanog RDF trojca u sljedećem primjeru govori da je *res:Barac, Antun* suautor od *res:Nazor, Vladimir*; te omogućuje OWL procesoru zaključiti i da je *res:Nazor, Vladimir* suautor književnika *res:Barac, Antun*, iako to nigdje u bazi nije direktno zapisano preko RDF trojca:

res: Barac, Antun *pro:jeSuautor* *res: Nazor, Vladimir*.

res: res: Nazor, Vladimir *pro: jeSuautor* *res: Barac, Antun*.

Slika 20 ilustrira definiranje *owl:SymmetricProperty* odnosa i njegovu funkciju u odnosu *pro: jeSuaautor*.



Slika 20: Definiranje simetričnog odnosa

owl:TransitiveProperty je klasa u koju se dodaju odnosi koje OWL procesor prepoznaje kao prijenosne. Klasa *owl:TransitiveProperty* je podklasa od *owl:ObjectProperty*, stoga *rdfs:range* prijenosnog svojstva može biti samo resurs. Prijenosno ili tranzitivno svojstvo može se objasniti na sljedeći način: ako postoji *Resurs1* i poveže se sa *Resursom2* pomoću odnosa *T*, a *Resurs2* se poveže sa *Resurs3* istim odnosom *T*, onda je i *Resurs1* povezan sa *Resursom3* tim istim odnosom *T*. U tom slučaju odnos *T* je tranzitivan odnos. Primjerice, ako imamo odnos *pro:jePrijevod* koji govori o tome kako je određeno djelo prijevod nekog djela, taj odnos može omogućiti uspostavljanje veze prevedenog djela (prema nekom drugom prijevodu izvornika) s izvornim djelom (izvornikom). Dakle, neka postoji odnos *pro:jePrijevod* koji je definiran kao prijenosni ili tranzitivni odnos:

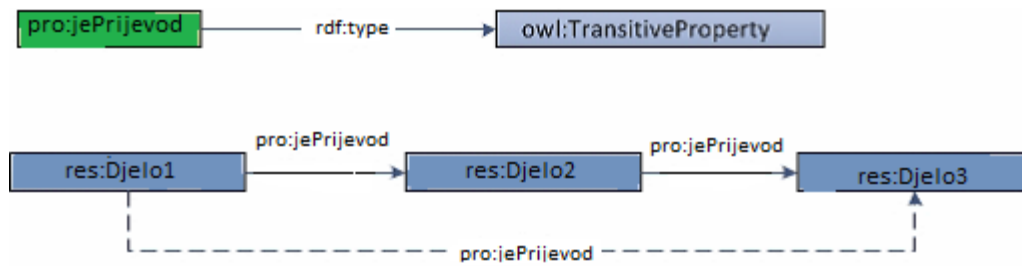
```
pro:jePrijevod    rdf:type    owl:TransitiveProperty .
```

Preko odnosa *pro:jePrijevod* moguće je povezati resurs *res:Djelo1* i resurs *res:Djelo2*, te isto tako resurs *res:Djelo2* i resurs *res:Djelo3*:

```
res:Djelo1 pro:jePrijevod res:Djelo2 .
```

```
res:Djelo2 pro:jePrijevod res:Djelo3 .
```

Ovime tranzitivnim odnosom indirektno je povezan podatak da je resurs *res:Djelo1* vezan preko odnosa *pro:jePrijevod* sa resursom *res:Djelo3*, odnosno OWL procesor prepoznaje kako je *res:Djelo1* ustvari indirektni prijevod *res:Djelo3* iako to nigdje nije direktno zapisano preko RDF trojca. Slika 21 ilustrira takvu vezu.



Slika 21: Definiranje tranzitivnog svojstva

owl:FunctionalProperty je podklasa od *rdf:Property* te prema tome *rdfs:range* funkcionalnog odnosa za razliku od simetričnog i tranzitivnog odnosa može biti resurs, literal ili datatype. Dodajući odnos u klasu *owl:FunctionalProperty*, taj odnos može biti povezan sa maksimalno jednom vrijednosti, odnosno maksimalno jednim resursom. Neka je odnos *pro:jeGodinaRođenja* svrstan u klasu *owl:FunctionalProperty*:

pro:jeGodinaRođenja *rdf:type* *owl:FunctionalProperty* .

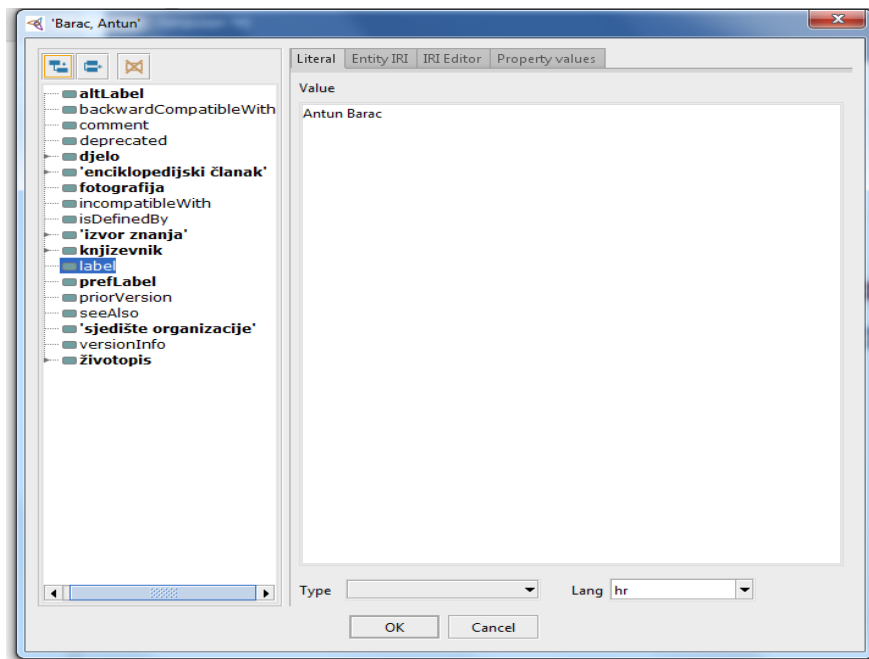
Tako funkcionalno definirani odnos *pro:jeGodinaRođenja* može sadržavati samo jednu vrijednost, tj. *res:Matoš, Antun Gustav* *pro:jeGodinaRođenja* "1873"^^int.

Postoji i klasa *owl:InverseFunctionalProperty* koja neće biti korištena u ontologiji književnosti, ali ju također treba spomenuti. Odnos unutar te klase ima isto svojstvo kao i odnos unutar *owl:FunctionalProperty*, ali sa inverznom karakteristikom, što znači da resurs na mjestu objekta može biti povezan maksimalno sa jednim subjektom.

4.5.2 Opisni odnosi

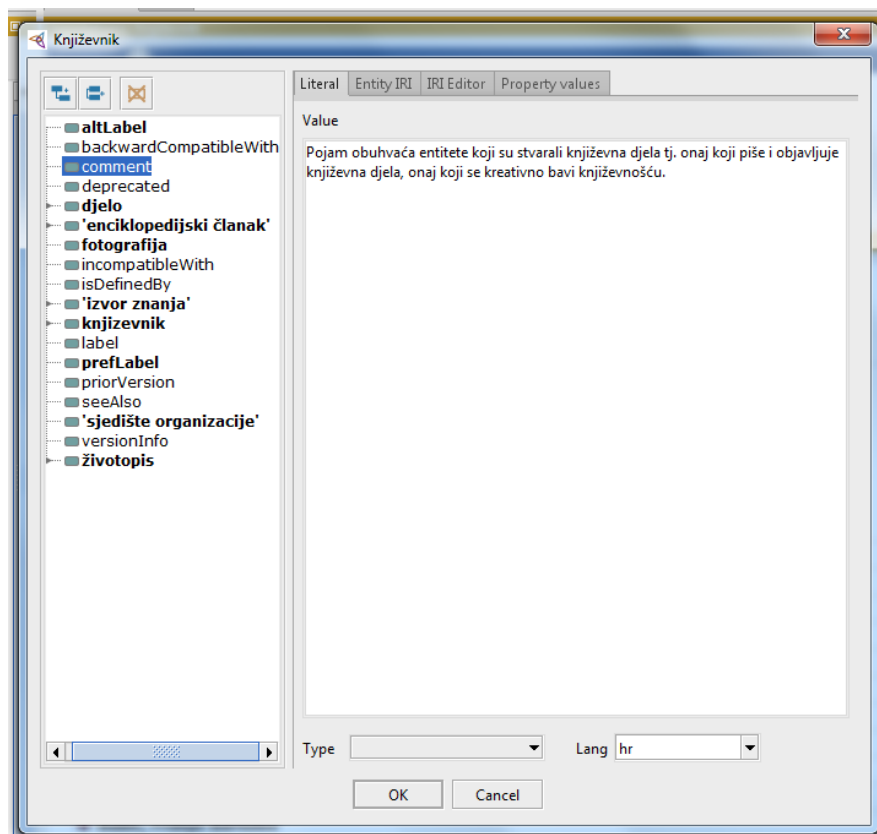
Nakon stvaranja klasa, individua i odnosa, istima je potrebno dodati lako čitljiva imena za čovjeka, odnosno doslovne vrijednosti preko *rdfs:label* i *rdfs:comment* RDFS odnosa. Navedeni odnosi se koriste za naziv (*rdfs:label*) i opis resursa (*rdfs:comment*) čineći ih lako čitljivim ljudima jer se iz URI-a ponekad ne može saznati koji točno pojam predstavlja. Objekti ovih dviju relacija su doslovne vrijednosti i njima je poželjno uvijek dodijeliti jezik na kojem su napisane, a algoritmi i ljudi obično traže resurse upravo preko tih relacija.

Klikom na gumb *Annotations* u Protege programu otvara se novi prozor u kojem je potrebno odabrati *rdfs:label*, te pod *Value* upisati doslovnu vrijednost. Pri tome je moguće odrediti i jezik na kojem je napisana (hr), te vrstu doslovne vrijednosti. Slika 22 prikazuje taj prozor.



Slika 22: Dodavanje opisnog odnosa naziva (*rdfs:label*) u Protégé-u

Na isti način, odabirom relacije *rdfs:comment* upisuje se komentar o resursu, odnosno dodatno se opisuje resurs na čovjeku razumljiv način. Slika 23 prikazuje taj prozor.



Slika 23: Dodavanje opisnog odnosa komentar (rdfs:comment) u Protégé-u

Prepoznajući potrebu za kontroliranim rječnikom terminologije koja se pojavljuje u ontologiji književnosti, individue (owl:NamedIndividual) se u ontologiji mogu redefinirati usporednom upotrebom SKOS rječnika. Ova mogućnost dozvoljava bolje definiranje semantičkih odnosa između termina. Ovaj postupak se u ontologiji književnosti može ostvariti pridodavanjem dodatnog *rdf:type skos:Concept* individui, dok ona i dalje ostaje član izvorne ontologijske klase te se na taj način resurs (individua) može identificirati kao član višestrukih klasa uključujući i *owl:Class* i *skos:Concept*, kao što je prikazano na Slici 24.

```

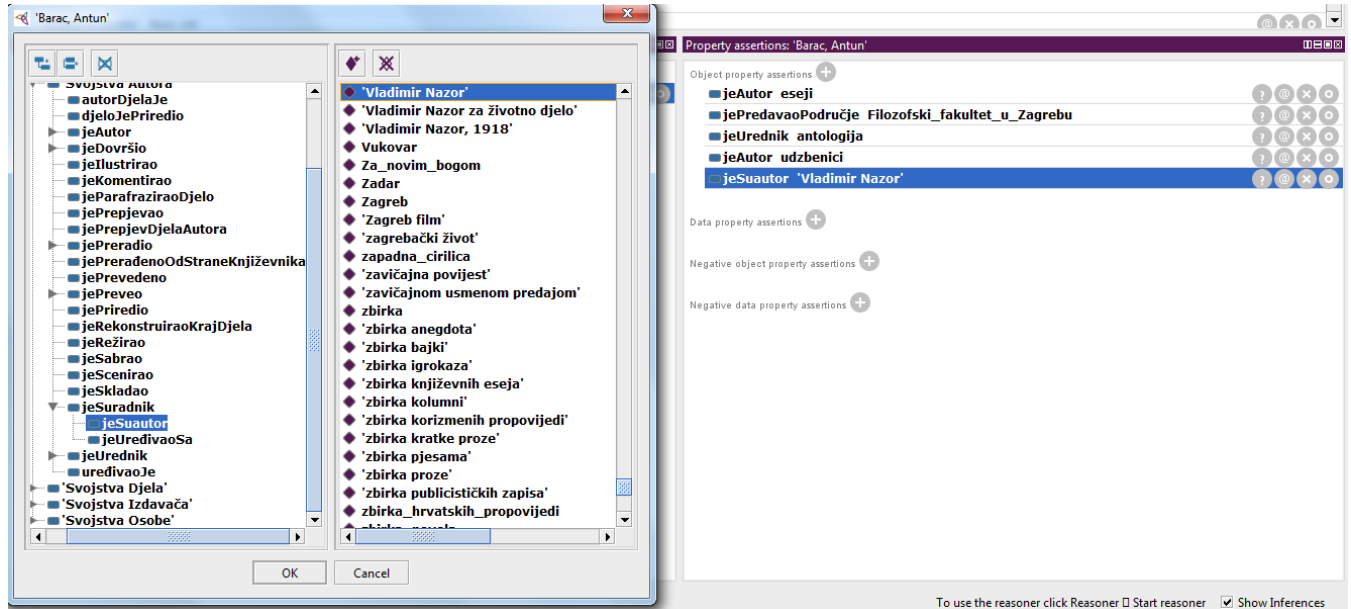
<Named Individual rdf:about="http://knjizevnost.org/vocab/Pripovijetka">
  <rdf:type rdf:resource="knj:Proza"/>
  <rdf:type rdf:resource="skos:Concept"/>
  <rdfs:label xml:lang="en">Pripovijetka</rdfs:label>
</Named Individual>

```

Slika 24: Pripadnost individue višestrukim klasama

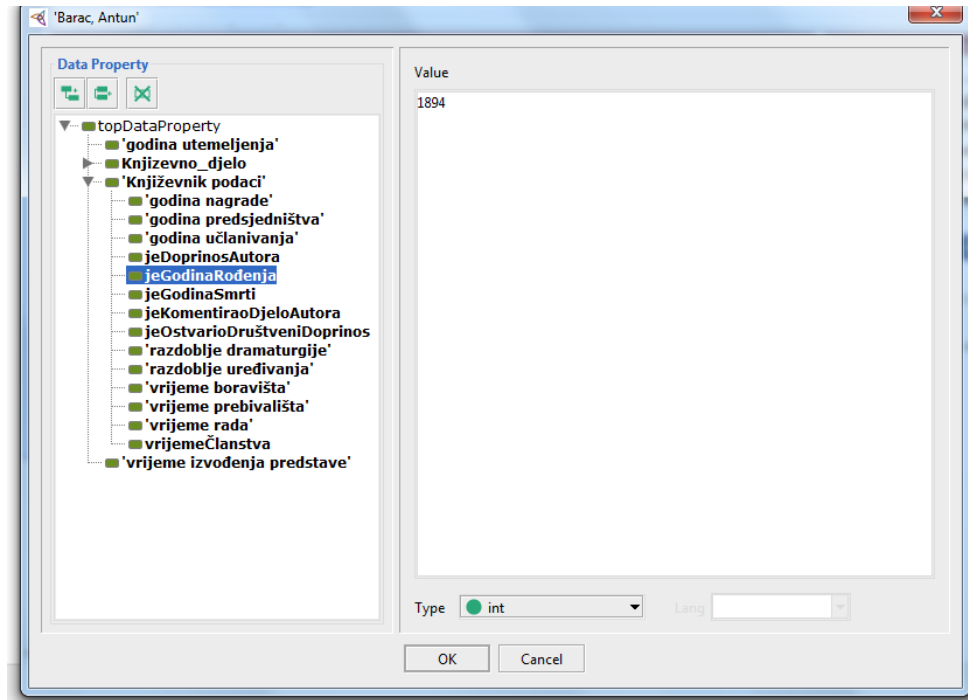
Individui *pripovijetka* se u ontologiji mogu dodijeliti i npr. sljedeći termini *humoreska*, *kratka priča*, *pripovijest*, *priča*, *uokvirena pripovijetka*. Stoga se za deskriptor može odrediti *skos:prefLabel* odnosom termin *pripovijetka* (*skos:prefLabel "pripovijetka"@hr*), a ostali nazivi mogu predstavljati alternativne termine za deskriptor *pripovijetka* koristeći *skos:altLabel* odnos (*skos:altLabel "kratka priča"@hr*, *skos:altLabel "humoreska"@hr*, *skos:altLabel "pripovijest"@hr*, *skos:altLabel "priča"@hr*, *skos:altLabel "uokvirena pripovijetka"@hr*).

Nakon dodjeljivanja čovjeku čitljivih naziva te određivanja naziva deskriptora i alternativnih termina, slijedi postupak povezivanja podataka i dodjeljivanja doslovnih vrijednosti. U kartici *Individuals* u Protege programu odabirom pojedine individue moguće ju je povezati sa drugom individuom ili joj dodijeliti doslovnu vrijednost. Nakon odabira individue, u dijelu prozora *Property assertions* klikom na gumb *Object property assertions* otvara se novi prozor u kojem se odabire odnos i individua sa kojom se želi povezati odabrana individua. Slika 25 prikazuje taj postupak.



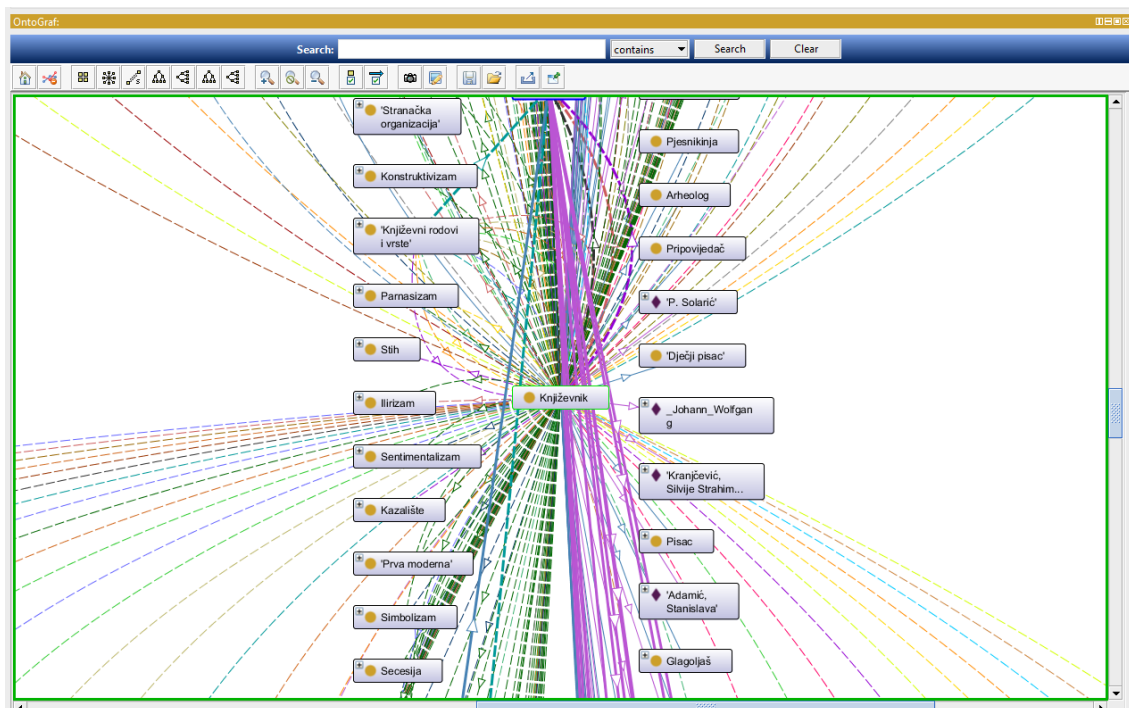
Slika 25: Povezivanje individua u Protégé-u

Na isti način klikom na gumb *Data property assertions* dodaje se odnos i upisuje se doslovna vrijednost koja se želi povezati sa odabranom individuom. Slika 26 prikazuje taj prozor.



Slika 26: Dodavanje doslovnih vrijednosti individui u Protégé-u

Za vizualizaciju ontologije se može koristiti već ugrađena mogućnost OntoGraf-a u Protégé programu ili se može instalirati jedan od dodataka kao što je OWL Viz. Slika 27 prikazuje vizualizaciju pomoću OntoGraf-a.



Slika 27: Vizualizacija ontologije pomoću OntoGraf-a

Značajno je napomenuti kako Protégé također ima ugrađene neke od procesora za zaključivanje odnosno *reasoner*. Aktivacijom jednog od njih Protégé povezuje podatke koji nisu direktno definirani u ontologiji, već su novi odnosi i RDF trojci izvedeni iz postojećih semantičkih veza.

Dobivenu ontologiju je moguće spremirati preko neke od ponuđenih načina serijalizacije. Preporuka je *Turtle* format jer je jednostavan i najrazumljiviji čovjeku, dok je s druge strane kompatibilan sa svim ostalim aplikacijama.

4.6 Jednakost i nejednakost

Pomoću OWL odnosa mogu se definirati iste klase, relacije i individue koje se pojavljuju unutar vlastite baze i drugih baza, odnosno između različitih ontologija te na taj način postići interoperabilnost dobivene ontologije književnosti Hrvatske enciklopedije s već postojećim ontologijama na internetu (npr. DBpedija, Freebase, Yago i sl.).

Prvi i osnovni odnos kojim je moguće definirati da su dvije individue različitog imena ustvari isti resurs je *owl:sameAs*. Neka postoje dva različita URI-a književnika Antuna Barca: *res:Barac, Antun* i *res:Antun_Barac(Književnik)*. Pomoću relacije *owl:sameAs* moguće je reći OWL procesoru da se ustvari radi o istom resursu. Tada je moguće izvući sve ostale odnose vezane uz jedan i drugi URI, te ih prikazati kao zajedničke:

```
res: Barac, Antun    owl:sameAs    res: Antun_Barac(Književnik).
```

Kada je potrebno striktno definirati različitost individua koristi se relacija *owl:differentFrom*, kojom se naglašava kako se ne radi o istom resursu. Tako je na primjer moguće razlikovati resurs *res: Babić, Ante* koji je publicist i *res: Babić, Ante* koji je ekonomist.

```
res: Babić, Ante    owl:differentFrom    res: Babić, Ante .
```

Za definiranje jednakosti klasa koristi se relacija *owl:equivalentClass*, dok se za definiranje jednakosti odnosa koristi se *owl:equivalentProperty*. Treba napomenuti da se ovdje ne deklariraju klase kao isti pojmovi, već klase sa istim individuama, a u stvarnom svijetu mogu biti dva različita pojma, odnosno resursa.

res:Književnici owl:equivalentClass res:Pisci .
pro:jeautor owl:equivalentProperty pro:jenapisao .

Kao što je već navedeno, ovi OWL odnosi omogućuju uspostavljanje interoperabilnosti ontologije književnosti Hrvatske enciklopedije s velikim npr. DBpedia (Yago, Freebase i sl.) projektima semantičkog weba:

- (a) res: *Barac, Antun* owl:sameAs dbpedia: *Antun_Barac*,
- (b) res:Književnici owl:equivalentClass dbpedia:Writer
- (c) pro:jeautor owl:equivalentProperty dbpedia:create

U navedenom primjeru uspostavljanja interoperabilnosti ontologije hrvatske književnosti s DBpedia projektom uočljivo je uspostavljanje jednakosti između dva različita resursa tj. jednakosti na razini individua pomoću *owl:sameAs* odnosa koje mogu pripadati klasama različitih ontologija, dok se uspostavljanje jednakosti između klasa i odnosa dviju različitih ontologija primjenom OWL odnosa *owl:equivalentProperty* i *owl:equivalentClass* uspostavlja jednakost tj. interoperabilnost (odnosno mapiranje, pridruživanje) dviju ontologija. Dakle, upotrebom odnosa (a) uspostaviti će se jednakost resursa tj. članaka Hrvatske enciklopedije s Wikipedijinim člancima (tj. DBpedijinim resursima), a primjenom odnosa (b) i (c) uspostaviti će se jednakost ontologije književnosti dobivene iz Hrvatske enciklopedije s DBpedia ontologijom dobivenom iz Wikipedije.

4.7 Verifikacija ontologije

U početku razvoja ontologija Protege software je imao poseban modul za verifikaciju ontologija, tzv. OntoClean modul. Razvojem alata verifikacija se ugradila u njihovu postojeću strukturu te se sada automatski provjerava ispravnost ontologije.

Isto se ugradilo u Virtuoso Triple Store software koji automatski provjerava ontologiju i ne dopušta da se izvode neispravne ontologije. Stoga će dobivena ontologija biti automatski verificirana u Protege alatu, kao i u Virtuoso Triple Store software-u.

4.8 Zaključak

Poglavlje je dalo kratki uvod u primijenjene standarde semantičke mreže (tj. rječnike RDFS, OWL, SKOS) prilikom izgradnje ontologije književnosti Hrvatske enciklopedije, čime će se dobivena ontologija uklopiti i učiniti sastavnim djelom semantičke mreže podataka. Ukratko su se prikazali temeljni postupci izgradnje ontologije u Protege programu te se ukazalo na mogućnost postizanja interoperabilnosti kako njezinih resursa pojedinačno, tako i cjelokupne ontologije sa postojećim semantičkim ontologijskim projektima (DBpedia, Yago, Freebase itd.). Uspostavljanje interoperabilnosti ontologije književnosti s ostalim projektima na semantičkoj mreži omogućit će povezivanje relevantnog i bogatog znanja *Hrvatske enciklopedije* u „globalnu mrežu znanja“ koja nastaje i razvija se kroz projekte semantičke mreže.

5 ENCIKLOPEDIČKA DJELA

Poglavlje *Enciklopedička djela* dati će osnovne informacije o povijesnom razvoju enciklopedije u svijetu, razvoju hrvatske središnje leksikografske ustanove Leksikografski zavod „Miroslav Krleža“ (LZMK), izdavačkoj djelatnosti općih i strukovnih enciklopedijskih izdanja LZMK obuhvaćajući izdanja na papirnatom mediju i ona u mrežnom okruženju.

5.1 Povijesni razvoj enciklopedije u svijetu

Enciklopedija (prema grč. *ἐγκύκλιος*: koji ide redom; opći + -pedija) je djelo u kojem se, abecednim ili kakvim drugim metodičkim slijedom, okupljaju i sustavno obrađuju činjenice i spoznaje o svim ljudskim znanjima (opća enciklopedija) ili pak sva građa jedne znanosti, umjetnosti, područja (strukovna enciklopedija ili posebna enciklopedija). Nezaobilazni prethodnici i začetnici enciklopedijske literature su antički leksikografi i glosografi. Oni su već polovicom 3. st. pr. Kr. obavili dvije nužne predradnje: izdvajanje pojedinih imena i natuknica i njihovo alfabetsko (ili abecedno) raspoređivanje. U bizantskom razdoblju, uz bogate glosografske priručnike, nastaju i sustavni popisi geografskih imena, kada dolazi do pojačana zanimanja za etimologijske popise, koji nastaju oko 9. st. Focije u 9. st. u svojem *Leksikonu* pokazuje izrazite enciklopedičke odlike. Međutim, sam središnji pojam *enciklopedija* ne susreće se ni u staroj grčkoj tradiciji, kao ni u rimskoj latinskoj. Pojam *enciklopedija* u značenju „sustavna i opća naobrazba“ vjerojatno potječe od Hipije iz Elide, sofista iz 5. st. pr. Kr. Iako antika nije poznavala pojam enciklopedije u novijem smislu, neki povjesničari kulture drže da je već i Aristotelovo cjelokupno filozofsko djelo u svojoj biti enciklopedijsko. Stari su se Rimljani, u svojim djelima praktične naravi namijenjenima djeci za udžbenike, dobro približili novijemu pojmu enciklopedije. To potvrđuju djela latinskih pisaca kao što su Katon i Varon, pisci o najrazličitijim područjima (od poljodjelstva i ratnoga umijeća do liječništva i govorništva). Djelo Plinija Starijega (1. st.), *Prirodoznanstvo* (*Historia naturalis*) u 37 knjiga je primjerno enciklopedijsko djelo, u kojem je zahvaćeno obilje područja, od zvjezdoznanstva i botanike do ljekarništva i antropologije.

Europsko srednjovjekovlje nastavlja tradiciju prije svega latinskih enciklopedijskih djela. Istina, to je manje od modernoga poimanja enciklopedije kao sustava, razredbe i razradbe različitih znanstvenih područja, ali i nešto više od polazne poučne obradbe svega što čovjek može naučiti u svojem životu.

Marcijan Kapela (*Martianus Minnaeus Felix Capella*) iz Kartage (prva polovica 5. st.) sastavio je enciklopedijski roman u prozi i stihu, koji se inače javlja i pod naslovom *Satyricon*, ali uglavnom kao *Svadba Merkura i Filologije* (*De nuptiis Mercurii et Philologiae*). To je djelo u 9 svezaka, koje je po svojoj naravi alegorijsko i mitologijsko, dijelom i poučno, ali značajem ipak enciklopedično.

Izidor Seviljski (570–636) bio je seviljski biskup, koji je ostavio nedovršeno svoje djelo *Knjiga etimologijâ* (*Liber ethimologiarum*), posmrtno je podijeljeno u 20 knjiga i objavljeno. U literaturi se to djelo naziva *Etimologije* (*Etymologiae*) ili *Izvori*, doslovno *Podrijetla* (*Origines*). Djelo nije tipičan jezikoslovni priručnik, nego prije suma sveukupnog antičkog znanja s različitih područja. Tako je tu i ljekarništvo i liječništvo, ali i 7 slobodnih (ili slobodnjačkih) umijeća, te pravo, crkva, jezik, ljudi, životinje, elementi, kovine, ratne vještine, poljodjelstvo, povrtlarstvo, jela i pića. Za to se Izidor služio i *Biblijom*, a i klasičnim piscima Plinijem i Vergilijem te Petronijem i Augustinom. *Etimologije* su zato tipična srednjovjekovna enciklopedija, koja je bila glavni prenositelj rimske baštine čitateljima europskoga srednjovjekovlja.

Hraban Maur (Hrabanus Magnentius Maurus, 780–856) bio je Izidorov sljedbenik, koji se u svojem glavnom djelu *O svijetu* (*De Universo*, u 22 knjige) uvelike služio svojim prethodnikom ispustivši mnoge antičke jedinice, a dodajući mnoge svetopisamske natuknice. Cijenjen zbog zasluga za širenje znanja među istočnim Francima (a bio je i Alkuinov učenik), stekao je naziv *Praeceptor Germaniae*. Iako je njegovo djelo kompilatorske naravi, taj je franački teolog iz Fulde po svojem djelu i utjecaju jedan od najistaknutijih srednjovjekovnih enciklopedičkih pisaca. Njegovo glavno enciklopedijsko djelo poznato je još i pod trima drugim naslovima: *O podrijetlu stvari* (*De origine rerum*), *O sveopćoj naravi* (*De universali natura*) i *O naravi stvari* (*De natura rerum*). Poznavao je dobro i grčki, hebrejski i sirski.²⁸⁶

Vincent od Beauvaisa (1190–1264) smatra se sastavljačem najpoznatije srednjovjekovne enciklopedije. Njegovo glavno djelo se zove *Veće zrcalo* – *Speculum maius* (ili *Knjižnica svijeta* – *Bibliotheca mundi*). Podloga njegova djela je nekih 2000 različitih izvora, a ono je po strukturi trodijelno: *Naravno zrcalo* (*Speculum naturale*), *Zrcalo nauka* (*Speculum doctrinale*) i *Povijesno zrcalo* (*Speculum historiale*). Kako je u njemu sustavno obrađena prirodna znanost te govorništvo, pravo i povijest, smatra se najvećim enciklopedičkim priručnikom 13. stoljeća, uz četvrti dodatni dio *Čudoredno zrcalo* (*Speculum*

²⁸⁶ usp. Enciklopedija // Hrvatska enciklopedija: mrežno izdanje. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=17879> (5.06.2016.).

morale). Djelo je inače potkrijepljeno mnogobrojnim navodima i izvodima iz arapskih i židovskih pisaca uz grčke i latinske.

Talijanski pisac i Danteov učitelj Brunetto Latini (1220–94) ubraja se među najistaknutije enciklopedičke pisce po svojem djelu *Knjiga riznice* ili *Riznica* (*Li Livres dou Trésor*), pisan na starofrancuskom jeziku u 3 dijela. U tom je djelu ne samo enciklopedijska obradba biblijskoga predmetnika, već se obrađuju i povijest, etika, zvjezdoznanstvo i zemljopis.

Bartholomaeus Anglicus (Bartholomew the Englishman ili Bartolomej Englez, 1230–50), engleski franjevac, u 13. st sastavio je svojevrsnu enciklopediju o značajkama ili svojstvima pod nazivom *O svojstvima stvari* (*De proprietatibus rerum*). Nije doživio tiskanje svojega djela. Djelo je bilo prevedeno na engleski 1398., dok se latinski pretisak pojavio tek 1470. Ipak je prije 1500. to djelo doživjelo čak desetak izdanja, zbog neobičnosti u izboru predmeta i načinu obradbe, jer se bavilo i otajstvenom naravi anđela, kao i svojstvima alkoholnih napitaka.

Gregor Reisch, katolički teolog i filozof, objavio je 1503. u Freiburgu vrlo popularno enciklopedijsko djelo (*Filozofski biser – Margarita philosophica*), koje se bavi *trivijem* i *kvadrivijem*, podrijetlom naravnih pojava, ali i kat. ćudorednicom.

Raphael Maffei, talijanski erudit, u 16. st. je u Rimu pisao enciklopediju *XXXVIII knjiga bilježaka o gradskim stvarima* (*Commentariorum rerum urbanorum libri XXXVIII*, 1506), koja zauzima iznimno mjesto i ima posebnu važnost u novijoj enciklopedici upravo po tome što iznimnu pozornost posvećuje novovjekim zemljopisnim otkrićima i sustavnoj obradbi životopisa istaknutih pojedinaca.

U razdoblju prije pojave glasovitih enciklopedista posebno mjesto ima Johann Heinrich Alsted (1588–1638), njemački teolog i filozof, koji je djelovao je kao profesor i pedagog. Najvažnije mu je djelo *Enciklopedija razlučena u sedam svezaka* (*Encyclopaedia septem tomis distincta*, 1630), koje je naišlo na izniman prijam kod čitateljstva. Njegova enciklopedija ne zaobilazi nijedno važnije područje prirodnih i duhovnih znanosti te različitih obrta i umijeća, ali je imala i posebne dodatke, npr. paradoksologiju (umijeće tumačenja paradoksa), dipnosofistiku (umijeće gozbenog umovanja) te čak i tabakologiju (kao nauk o naravi, porabi i zloporabi duhana).

U 17. st. pojavilo se nekoliko enciklopedijskih rječnika na kulturnom području, što u stihu što pak u prozi, što dovršenih što opet nedovršenih, ali ti priručnici imaju više kulturno-povijesno značenje. Neka od tih djela bila su vrlo ambiciozna po nakani, kao npr. *Opći*

povijesno-zemljopisno-kronološko-pjesničko-filološki leksikon (Lexicon universale historico-geographico-chronologico-poëtico-philologicum) baselskoga profesora povijesti J. J. Hofmanna (1635–1706), ali ne i po ostvarenju, zbog površne i netočne obradbe uključenih natuknica. Za razliku od takva djela, talijanski franjevac, erudit i kozmograf Vincenzo Maria Coronelli (1650–1718), koji je veći dio života proveo u Veneciji (a bio je general reda Male braće), započeo je i rad na prvoj enciklopediji sa sustavnim abecednim rasporedom građe. Trideset je godina radio na tome djelu (naslovljenome *Crkveno-svjetovna opća knjižnica – Biblioteca universale sacro-profana*), koje je bilo predviđeno u 45 svezaka, ali ih je bilo objavljeno tek 7. Pa ipak je to važno djelo i po primjeni živoga modernog talijanskoga jezika. Nažalost, i ta je važna enciklopedija u povijesno-razvojnome smislu bolovala od površnosti i netočnih podataka.

Razdoblje prosvjetiteljstva ponekad se čak poistovjećuje s enciklopedijom, pa se i opisuje kao doba ili razdoblje enciklopedije, a samo se stoljeće smatra stoljećem velike enciklopedijske revolucije. Osamnaesto je stoljeće nazvano „dobom enciklopedije“²⁸⁷, „enciklopedijskim stoljećem“²⁸⁸ ili čak i zlatnim razdobljem europske enciklopedike. Proizvodnjom i širenjem znanja nastaju novi problemi. Jedan od njih je kako upravljati brzo rastućom količinom javne informacije i znanja, kako je selekcionirati, klasificirati i pohraniti. Pojava periodične literature (novina i časopisa) dodatno je pojačala percepciju preobilja. Sve se više čita pa se 18. stoljeće smatra „revolucijom čitanja“²⁸⁹. Rolf Englesing napravio je razlikovanje „intenzivnoga“ i „ekstenzivnoga“ čitanja²⁹⁰. Za razliku od intenzivnoga čitanja, zatvorenoga i ponavljajućega čitanja maloga broja knjiga (npr. Biblije), ekstenzivno čitanje označava brzo konzultiranje ili letimično čitanje velikoga broja tekstova. Taj drugi način u 18. stoljeću postupno zamjenjuje prvi. Prekomjernost je stvorila problem usvajanja znanja i promijenila čitalačke navike, a različite vrste rječnika, leksikona i enciklopedija bile su rješenje toga problema. Te su knjige istodobno mogle potaknuti ekstenzivno konzultiranje²⁹¹ jer nisu bile priređene za cjelovito čitanje nego za brzo konzultiranje određenoga dijela (otud prevlast abecednoga načela uređenja građe). Istodobno se zbilila velika promjena u prirodi i

²⁸⁷ Yeo, Richard. *Encyclopaedic Vision: Scientific Dictionaries and Enlightenment Culture*. Cambridge: Cambridge University Press, 2001. Str. 277.

²⁸⁸ Barzun, Jacques. *Od osvita do dekadencije: 500 godina zapadne kulture*. Zagreb: Masmedia, 2003. Str. 429.

²⁸⁹ Burke, Peter. *A Social History of Knowledge: From Gutenberg to Diderot*. London: Polity Press u suradnji s Blackwell Publishers, 2000. Str. 86

²⁹⁰ Englesing, Rolf. *Der Bürger als Leser. Lesergeschichte in Deutschland, 1500–1800*. Stuttgart: J. B. Metzlersche i A. E. Poeschel, 1974. Str. 113.

²⁹¹ Burke, Peter, op. cit. (bilj. 289), str. 182.

strukturi znanja, koja je rezultat znanstvene revolucije 18. stoljeća. U tom smislu velik znanstveni preokret 18. stoljeća bio je uvod u enciklopedijsku revoluciju 18. stoljeća. Baconov nacrt u *Instauratio magna* (1620) uvodi posve novu klasifikaciju znanja i perspektivu enciklopedije. Leksikoni i enciklopedije 18. stoljeća nastojat će te rezultate usvojiti te su najvažniji od njih isticali potrebu da se, unatoč logici abecednoga reda, njihova građa shvaća kao cjelovita i jedinstvena, pa se one trebaju pažljivo, čak sustavno čitati. Enciklopedije jednostavno nisu htjele biti samo izvor prve obavijesti, nego načini na koje informacija može biti uređena i obuhvaćena kao znanje, te time i sredstvo obučavanja. Jedan od načina da se ta sustavnost i cjelovitost istakne bili su sustavi uputnica i indeksi. Zato su „glavne enciklopedije prosvjetiteljstva težile kompromisu između sustavnog i abecednog uređenja“²⁹². Na kraju 17. stoljeća Brockhaus je počeo objavljivati svoj prvi *Konversations-Lexikon* (1796–1811). U Nizozemskoj, u Amsterdamu, francuski tiskar i zastupnik radikalnoga prosvjetiteljstva Jean-Frédéric Bernard objavio je prvu svjetsku enciklopediju religija u sedam svezaka, *Ceremonije i religijski običaji svih naroda svijeta* (1723–35), s prekrasnim ilustracijama Bernarda Picarta, pa je englesko izdanje (1733–77) te vrlo utjecajne knjige nosilo potpis obojice. Najznačajnije dvije edicije tog vremena su: Chambersova *Enciklopedija* i Zedlerov *Leksikon*. Ephraim Chambers (u. 1740) je objavio i posvetio engleskomu kralju enciklopedijsko djelo u 2 sveska 1728. pod naslovom *Ciklopedija ili opći rječnik umjetnosti i znanosti koji sadržava objašnjenja naziva i opis stvari koje se njima označuju u nekoliko umjetnosti, kako slobodnih tako tehničkih, i nekoliko znanosti, kako ljudskih tako i božanskih* (*Cyclopaedia or an Universal Dictionary of Arts and Sciences containing an Explanation of the Terms and an Account of the Things Signified thereby in the several Arts, Liberal and Mechanical, and the several Sciences, Human and Divine*). Chambers ističe kako njegovo djelo nije uradak jednog uma, nego je proizvod sveukupne zajednice ljudskoga znanja. Njegova je *Ciklopedija* bila dugo vremena veoma popularna. Johann Heinrich Zedler (1706–60), knjižar iz Leipziga, objavio je *Veliki sveobuhvatni opći leksikon svih znanosti i umjetnosti koje jesu do sada ljudski um i umješnost izumili i poboljšali* (*Grosses vollständiges Universal-Lexikon Aller Wissenschaften und Künste welche bisher durch menschlichen Verstand und Witz erfunden und verbessert worden*) koji je izišao je u 64 sveska u razdoblju 1732–50. Na tom je djelu sudjelovao veći broj urednika i suradnika. Zedlerovo djelo poznato je kao do tada najveća i najobuhvatnija enciklopedija, koja je po strukturi golemi enciklopedijski leksikon. U njemu su obuhvaćene humanističke, prirodne i

²⁹² Yeo, Richard, op. cit. (bilj. 287), str. 114.

primijenjene znanosti te biografije istaknutih pojedinaca. I danas je taj univerzalni leksikon pouzdan izvor za mnoge korisne podatke. Osamnaesto stoljeće je proizvelo dva djela po kojima je zapravo kasnije i nazvano zlatnim dobom enciklopedistike – slavnu francusku *Enciklopediju* te njezinu britansku suparnicu *Britansku enciklopediju*. Ipak, nazivi zlatno doba ili doba enciklopedije odnose se ponajprije na prvo od njih, francusku *Enciklopediju*, djelo koje je bilo srž enciklopedijske revolucije. Dok je ona odmah postala slavna i čudesno utjecajna, škotska i britanska, a kasnije američka *Britanska enciklopedija* bit će prihvaćena kao vrijedno djelo tek sa svojim trećim izdanjem (1788–97), a u potpunosti u prvoj polovici 19. stoljeća. Iako nije imala tako snažan početak i utjecaj, *Britanska enciklopedija* postigla je tijekom dugoga razdoblja izlaženja velik enciklopedijski i ideološki utjecaj. Chambersovo djelo (*Ciklopedija*) poslužilo je kao osnova za najpoznatiju među europskim enciklopedijama, za glasovitu *Francusku enciklopediju*, koja je po svojoj važnosti i značenju postala paradigma moderne enciklopedije. Iako je zamišljena kao francuska inačica Chambersove enciklopedije, francuska je *Enciklopedija* otišla u drugom smjeru, koji je postavio Bayleov *Povijesni i kritički rječnik* (1697). Britanska enciklopedija vezala se za enciklopedijsku tradiciju koju obilježavaju Harrisov *Tehnički leksikon* (1704), kao znanstveno-tehničko, ali i humanističko djelo, te vrlo popularna Chambersova *Ciklopedija* (1728). Djelo je proizvod zajedničkoga rada većega broja enciklopedijskih pisaca, koji su u povijesti i ostali pod imenom *enciklopedisti* (kako se npr. nazivaju i petorica najpoznatijih: Rousseau, Voltaire, Diderot, D’Alambert i Helvetius). *Francuska enciklopediju* izvorno je naslovljena *Enciklopedija ili razumski rječnik znanosti, umjetnosti i obrta, sastavljen od udruge učenjaka (Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers, par une société de gens de lettres)*. Njegov je stvarni začetnik i voditelj bio Denis Diderot. Enciklopedija je izlazila od 1751. do 1772, a sama ideja o takvu pothvatu nastala je 1740. Kada je počelo (1745) prevođenje Chambersove enciklopedije na francuski, došlo je do dogovora da se dotični prijevod upotrijebi, ali i uveliko proširi i obogati priložima francuskih autora. Diderot je preuzeo vođenje projekta i uključio u njega francuske znanstvenike, filozofe i književnike. Nakon isusovačkih napadaja prvi su svesci bili i zabranjivani. Zabrana je poslije ukinuta, ali je uvedena cenzura priloga. Zabrane su i dalje slijedile, tako da se *Enciklopedija*, kada je bila dovršena, iako službeno zabranjena, širila cijelom Europom, prenoseći bitne ideje prosvjetiteljstva. Projekt je bio počeo s dvadesetak suradnika, ali je na kraju bio narastao do pedeset izravnih pisaca različitih članaka. Bila su obuhvaćena područja duhovnih znanosti (filozofije, teologije, književnosti i jezika), ali i teorijskih i primijenjenih znanosti (matematike, fizike i ekonomije) te povijest, politika i glazba. Posebnost *Francuske*

enciklopedije bilo je uvrštavanje i sustavna obradba područja rada, obrta, umijeća, tehnike i izuma, uz opise životno važnih prirodnih zakona i znanstvenih metoda. Vjerojatno je u suradništvo na enciklopediji po prvi put bilo uključeno i nekoliko obrtnika, tehničara i inženjera, koji su ili izravno pisali jedinice ili služili kao informanti za posebna radna ili tehnička područja. Popularnosti *Enciklopedije* pridonijelo je i obilje ilustracija vrhunske izradbe i učinkovite preglednosti. *Enciklopedija* je izvorno obuhvaćala 17 folio-svezaka između 1751. i 1765. i 11 svezaka s bakrorezima između 1762. i 1772. Ukupno su obuhvaćali 71 818 članaka i 2885 ilustracija. Tih 28 svezaka, objavljenih između 1751. i 1772., uredio je Diderot. No u enciklopediju se ubraja još sedam svezaka, koje on nije uredio: 1776–77. objavljeno je pet dodatnih svezaka (četiri tekstualna i jedan svezak ilustracija) te 1780. dva sveska indeksa, koje je Diderot odbio urediti. Ipak, s punim pravom, *Enciklopedija* se naziva Diderotovom. Temelj za čvrstu klasifikaciju znanja nju Diderot i d'Alembert našli su u Baconovoj podjeli znanja na tri opće grane, koje odgovaraju trima temeljnim čovjekovim sposobnostima: 1. povijest, koja se zasniva na pamćenju i obuhvaća povijest koja se odnosi na Boga (sveta i crkvena povijest), čovjeka i prirodu; 2. filozofiju, koja se zasniva na mišljenju i obuhvaća znanost o Bogu, čovjeku i prirodi, kao i fizikalno-matematičke znanosti, te 3. lijepe vještine, koje se zasnivaju na mašti i obuhvaćaju slikarstvo, kiparstvo, arhitekturu, pjesništvo i glazbu. Toj tematskoj organizaciji znanja, izraženoj u tradicionalnom drvu znanja, ili „enciklopedijskom načelu“, kako ga naziva d'Alembert, dodano je moderno načelo abecedne organizacije koje on u uvodnoj raspravi naziva „rječničkim načelom“, koje olakšava korištenje enciklopedije. U dva dodatna sveska pod naslovom »*Table analytique et raisonnée*« uvela je opsežan indeks, koji će *Britannica* usvojiti tek u 7. izdanju (1830–42), kad će biti otisnut poseban svezak s indeksom. Sustavi referenciranja (klasifikacijska mjesta i uputnice), koji su osmišljeni kako bi se ublažile slabosti abecedne organizacije, uspostavljajući međupovezanost abecediranjem raspršenih fragmenata znanja, pridonijeli su enciklopedističkoj izuzetnosti *Enciklopedije*²⁹³. Temeljna metoda enciklopedijske obradbe sastojala se od sustavna navođenja dokumentarno-faktografskih sastavnica uz filozofsku i kritičku ocjenu te uz znanstvene i umjetničke potkrjepe. Na području nacionalnoga gospodarstva članci su bili ilustrirani modernim stanjem sela, seljaštva i proizvodnje uopće, a usto je slijedila i razborita te obrazložena kritika suvremenoga poreznog sustava. Zapravo, nije bilo životnoga područja – ma kako ono bilo osjetljivo – a da ga enciklopedisti nisu obrađivali istodobno i dokumentirano i izrazito politički angažirano. Predvođeni Diderotom,

²⁹³ usp. Ravlić, Slaven. *Prosvjetiteljstvo i enciklopedija: dva prosvjetiteljstva i dvije enciklopedije* // *Studia lexicographica*, 7 (2013), 2(13), str. 36-49.

oni su se oborili na suvremeni religijski sustav. Izložili su oštroj kritici sve oblike praznovjerja i mračnjaštva, kojih primjeri nisu nikako bili rijetkost ni u suvremenome društvu i poretku. U osnovi je bila ideja da sva znanost, umjetnost i politika treba imati za cilj sreću, i to što veću sreću za što veći broj ljudi. Prema sudu povjesničara enciklopedizma *Francuska enciklopedija* nije imala za cilj samo informacije o svim područjima znanja i iskustva nego i oslobođenje pojedinca i zajednice od svih predrasuda i nepravda koje priječe ostvarenje njegove naravne i ciljane sreće u slobodi, blagostanju i stvaralaštvu. Naravno, zbog svega toga prilozi u enciklopediji bili su raznovrsni po razini, pa i točnosti; neki su bili suviše feljtonski i polemični, pa samim time i nedostatan leksikografski ili enciklopedijski. Mjestimice se žrtvovala točnost i objektivnost, a bilo je i brzine te previda, ali je u cjelini *Francuska enciklopedija* isprva bila ujedno i riznica znanja i politički pamflet, a i udžbenik za nekoliko naraštaja europskih reformatora i upornih popravljča zatečenih društvenih odnosa.

Francuska enciklopedija nazvana je ratnim strojem; napala je crkvu, despotsku vladavinu, bila je ujedno popularna kao nijedna enciklopedija u europskoj povijesti, ali i proganjana kao nijedna druga u 18. st., velikom dobu enciklopedizma i enciklopedista.²⁹⁴

U istome stoljeću objavljeno je (1771) i prvo izdanje britanske enciklopedije (*Britanska enciklopedija ili Rječnik umjetnosti i znanosti – The Encyclopaedia Britannica or Dictionary of Arts and Sciences*). Prvo izdanje, objavljeno u Edinburghu 1768–71. u tri sveska (s 2659 stranica i 160 tabli gravura), nije bilo impresivno, ni po opsegu ni po sadržaju. Nije sadržavalo ni biografske članke. Članci su uglavnom bili kompilacije (preuzimanja sadržaja) iz drugih knjiga, a u prvom svesku, nakon predgovora, objavljena je na dvije stranice lista publikacija iz kojih je enciklopedija sastavljena. Najvažnija inovacija bila je kombinacija makropedijskih i mikropedijskih članaka.²⁹⁵ Pojedina područja se razlučuju u sustav odvojenih rasprava, a obrađuju se u abecednom poretku, povezano s pripadnim znanostima. Tako važni predmeti ostaju na okupu i olakšava se povezanost s mnogobrojnim posebnim natuknicama. U susljednim izdanjima i dodatcima proširuje se područje povijesti i biografija znamenitih pojedinaca. Već je drugo izdanje (10 svezaka, 1777–84), koje je, nakon što je Smellie odbio uređivanje jer se protivio obuhvatiti i biografije, uredio James Tytler (polihistor i politički radikal, čiji su reformizam uspješno obuzdali vlasnici Bell i Macfarquhar), bilo znatno poboljšano (sadržavalo je više rasprava i novih članaka te proširenje i aktualizaciju postojećih, a Bell ga je ilustrirao s 340 tabli). Treće izdanje (18 svezaka, 1788–97), koje je počeo uređivati Macfarquhar, a nakon njegove smrti (1793) nastavio George Gleig (svećenik,

²⁹⁴ usp. Enciklopedija, op. cit. (bilj. 286).

²⁹⁵ Ravlić, Slaven, op. cit. (bilj. 293), str. 49.

od 1808. biskup Brechina i primas Škotske episkopalne crkve), definitivno je učvrstilo reputaciju *Enciklopedije* kao pouzdane publikacije koja je proizvod identificiranih stručnjaka i opsežnih studija (npr. članak »kemija« imao je 261 stranicu, a sastavio ga je James Tytler). Izdanju su 1801. dodana dva dopunska sveska (koje je uredio Gleig), a cijelo izdanje s dopunskim svescima objavljeno je 1803. godine. U trećem izdanju surađivalo je nekoliko priznatih znanstvenih autoriteta, kao što je Thomas Thomson, koji je napisao nekoliko opsežnih članaka (kemija, mineralogija i dr.) i uveo upotrebu kemijskih simbola (u dodatnom svesku iz 1801., u članku kemija od 190 stranica, koji je zamijenio Tytlerov članak), te John Robison, tajnik Kraljevskoga društva Edinburgha, koji je napisao više novih članaka iz prirodne filozofije (astronomija, dinamika, elektricitet, magnetizam, mehanika i dr.). Tom izdanju dodana je lista autora, ali najveći broj članaka su napisali (kompilirali) Tytler i Gleig. Izdanje je bilo vrlo kvalitetno ilustrirano: dodane su i table u boji. Treće izdanje bilo je i iznimno profitabilno: prodano je u otprilike 10 000 primjeraka i zaradilo oko 42 000 funta. To izdanje donosi novi fenomen: piratsko izdanje. Osim legitimnoga izdanja tiskanoga u Edinburghu, tiskano je i piratsko izdanje u Dublinu (tiskao ga je u 2000 primjeraka James Moore), te izdanje u Philadelphiji (prva američka enciklopedija), koje se u potpunosti temeljilo na *Britanskoj enciklopediji* (tiskao ga je Thomas Dobson). *Britanska enciklopedija* počinje se shvaćati kao zastava Britanskoga Imperija i kao „nacionalna enciklopedija“ u smislu prijenosnika britanskih vrijednosti u kolonije²⁹⁶, dakle širenja britanskih vrijednosti i tradicije kao univerzalnih. U 19. je stoljeću stekla svjetsku reputaciju te ju je stalno povećavala. Važni suradnici bili su Archibald Constable, James Mill, David Ricardo, Thomas Malthus, Walter Scott. Nova izdanja uglavnom su bili ispravljani i ažurirani pretisci prethodnih – četvrto izdanje u 20 knjiga objavljeno je 1810. godine, a peto izdanje, koje je bilo pretisak četvrtoga s ispravcima objavljeno je 1817. Peto izdanje dopunjeno je posebnim *Dodatkom (Supplement)*, koji je objavljen u šest svezaka 1824 (kasnije nazvano *Supplement to the Fourth, Fifth and Sixth Editions of the Encyclopaedia Britannica*), a uredio ga je Macvey Napier. *Dodatak* je bio glavna inovacija jer je bio više kolektivno djelo eksperata, sastavljeno od autorskih rasprava najznačajnijih britanskih i stranih znanstvenika. Opsežne rasprave objavljene u *Britannici* početkom 19. stoljeća često su se tiskale odvojeno kao udžbenička literatura za studente Edinburškoga sveučilišta. Sedmo izdanje (21 svezak, 1830–42) donijelo je još jednu važnu inovaciju: poseban svezak s indeksom, koji će otada sadržavati svako naredno izdanje. Nakon devetoga izdanja (24 sveska, 1875–89) vlasništvo nad

²⁹⁶ Yeo, Richard, op. cit. (bilj. 287), str. 1.

enciklopedijom prešlo je u SAD (1901), iako je ona i dalje bila britanski proizvod, ne samo po tome što je daleko najveći broj suradnika bio britanski (pa je tako od 1100 suradnika samo njih 70 bilo iz SAD-a, a 60 iz drugih europskih zemalja). Enciklopedijski se rad profesionalizirao i uspostavljeno je trajno uredništvo s urednicima i pomoćnim urednicima kako bi se ubrzao kontinuirani proces revizije i dopune, pa je *Britannica* bila prva enciklopedija koja je krenula u tom smjeru.²⁹⁷ Slavno 11. izdanje objavio je Cambridge University Press (29 svezaka, 1910–11), a značilo je zaokret prema manjim člancima, pa je ukupno sadržavalo oko 40 000 članaka, dvostruko više od 9. izdanja (17 000), iako nije imalo puno veći opseg. Od 1943. vlasništvo nad enciklopedijom prešlo je na Čikaško sveučilište (dobilo ga je kao poklon od generala Roberta Wooda, predsjednika tvrtke Sears, Roebuck and Company, dotadašnjega vlasnika). S 15. izdanjem *Britannice* (28 svezaka, 1974) došao je novi zaokret. To je izdanje, koje je posebno bilo usmjereno isticanju uloge enciklopedije u obrazovanju i razumijevanju a da ne smanji njezinu ulogu kao referentne knjige, obnovilo staru ideju cjeline i jedinstva znanja. Njezina tri dijela – Propedija (*Propaedia, or Outline of Knowledge*), Mikropedija (*Micropaedia, or Ready Reference and Indeks*) i Makropedija (*Macropaedia, or Knowledge in Depth*) – bila su nastojanje da se razumijevanje i učenje neke teme pokaže kao ulazak u krug u kojemu su kratki informativni članci o detaljima stvari kao i dugi članci (neki od njih imaju opseg veće knjige, npr. članak „Kina“ ima 146 stranica velikoga enciklopedijskoga formata) o općim temama uređeni tako da izrazi njihov međusobni odnos i njihov odnos prema cjelini znanja koji osobito ističe propedijski dio kao složeni indeks znanja. To je izdanje kasnije objavljivano s izmjenama i dopunama (1985. uveden je opći indeks). U ožujku 2012. uprava enciklopedije objavila je da neće biti više klasičnih papirnatih izdanja, koja obuhvaćaju manje od 1% prodaje, te da će se usredotočiti na izdanja na novim medijima. *Encyclopaedia Britannica* na početku 1990-ih počela se tiskati na CD-ROM-u, a potpuna verzija enciklopedije na dva diska objavljena je 1995., dok se 1999. pojavila na DVD-u. U isto vrijeme, 1994., *Britannica* se pojavila i kao internetska enciklopedija, pa ova dva oblika (elektronički i internetski) danas posve dominiraju, a korisnici ih često kombiniraju.²⁹⁸ *Britanska enciklopedija* nije nikad bila tako otvoreno i strastveno politički angažirana kako je to bez pridržaja bila prethodna *Francuska enciklopedija*. Potkraj 19. st. do desetog izdanja sve se više obogaćivala posebnijim

²⁹⁷ usp. Collison, Robert L. *Encyclopaedias: Their History Throughout the Ages*. 2. Izdanje. New York: Hafner Publishing Co., 1966. Str. 138–155.

²⁹⁸ usp. Ravlić, Slaven, op. cit. (bilj. 293), str. 49-51.

podacima, što će biti značajka i niza novijih izdanja i dodataka. Najistaknutiji urednici u 19. st. bili su Archibald Constable, James Mill, Ricardo, Malthus, Hazlitt i Sir Walter Scott. Uz tipična i klasična leksikografska djela u modelu enciklopedija, već u 18. st. počinju se javljati slični priručnici u obliku *Conversations-Lexicona*. Niz takvih izdanja počinje djelom *Razgovorni leksikon s osobitim obzirom na suvremeno doba (Conversations-Lexikon mit vorzüglicher Rücksicht auf die gegenwärtigen Zeiten)* u 6 svezaka u Leipzigu (1796–1808). To je djelo započeo G. R. Löbel, ali ga je 1808. kupio F. A. Brockhaus, pod čijim je imenom taj tip enciklopedije postao svjetski poznat, i to kao najuspješniji i najviše oponašan enciklopedijski priručnik. Slijedila su mnoga izdanja, i dotično se djelo od petnaestog izdanja 1928 javlja kao *Der Grosse Brockhaus*, u dvadesetak svezaka. Takvo djelo nije namijenjeno samo uskim znanstvenim krugovima nego i širokim pučkim slojevima, promičući obaviještenost i upućenost u obliku točnih i pristupačnih sažetaka i zemljopisnih pojedinosti Njemačke, a i druge srednjoeuropske (rom. i slav.) enciklopedije nastavljaju s enciklopedijama takva, tipično europskog modernog enciklopedijskog stila. Trinaest izdanja je objavljeno tijekom 19. st. Članci, često veoma kratki, smatrali su se izvrsnim i pouzdanim, posebno za njemačke predmete, davali su reference na najbolje knjige te su uključivali biografije živih osoba. Posljednja (2005-2006) potpuno printana verzija *Brockhaus Enzyklopaedie* je 21 izdanje s približno 24 500 stranica u 30 svezaka. U veljači 2008. F. A. Brockhaus je objavio prebacivanje na online enciklopediju i prekidanje tiskanog izdanja. Distribucija Brockhaus enciklopedije je potpuno prekinuta 2014. godine. U 20. st. izlaze djela u tradiciji velike Francuske ili Britanske enciklopedije, kao *Encyclopaedia Americana*, *Enciclopedia Italiana*, *Enciclopedia Espasa* (u 72 sveska), *Boljšaja Sovetskaja Enciklopedija* (u 65 svezaka).

5.2 Leksikografski zavod Miroslav Krleža

Leksikografski zavod Miroslav Krleža je središnja hrvatska leksikografska ustanova. Utemeljen je kao *Leksikografski zavod FNRJ* 5.10.1950. odlukom jugoslavenske vlade, a na mjesto ravnatelja je postavljen Miroslav Krleža na čiji je poticaj Zavod i osnovan te koji je navedenu dužnost prvog ravnatelja obnašao sve do svoje smrti 1981 (od 1984. ustanova u nazivu nosi njegovo ime). Leksikografski zavod je jedinstvena institucija u državi. Osnivanje Zavoda Krleža je mogao ostvariti u Zagrebu gdje je već postojala jaka hrvatska leksikografska tradicija i neposredno radno iskustvo na *Hrvatskoj enciklopediji* (1941-45) čiji je glavni

urednik M. Ujević, koji je postao Krležinim pomoćnikom i preuzeo vođenje organizacijsko-stručnih poslova. Nakon Krležine smrti, institucija se 20.01.1962. preimenovala u *Jugoslavenski leksikografski zavod*. Sabor SR Hrvatske 15. lipnja 1972. donosi odluku kojom preuzima osnivačka prava i obvezu financijske potpore Zavodu. Vlada Republike Hrvatske 6.02.1991. donosi Uredbu o *Leksikografskom zavodu „Miroslav Krleža“*²⁹⁹. Osnivačka su prava prenesena na Hrvatski sabor, a njegov ustroj i područje djelovanja uređeni su zakonom iz 2003.³⁰⁰, kojim je Zavod institucionaliziran kao javna ustanova u djelatnosti leksikografije i enciklopedike od osobitog interesa za Republiku Hrvatsku.

Zavod je izdao više od 250 različitih enciklopedija, rječnika, leksikona i drugih izdanja. Enciklopedije, atlasi, rječnici, leksikoni i bibliografije Leksikografskog zavoda tradicionalno su pouzdani i na prepoznatljiv način tumače činjenice iz hrvatskog nasljeđa i društvene zbilje.³⁰¹ Sve se veći naglasak stavlja na jednosveščana izdanja. Njima Zavod, kao ustanova koja je dala velik prinos održavanju i podizanju hrvatskoga intelektualnog standarda i čija se produkcija brojem i raznovrsnošću može mjeriti sa sličnom djelatnošću na najvećim svjetskim jezicima, nastoji odgovoriti aktualnim kulturnim potrebama i odnosima.³⁰² Osnovna je djelatnost Zavoda prikupljanje, obrada i provjera podataka, koji su uz dodatna temeljna istraživanja (radi razvoja znanstveno-strukovne osnove) koriste kao baza podataka za izradbu različitih edicija. Dodatne djelatnosti Zavoda odnose se na znanstveno usavršavanje u suradnji sa sveučilišnim ustanovama, razmjenu leksikografski verificiranih podataka sa srodnim inozemnim ustanovama te utvrđivanje općih leksikografsko-enciklopedijskih standarda. Misija Leksikografskog zavoda „Miroslav Krleža“ utemeljenja je osobito u suvremenoj funkciji enciklopedijske leksikografije koja postupcima sistematiziranja, sintetiziranja, indeksiranja, pohranjivanja i distribuiranja informacija i spoznaja daje mogućnost pouzdanog pristupa globalnom znanju, omogućuje sustavnu orijentaciju unutar brzorastućega broja informacija, čime jača mogućnosti široke i kreativne uključenosti korisnika u znanstvene, istraživačke i obrazovne procese. Misija je Zavoda i u razvoju, poznavanju i razumijevanju hrvatske prirodne i povijesne baštine te kulturne, znanstvene, političke i šire društvene

²⁹⁹ Uredba o Leksikografskom zavodu „Miroslav Krleža“. Dostupno na URL: http://narodne-novine.nn.hr/clanci/sluzbeni/1991_02_6_153.html (6.06.2016.).

³⁰⁰ Zakona o Leksikografskom zavodu „Miroslav Krleža“. Dostupno na URL: http://narodne-novine.nn.hr/clanci/sluzbeni/2003_06_96_1223.html (31.05.2016.).

³⁰¹ O Leksikografskom zavodu. Dostupno na URL: <http://www.lzmk.hr/hr/leksikografski-zavod/o-leksikografskom-zavodu> (31.05.2016.).

³⁰² Leksikografski zavod Miroslav Krleža // Hrvatska enciklopedija: mrežna verzija. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=35937> (31.05.2016.).

zajednice u njezinu prožimanju s univerzalnim iskustvom i znanjem. Projekti Leksikografskoga zavoda razvijaju se i kao bitan prinos u nastanku i održanju suvremenoga hrvatskog kulturnog identiteta i znanstvenoga standarda. Misija Zavoda osobito obuhvaća utvrđivanje općih leksikografsko-enciklopedičkih standarda, rad na standardizaciji znanstvenoga i stručnoga nazivlja, razvijanje velikih općih enciklopedija koje povezuju hrvatsku građu i teme s univerzalnima, prezentiranje hrvatske sastavnice i njezinu ugradnju u svjetsku znanstvenu i kulturnu građu na hrvatskom i drugim jezicima, razvijanje projekata o dosad leksikografski neobrađenim temama radi što boljega upoznavanja s različitim područjima znanosti i umjetnosti te rad na digitalnom povezivanju leksikografski obrađene građe u jedinstveni sustav međusobno povezanih jedinica koje se mrežno nadopunjuju u širini i dubini globalnoga znanja.

Vizija Leksikografskoga zavoda Miroslav Krleža od 2015. do 2020. jest:

- unaprijediti mrežne sustave opće i specijalističke enciklopedike radi razvijanja javnoga znanja visoke znanstvene relevantnosti te kreirati jedinstveni javno dostupni digitalni repozitorij leksikografskih sadržaja,
- dopuniti nacionalnu građu analitičkim i sintetskim člancima o najznačajnijim osobama i pojmovima,
- nastaviti suradnju sa znanstvenom zajednicom u izradbi leksikografskih projekata i organizaciji znanstvenih skupova radi potpore razvoju pojedinih znanstvenih disciplina i tema,
- popularizirati rezultate znanstvenih istraživanja objavljivanjem dijelova leksikografski obrađene građe zavodskih projekata na mreži,
- poduprijeti angažman znanstvenika Zavoda u znanstvenom radu, osobito na unapređivanju javnoga znanja i njegove leksikografske obradbe, i nastavnom radu izvan Zavoda,
- unaprijediti međunarodni profil, posebno suradnjom sa srodnim ustanovama u Europi, te inicirati mrežno povezivanje europskih općih enciklopedija, osobito onih »malih« naroda (jezika),
- razviti prepoznatljivost hrvatske nacionalne baštine ažuriranjem i proširivanjem mrežnih leksikografskih sadržaja na stranim jezicima.³⁰³

³⁰³ O Leksikografskom zavodu. Misija i vizija. Dostupno na URL: <http://www.lzmk.hr/hr/leksikografski-zavod/o-leksikografskom-zavodu/misija-i-vizija> (31.05.2016.).

Prethodno navedeno ukazuje na to kako ova doktorska disertacija direktno pridonosi ostvarivanju pojedinih elemenata navedene misije i vizije Leksikografskog zavoda „Miroslav Krleža“, kao što su: pružanje mogućnosti pouzdanog pristupa globalnom znanju, omogućivanje sustavne orijentacije unutar brzorastućega broja informacija, prezentiranje hrvatska sastavnice i njezina ugradnja u svjetsku znanstvenu i kulturnu građu na hrvatskom i drugim jezicima te digitalno povezivanje leksikografski obrađene građe u jedinstveni sustav međusobno povezanih jedinica koje se mrežno nadopunjuju u širini i dubini globalnog znanja, unapređenje mrežnih sustava opće i specijalističke enciklopedije radi razvijanja javnog znanja visoke znanstvene relevantnosti, stvaranje jedinstvenog javno dostupnog digitalnog repozitorija leksikografskih sadržaja te povezivanje europskih općih enciklopedija.

5.3 Hrvatska opća enciklopedija

Samu riječ enciklopedija u modernom značenju je u Hrvatskoj prvi put upotrijebio Pavao Skalić u djelu *Znalac enciklopedija ili kruga znanosti kako svetih tako i svjetovnih* (*Encyclopediae seu orbis disciplinarum tam sacrarum quam profanarum epistemon*, 1559). Prvi nedovršeni pokušaj enciklopedije je *Hrvatska enciklopedija* (1887–90) u dva sveska, objavljena u Osijeku, kojoj su autori bili Ivan Zoch i Josip Mencin. Krajem 1930-ih u Zagrebu se pokreće projekt izrade opće i nacionalne enciklopedije *Hrvatska enciklopedija* u nakladi Hrvatskoga izdavačkog bibliografskog zavoda u Zagrebu. Enciklopedija je bila zamišljena u 12 svezaka, ali je objavljeno svega pet svezaka („A-Automobil“, Automaši–Boito“, „Boja-Cleveland“, „Cliachit-Diktis“, „Dilatacije-Elektrika, 1941–45) ove nedovršene enciklopedije. Glavni urednik i pokretač bio je Mate Ujević. Izdavanju *Hrvatske enciklopedije* (1941-1945) prethodili su prospekti s oglednim arcima enciklopedije koji su obrazlagali vrijednost i potrebu stvaranja jednog takvog djela te iznosili koncepciju razmišljanja i rada uredništva. Analizirano je stajalište kako sastav hrvatskog jezika nije pogodan za analitički način prikazivanja kakav podrazumijevaju leksikoni jer pojmovi koji čine cjelinu ne slijede u abecednom nizu (kao što je slučaj u njemačkom jeziku) te je iz više kratkih natuknica nemoguće dobiti sustavan pregled. Stoga je trebalo stvoriti enciklopediju, a ne leksikon koji u tadašnjim okolnostima ne bi mogao zatvoriti prazninu u sistematiziranju svih grana ljudskoga znanja na hrvatskom jeziku. Kao jedan od važnih motiva uredništva isticana je potreba da se

Hrvati oslobode stranih leksikona i enciklopedijskih djela.³⁰⁴ Stručno uredništvo sastojalo se od 85 eksperata. Cjelokupni sadržaj enciklopedije podijeljen je u 57 glavnih struka i 7 pomoćnih. Otprilike dvije trećine Enciklopedije trebalo je pokrivati opće teme, a jedna trećina specijalne hrvatske teme.³⁰⁵ Usljed političkih prilika po svršetku Drugog svjetskog rata ova enciklopedija je pala u zaborav. Mnogi suradnici i glavni urednik Mate Ujević su bili osnova oko koje je Miroslav Krleža okupio Leksikografski zavod FNRJ, današnji *Leksikografski zavod »Miroslav Krleža«* (LZMK), koji se osniva 1950. u Zagrebu.

Enciklopedija univerzalnoga tipa *Opća enciklopedija* od 1955. do 1982. izišla je u tri izdanja, prvo kao *Enciklopedija Leksikografskog zavoda* (prvo izdanje 1955-64. godine, sedam svezaka; drugo izdanje 1966-69. godine, šest svezaka), a zatim pod imenom *Opća enciklopedija Jugoslavenskog leksikografskog zavoda* (prvo, tj. ukupno treće izdanje 1977-88. godine, osam svezaka i jedan dopunski). U posljednjem izdanju (glavni urednik J. Šentija) od 8 svezaka (uz dopunski deveti) u 50 000 članaka razvrstanih abecednim redom, sintetizira na hrvatskom jeziku sveukupno ljudsko znanje i umijeće. *Enciklopedija Jugoslavije* (1955–71) izišla je u 8 svezaka, a glavni joj je urednik bio M. Krleža. Od 1980. do 1990. izašlo je 6 svezaka na hrvatskom jeziku (glavni urednici: M. Krleža, I. Cecić, J. Sirotković). Istodobno su izašli prvi svesci i na srpskoj ćirilici te na ostalim jezicima naroda i narodnosti SFRJ (slovenskom, makedonskom, albanskom i mađarskom). Zadaća je te enciklopedije bila dati – u obliku tipičnoga modela *encyclopaediae particularis* ili *encyclopaediae nationalis* – pregled civilizacije i kulture, ekonomije i politike svih južnoslavenskih naroda u republikama i autonomnim pokrajinama Jugoslavije. Zbog niza okolnosti koje su nastale raspadom Jugoslavije, ta je enciklopedija ostala nedovršena.

Godine 1994. tiskan je ogledni arak nove *Hrvatske enciklopedije* urednika Dalibora Brozovića. Međutim, nakon konceptijskih promjena, izdanje je prošireno „općom“ sastavnicom te je 1999. godine tiskan prvi svezak *Hrvatske enciklopedije*. Kao nacionalni projekt od najviše kulturne, obrazovne i društvene važnosti *Hrvatska enciklopedija* (glavni urednici D. Brozović, 1999–2001; A. Kovačec, 2002–05; S. Ravlić, 2006–09); izlazi od 1999. do 2009. u 11 svezaka. U tom je izdanju na ukupno 9272 stranice velikog enciklopedijskog formata objavljeno 67 077 članaka (s ukupno 1 059 000 redaka). Nastavljajući petostoljetnu

³⁰⁴ usp. Švab, Mladen. Mate Ujević: utemeljitelj suvremene hrvatske enciklopedistike / Radovi Leksikografskog zavoda „Miroslav Krleža“, 2(1992), str. 14.

³⁰⁵ Mate Ujević (1901.-1967.): utemeljitelj i glavni urednik Hrvatske enciklopedije. Dostupno na URL: www.croatianhistory.net/etf/ujevic.html (26.05.2016).

tradiciju hrvatske leksikografije, *Hrvatska enciklopedija* nastojala je sintetizirati sveukupnu povijest i kulturu hrvatskog naroda, ali unutar globalne okvirnice, u kojoj su prstenasto i spiralno – i bliži i daljnji, i srodni i nesrodni, pojedinci i zajednice, narodi i države. U tom je smislu *Hrvatska enciklopedija* u tradiciji klasične europske enciklopedije (francuskog ili britanskog tipa), ali i sa svim potrebnim značajkama posebne, nacionalne enciklopedije, s obiljem biografskih i povijesnih podataka. Pa ipak, 70% prostora je posvećeno sveopćemu pojmovniku zemlje i svemira, a 30% užemu nacionalnom predmetniku. Cjelokupan enciklopedijski niz obuhvaća 11 svezaka, s 1 000 000 redaka i 10 000 likovnih priloga. Na pojedinim svescima surađivalo je 20 do 30 radnika Zavoda i 300 do 400 vanjskih suradnika. Ona nastavlja veliku enciklopedijsku tradiciju koju je institucionalizirao Leksikografski zavod, kakvu imaju velike europske zemlje i tek neke srednje i manje, a obilježile su je opće enciklopedije: od nedovršene *Hrvatske enciklopedije* (5 svezaka, 1941–45, glavni urednik Mate Ujević), *Enciklopedije Leksikografskog zavoda* (7 svezaka, 1955–64, glavni urednici Marko Kostrenčić, Miroslav Krleža i Miljenko Protega) do *Opće enciklopedije* (8 svezaka, 1977–82, glavni urednik Josip Šentija).³⁰⁶ Od 2013. objavljuje se mrežno izdanje *Hrvatske enciklopedije* koje se temelji se na prethodno navedenom knjižnom (tiskanom) izdanju objavljenom u 11 svezaka od 1999. do 2009. Glavni urednik mrežnog izdanja *Hrvatske enciklopedije* je Slaven Ravlić, a pomoćnici glavnog urednika su Sanja Fabijanić, Zdenko Jecić i Mirjana Mataija. Odlučeno je da se daljnji rad u okviru *Hrvatske enciklopedije*, kao temeljnog programa Leksikografskog zavoda, nastavi u obliku računalne baze iz koje je proizašlo mrežno izdanje enciklopedije te druga elektronička i multimedijaska izdanja. Time je rad na općoj zavodskoj enciklopediji dobio trajan i sustavan karakter. U jesen 2011. započeo je rad na projektu mrežnog izdanja revizijom abecedarija i pisanjem novih članaka, a u 2012. i ažuriranje i dopune tekstova u računalnoj bazi. Pokretanjem mrežnog izdanja *Hrvatske enciklopedije* ulazi se u novu fazu, u kojoj slijede i određene konceptijske inovacije, prije svega u eksplicitnijem autorskom profiliranju dijela makropedijskih i preglednih članaka (potpisivanje članaka, izabrana literatura), ali i u drugim člancima (linkovi na besplatno dostupne značajne knjige i članke, npr. u bazi Hrčak, i sl.), te u ilustracijskom dijelu enciklopedije (videomaterijal i dr.). Mrežno izdanje *Hrvatske enciklopedije* profesionalno je uređivana enciklopedija besplatno dostupna i komunikacijski otvorena svojim čitateljima (korisnicima). Otvorena je za inicijative, prijedloge i doprinose svojih čitatelja (korisnika), koji mogu postati i njezini suradnici. Svaki je doprinos dobrodošao: u obliku komentara,

³⁰⁶ Enciklopedija, op. cit. (bilj. 286).

prijedloga za poboljšanje ili ispravljanja netočnosti, ali i kao posve suradnička dopuna postojećih članaka ili iniciranje novih. Time je enciklopedija otvorena za komunikaciju s korisnicima i za njihove doprinose, a ispunjava svoju temeljnu zadaću – istinita, pouzdana, uravnotežena i razumljiva informacija o svemu relevantnom.³⁰⁷ Osim mrežnog izdanja Hrvatske enciklopedije, svakako je nezaobilazno ne dotaknuti se još jedne mrežne opće enciklopedije tj. *Proleksis enciklopedije*³⁰⁸ Predstavlja prvu veću hrvatsku *opću i nacionalnu online enciklopediju* koja 2010. godine izlazi na hrvatskom jeziku. Svima je dostupna besplatno, kao enciklopedija slobodnog pristupa. Nalazi se na web stranicama Zavoda, kao cjelina, uz mogućnost pretraživanja i vidljivost enciklopedijskih članaka kao pojedinačnih jedinica univerzalne mrežne građe. *Proleksis enciklopedija* je klasična, urednički nadzirana enciklopedija, ali je po mogućnostima novih komunikacija, i moderna elektronička enciklopedija, otvorena za vanjsku suradnju njenih korisnika. *Proleksis enciklopedija* zasniva se ponajprije na tiskanom izdanju *Opće i nacionalne enciklopedije u 20 knjiga*³⁰⁹ koju je, uz leksikografsku potporu Leksikografskog zavoda Miroslav Krleža, izradila tvrtka PRO LEKSIS. Glavni urednik tog izdanja je Antun Vujić. Knjige *Opće i nacionalne enciklopedije* izlazile su kontinuirano od 2005. do 2007. (s dopunskom 21. knjigom, 2009). Izlaskom svih tih enciklopedijskih knjiga ostvareni su enciklopedijsko-leksikografski resursi koji obuhvaćaju oko šezdesetak tisuća članaka, odrednica i natuknica. Nakon objavljivanja *Opće i nacionalne enciklopedije* u tiskovnom obliku, mogućnost i potreba prelaska enciklopedije na elektronički medij značili su logičan nastavak početne ideje. Tako je izrađeno i DVD izdanje³¹⁰ (2009), a zatim i eksperimentalno online izdanje *Proleksis enciklopedije*, u suradnji s CARNetom za korisnike te Hrvatske akademske i istraživačke mreže. Kroz godinu dana rada to se izdanje usavršavalo informatički i sadržajno. Dogovorom između tvrtke PRO LEKSIS i Leksikografskog zavoda „Miroslav Krleža“ koncem 2010. godine odlučeno je da daljnju brigu o uređivanju i izdavanju *Proleksis enciklopedije* preuzme Zavod, te su Zavodu prenesena i sva izdavačka prava, bez naknade. *Proleksis enciklopedija*, koja je i nastala uz raniju leksikografsku potporu Zavoda, tako sada postaje jedno od njegovih online izdanja, svojevrsan uvod u leksikografsko-enciklopedičku mrežnu građu samog Leksikografskoga

³⁰⁷ Predgovor mrežnom izdanju Hrvatske enciklopedije. Dostupno na URL:

<http://www.enciklopedija.hr/predgovor.aspx> (31.05.2016.).

³⁰⁸ *Proleksis enciklopedija*. Dostupno na URL: <http://proleksis.lzmk.hr> (31.05.2016.).

³⁰⁹ *Enciklopedija: opća i nacionalna: u 20 knjiga* / [glavni urednik Antun Vujić]. Zagreb: PRO LEKSIS: Večernji list, 2005-2007.

³¹⁰ *Enciklopedija [Elektronička građa]: opća i nacionalna u 20 knjiga* / [glavni urednik Antun Vujić]. Zagreb: PRO LEKSIS: Večernji list, 2009.

zavoda „Miroslav Krleža“ kao središnje hrvatske leksikografske kuće.³¹¹ Pregled općih enciklopedijskih izdanja LZMK prikazan je Tablicom 6.

Tabela 6: Pregled općih enciklopedijskih izdanja LZMK

NASLOV	BROJ SVEZAKA	RAZDOBLJE IZDAVANJA	UREDNIK	NAZIV NAKLADNIKA
<i>Hrvatska enciklopedija</i>	2 sv.	1887 1890	obrađuju Ivan Zoch i Josip Mencin	Tiskom Dragutina Laubnera
<i>Hrvatska enciklopedija</i>	5 sv.	1941-	glavni urednik Mate Ujević	Naklada konzorcija Hrvatske Enciklopedije
		1941		Naklada Hrvatskog izdavačkog bibliografskog zavoda
		1942		Naklada Hrvatskog izdavačkog bibliografskog zavoda
		1942		Naklada Hrvatskog izdavačkog bibliografskog zavoda
		1945		Naklada Hrvatskog izdavačkog bibliografskog zavoda
<i>Enciklopedija Leksikografskog zavoda</i>	7 sv.	1.izdanje 1955-1964	redaktori Marko Kostrenčić, Miroslav Krleža, Miljenko Protega	Leksikografski zavod FNRJ Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije</i>	8 sv.	1.izdanje 1955-1971	glavni redaktor Miroslav Krleža	Leksikografski zavod FNRJ
<i>Enciklopedija Leksikografskog zavoda</i>	6 sv.	2.izdanje 1966-1969	glavna redakcija Nada Bogdanov, Darko Grlić, Mladen Iveković, Kruno Krstić, Rikard	Jugoslavenski leksikografski zavod

³¹¹ usp. O Proleksis enciklopediji. Dostupno na URL: <http://proleksis.lzmk.hr/naslovnica/> (3.06.2016.).

			Podhorsky, Borko Vranjican	
<i>Opća enciklopedija Jugoslavenskog leksikografskog zavoda</i>	9 sv. (8 sv. + 1 dopunski sv.)	1977 - 1988	glavni urednik Josip Šentija	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije</i>	6 sv.	2.izdanje 1980-1990	glavni urednik Miroslav Kreža, Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije (ćirilica)</i>	2 sv.	1983-1985	glavni urednik Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije (makedonski jezik)</i>	2 sv.	1983-1985	glavni urednik Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije (slovenski jezik)</i>	4 sv.	1983-1989	glavni urednik Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije (albanski jezik)</i>	2 sv.	1984-1987	glavni urednik Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Enciklopedija Jugoslavije (mađarski jezik)</i>	2 sv.	1985-1988	glavni urednik Ivo Cecić, Jakov Sirotković	Jugoslavenski leksikografski zavod
<i>Hrvatska enciklopedija</i>	11 sv.	1999-2009	glavni urednik Dalibor Brozović (1999–2001); A. Kovačec (2002–05); Slaven Ravlić (2006–09)	Leksikografski zavod „Miroslav Kreža“

5.4 Ostala enciklopedijska izdanja LZMK

Leksikografski zavod je od svog osnutka objavio, osim izdanih općih enciklopedija prikazanih u prethodnom odlomku, i cijeli niz izdanja strukovnih enciklopedija za pojedino područje.

Tako su do smrti Miroslava Kreže, odnosno do 1981. godine objavljena sljedeća izdanja. *Pomorska enciklopedija* (1. izdanje u 8 svezaka) od 1954. do 1964., te 2. izdanje u 8 svezaka od 1972. do 1989 (do 1981 je izašlo 5 svezaka ovog 2. izdanja). Glavni urednici su

bili Mate Ujević (1. izdanje) te Vladislav Brajković i Petar Mardešić (2. izdanje). Kao jedna od rijetkih posebnih enciklopedija o svim aspektima mora i pomorstva, ona sustavno, s obiljem najraznovrsnijih podataka, zahvaća sve teorijske i praktične strane i pomorstva i brodogradnje. Takvo je djelo po sustavu i načinu obradbe rijetkost i među najpoznatijim pomorskim narodima. *Medicinsku enciklopediju* (1. izdanje u 10 svezaka) od 1957. do 1965., te 2. izdanje u 8 svezaka (od čega su dva dopunska sveska) od 1967. do 1986. (do 1981 je izašlo 6 svezaka i jedan dopunski svezak). Glavni su joj urednici bili Ante Šercar (1. izdanje) te Mirko Dražen Grmek i Ivo Padovan (2. izdanje). *Muzička enciklopedija* izlazi u 2 sveska od 1958. do 1963. (1. izdanje), te u 3 sveska od 1971. do 1977 (2. izdanje). Glavni urednici su Josip Andreis (1. izdanje) te Krešimir Kovačević (2. izdanje). Među enciklopedijske rijetkosti na širem području svjetske enciklopedikeje objavljena je *Šumarska enciklopedija* u 2 sveska od 1959. do 1963. (1. izdanje), te u 3 sveska od 1980. do 1987. (2. izdanje, od čega je do 1981. izašao 1 svezak). Glavni urednici su Aleksandar Ugrenović i Zvonimir Potočić (1. izdanje) te Zvonimir Potočić (2. izdanje). *Enciklopedija likovnih umjetnosti* izašla je u 4 sveska od 1959. do 1966. Glavni urednici su Andre Mohorovičić, Slavko Batušić, Mirko Šeper. *Tehnička enciklopedija* izlazila je u 13 svezaka od 1963. do 1997., od čega je do 1981. izašlo 7 svezaka. Glavni urednici su Rikard Podhorsky, Živan Viličić, Hrvoje Požar i Duško Štefanović. Zahvaća sva područja tehnike, prirodne znanosti vezane za tehničku praksu pa za matematiku, fiziku i kemiju vezane za naprave, uređaje, strojeve i najrazličitija vozila i prometala, tako da zahvaća i cijeli povijesni i moderni prometni, komunikacijski i informatički sustav. U strukturalnome smislu ona je praktična kombinacija mikropedijske i makropedijske metode. Objavljena je u 13 svezaka, od kojih posljednji sadržava i detaljno kazalo pojmova za svih 13 svezaka. Uz mnogobrojne članke, crteže, dijagrame i fotografije taj priručnik donosi i 10 000 naslova izvorne strukovne literature. *Otorinolaringologija* je izašla u 2 sveska od 1965. do 1966., a glavni urednik je Ante Šercer. Poljoprivredna enciklopedija je izašla u 3 sveska od 1967. do 1973., a glavni urednik je Mladen Josifović. Kao dodatak u trećoj knjizi toga djela nalaze se mnogobrojne biografije i iscrpno predmetno kazalo. U svojem širokom programu enciklopedijske obradbe različitih područja LZMK nije zaobišao ni svijet športa, pa je objavio i iscrpnu enciklopediju svih područja športa i tjelesnog odgoja – *Enciklopediju fizičke kulture* u 2 sveska (1975–76), a glavni urednik Marijan Flander.

Leksikografski zavod „Miroslav Krleža“ je nastavio bogato objavljivati svoja enciklopedijska izdanja i nakon smrti Miroslava Krleže. Prikaz ostalih objavljenih strukovnih

enciklopedija nakon 1981. godine slijedi u nastavku. *Likovna enciklopedija Jugoslavije* objavljena je u 2 sveska (1984–87), a glavni urednik je Žarko Domljan. *Filmska enciklopedija* je objavljena u 2 sveska (1986-90), a glavni urednik je Ante Petravić. *Krležijana* je objavljena u 3 sveska (1993-99), a glavni urednik je Velimir Visković. Iako moderna enciklopedika (npr. talijanska i ruska) poznaje enciklopedijska djela posvećena istaknutim piscima nacionalnih književnosti, hrvatsko izdanje enciklopedije *Krležijana* ipak je nedvojbeno posebnost. Odužujući se svojem utemeljitelju (čije ime i sama nosi) leksikografska ustanova LZMK zahvaća istodobno i cjelokupno djelo najvećega hrvatskog pisca 20. stoljeća, poligrafa i erudita M. Krleže. Jednostavno rečeno, nastojalo se popisati i opisati sve što je Krleža napisao i sve što se ikad i igdje napisalo o Krleži, uz sustavan popis svih tuzemnih i inozemnih izdanja i prijevoda njegova djela te literature o Krleži za razdoblje od 1914. do 1993. Djelo je objavljeno u 3 sveska (1993–99), a posljednji svezak sadržava bibliografiju i literaturu. *Enciklopedija hrvatske umjetnosti* izdana je u 2 sveska (1995–96), a glavni urednik je Žarko Domljan. *Hrvatska likovna umjetnost* izdana je u 8 svezaka 2005. godine, a glavni urednik je Žarko Domljan. Iste godine je objavljena i *Istarska enciklopedija* u jednom svesku, a glavni urednici su joj Miroslav Bertoša i Robert Matijašić. *Hrvatska književna enciklopedija* je objavljena u 4 sveska (2010-12), a glavni urednika je Velimir Visković. Tablica 7 daje prikaz svih strukovnih enciklopedijskih izdanja.

Tabela 7: Prikaz strukovnih enciklopedijskih izdanja LZMK

Naslov	Broj svezaka	Godina izdavanja	Urednik
Pomorska enciklopedija	8 sv.	1.izdanje 1954-1964.	glavni redaktor Mate Ujević
Medicinska enciklopedija	10 sv.	1.izdanje 1957-1965.	glavni redaktor Ante Šerecr.
Muzička enciklopedija	2 sv	1.izdanje 1958-1963.	glavni redaktor Josip Andreis
Šumarska enciklopedija	2 sv.	1.izdanje 1959-1963	glavni redaktori Aleksandar Ugrenović, Zvonimir Potočić.
Enciklopedija likovnih umjetnosti	4 sv.	1959-1966.	glavni redaktori: Andre Mohorovičić, Slavko Batušić, Mirko Šeper.
Tehnička enciklopedija	13 sv.	1963-1997.	glavni redaktor Rikard Podhorsky (31.12.1974), Živan Viličić (v. d. od 1.1.1975), Hrvoje Požar (1979-1992), Duško Štefanović (1997)
Otorinolaringologija	2 sv.	1965-1996	glavni redaktor Ante Šerecr

Poljoprivredna enciklopedija	3 sv.	1967-1973	glavni redaktor Mladen Josifović
Medicinska enciklopedija	8 sv. (6 sv., dopunski svezak, drugi dopunski svezak)	2. prerađeno izdanje 1967-1986	glavni redaktor Ante Šerčer, glavni urednik dopunskog sveska Mirko Dražen Grmek, glavni urednik drugog dopunskog sveska Ivo Padovan
Muzička enciklopedija	3 sv.	2. izdanje 1971-1977	glavni urednik Krešimir Kovačević
Pomorska enciklopedija	8 sv.	2. izdanje 1972-1989	glavni urednici Vladislav Brajković, Petar Mardešić
Enciklopedija fizičke kulture	2 sv.	1975-1977	glavni urednik Marijan Marjan Flander
Šumarska enciklopedija	3 sv.	2. izdanje 1980-1987	glavni urednik Zvonimir Potočić
Likovna enciklopedija Jugoslavije	2 sv.	1984-1987	glavni urednik Žarko Domljan
Filmska enciklopedija	2 sv.	1986-1990	glavni urednik Ante Peterlić
Krležijana	3 sv.: Sv. 1: A-LJ, Sv. 2: M-Ž, Sv. 3.: Bibliografija Miroslava Krleže	1993-1999	glavni urednik Velimir Visković
Enciklopedija hrvatske umjetnosti	2 sv.	1995-1996	glavni urednik Žarko Domljan
Hrvatska likovna enciklopedija	8 sv.	Izdanje Leksikografskog zavoda i Vjesnik d.o.o. // izdanje priredile Zdenka Ožić, Ozana Martinčić i Bojana Franić] 2005	glavni urednik Žarko Domljan
Istarska enciklopedija	1 sv.	2005	urednici Miroslav Bertoša, Robert Matijašić
Hrvatska književna enciklopedija	4 sv.	2010-2012	glavni urednik Velimir Visković

Kao što je izdana mrežna opća Hrvatska izdanja, tako su uslijedila i mrežna izdanja strukovnih enciklopedija. Tako je Zavod neke od svojih strukovnih izdanja objavio na Portalu znanja.³¹² Osim strukovnih enciklopedija, *Istarske enciklopedije* i *Krležijane*, portal obuhvaća još pet izdanja poput *Hrvatskog obiteljskog leksikona*, *Filmskoga leksikona*, *Nogometnoga leksikona*, *Medicinskog leksikona* te izbor članaka iz *Hrvatskoga biografskog leksikona*. Sadržaj portala trenutačno čini 60 652 članaka. To je sadržaj koji Zavod u cjelini nudi svim korisnicima interneta u Hrvatskoj (a i svijetu), a koji će se stalno nadopunjavati – novim člancima, preuzimanim iz već objavljenih izdanja. U tom smislu portal je početak i nešto drukčijeg djelovanja Zavoda koji znanje što ga prikuplja i leksikografski obrađuje neće prezentirati samo na papiru već i sve više, a naposljetku i primarno, u elektroničkom obliku. Izdanja su međusobno povezana zajedničkim abecedarijem, načinjenim na temelju *Hrvatskog obiteljskog leksikona*, što čini okosnicu portala na koju se povezuju istovjetne natuknice iz ostalih izdanja. *Medicinski leksikon* kao sedmo izdanje portala nije povezan u zajednički abecedarij. Natuknice koje su specifične za pojedina izdanja u osnovni su abecedarij portala dodane kao natuknice bez teksta. Za osnovno je pretraživanje portala dostupan cjelokupan sadržaj *Hrvatskog obiteljskog leksikona* te natuknice iz povezanih izdanja. Na taj su način povezane *Istarska enciklopedija*, *Filmski leksikon*, *Nogometni leksikon*, *Krležijana* te članci *Hrvatskoga biografskog leksikona*. Sva izdanja dostupna su za pretraživanje i kao zasebne cjeline.³¹³ Navedeno ukazuje kako je portal pokušaj umrežavanja strukovnih enciklopedija s ostalim izdanjima LZMK. U tom postupku se uočava kako se tradicionalna enciklopedička organizacijska struktura s papirnatom medija pokušava prenijeti i prilagoditi organizaciji enciklopedičkih mrežnih izdanja, što ukazuje na potrebu preispitivanja i istraživanja organizacije enciklopedičkog znanja u mrežnom okruženju što je jedna od temeljnih zadaća ove doktorske disertacije.

³¹² Portal znanja. Dostupno na URL: http://enciklopedija.lzmk.hr/o_portalu.aspx (03.6.2016.).

³¹³ usp. O portalu. Dostupno na URL: http://enciklopedija.lzmk.hr/o_portalu.aspx (03.6.2016.).

5.5 Zaključak

Peto poglavlje *Enciklopedička djela* dalo je uvod u povijesni razvoj enciklopedije u svijetu kako bi se u potpunosti shvatio kontekst kroz koji je enciklopedija morala proći da bi zadobila današnje poznate odlike suvremenog enciklopedičkog djela. Poglavlje je dalo osnovne informacije o razvoju hrvatske središnje leksikografske ustanove, Leksikografskom zavodu „Miroslav Krleža“ (LZMK), koji obavlja djelatnost leksikografije i enciklopedike od osobitog interesa za Republiku Hrvatsku. Ukazalo se u ostvarivanju kojih pojedinih elemenata misije i vizije LZMK može pridonijeti ova doktorska radnja. Prikazana je izdavačka djelatnost općih i strukovnih enciklopedijskih izdanja LZMK, obuhvaćajući izdanja na papirnatom mediju i ona u mrežnom okruženju. Analizom izdavačke djelatnosti mrežnih izdanja, utvrđeno je kako se u njihovom ostvarivanju uglavnom radi o prilagodbi tradicionalne organizacije enciklopedičkog znanja s papirnatog medija u mrežno okruženje. Upravo će ova doktorska radnja istražiti koje bi koristi od primjene ontologijskih principa semantičkog weba imao LZMK, kao i sami korisnici ovih vrijednih izvora znanja.

6 ENCIKLOPEDIČKA ORGANIZACIJA ZNANJA

6.1 Uvod

Riječ enciklopedija dolazi od grčkih riječi *ἐγκύκλιος* - koji ide redom, opći i *παιδεία* - odgoj, obrazovanje, dakle sustavna i opća naobrazba. Enciklopedija se definira kao “djelo u kojem se, abecednim ili kakvim drugim metodičkim slijedom, okupljaju i sustavno obrađuju činjenice i spoznaje o svim ljudskim znanjima (opća enciklopedija) ili pak sva građa jedne znanosti, umjetnosti, područja (strukovna ili posebna enciklopedija)”³¹⁴

Enciklopedička organizacija znanja organizirano prezentira znanja kako bi se korisniku omogućilo brzo i lako dolaženje do željene informacije. Takvo organizirano prezentiranje znanja tradicionalnih enciklopedičkih djela rezultiralo je njihovim ujednačenim osnovnim svojstvima, načinom izlaganja i organizacijom sadržaja. Za razliku od većine literarnih djela koja imaju linearnu organizaciju teksta tj. čitaju se „od početka prema kraju“, enciklopedička djela omogućavaju pretraživanje željene informacije svojom nelinearnom organizacijom kojom je znanje podijeljeno na manje, pregledne dijelove. Korisnik pristupa znanju poštujući metodički slijed kojim je ono nanizano. Nadalje, slijedeći uputnice kojima su dijelovi znanja povezani korisnik sam određuje slijed sadržaja prema svom interesu.

Organizacijska struktura tradicionalne enciklopedije zasniva se na slijedećem procesu³¹⁵:

- razdioba nepregledne količine znanja u pregledne dijelove tj. članke, a ukoliko su oni još uvijek odveć veliki (nepregledni), dodatno se dijele na naslovljena poglavlja,
- dodjeljivanje naziva člancima – natuknica ili lema, kako bi se traženi dijelovi znanja (članci) mogli pronaći,
- nizanje članaka nekim logičkim slijedom: abecedni redoslijed, konceptualni redoslijed (sustavni, tematski red ili poredak), kronološki redoslijed (vremenski red ili poredak), prostorni redoslijed (red ili poredak) , mješoviti redoslijed,
- uspostavljanje veza među člancima (umrežavanje članaka),
- izrada indeksa.

³¹⁴ Enciklopedija, op. cit. (bilj. 286)

³¹⁵ Michel, P.; Herren, M.. Unvorgreifliche Gedanken zu einer Theorie des Enzyklopadischen - Enzyklopadien als Indikatoren für Veränderungen bei der Organisation und der gesellschaftlichen Bedeutung von Wissen // Allgemeinwissen und Gesellschaft, Akten des internationalen Kongresses über Wissenstransfer und enzyklopadische Ordnungssysteme, Prangins, 2003. Dostupno na URL: <http://www.enzyklopaedie.ch/kongress/aufsaeetze/vorwort.pdf> (5.01.2015.).

Usporedbom prethodno navedene organizacijske strukture tradicionalnih s elektroničkim enciklopedijama u web okruženju uočavaju se određene razlike proizašle iz različitih vrsta medija: npr. uspostavljanje poveznica na naslovljena poglavlja kako bi se omogućilo navigaciju po određenom članku, osiguranje sinonima za odabranu natuknicu kako bi se omogućilo bolje pretraživanje, pristupanje člancima pomoću tražilica pa redosljed članaka naizgled ne igra bitnu ulogu jer nisu doslovno fizički nanizani (mogu se istovremeno ostvariti sve prethodno navedene vrste logičkog nizanja članaka tj. web okruženje može omogućiti istovremeno predstavljanje informacije abecednim, konceptualnim, kronološkim, prostornim ili mješovitim redosljedom, kao i istovremenu mogućnost prikazivanja informacije s prostorno-vremenskom komponentom), elektroničke poveznice (hiperveze, linkovi) prilikom umrežavanja članaka i dodatne mogućnosti umrežavanja članaka (rubrika »vidi još«, liste članaka svrstanih u pojedinu kategoriju), pretraživanje informacija odvija se uglavnom preko automatski izdvojenih kazala itd. Prethodna kratka usporedba ukazuje da su za organizaciju znanja u web okruženju potrebni *sustavi za organizaciju znanja* (engl. knowledge organization systems - KOS) prikladni za web okruženje. KOS ima jedan cilj: organizirati sadržaj kako bi se podržalo pretraživanje relevantnih jedinica građe iz digitalne zbirke. Prema Hodge³¹⁶ termin *sustavi za organizaciju znanja (KOS)* uvela je *Networked Knowledge Organization Systems Working Group* prilikom svog prvog sastanak na *ACM Digital Libraries '98 Conference* in Pittsburgh, Pennsylvania. Termin *sustavi za organizaciju znanja* ima namjenu obuhvati sve vrste shema za organizaciju informacija i promicanje upravljanja znanjem. Sustavi za organizacija znanja uključuju klasifikacijske i kategorizacijske sheme koje organiziraju materijale na općoj razini (kao što su knjige na policama), predmetne odrednice koje pružaju detaljniji pristup, i normativne datoteke koje kontroliraju varijante verzije ključnih informacija (kao što su geografska i osobna imena). Sustavi za organizaciju znanja također uključuju visoko strukturirane rječnike, kao što su tezaursi, i manje tradicionalne sheme, kao što su semantičke mreže i ontologije. Hodge prikazuje vrste KOS-a grupirane u tri opće kategorije, priznajući kako popis nije sveobuhvatan. Kategorije u koje grupira vrste KOS-a su liste termina koje ističu popis termina često s definicijama, klasifikacije i kategorije koje naglašavaju stvaranje skupova predmeta, liste odnosa koje naglašavaju veze između termina i pojmova. Svi ovi primjeri sustava za organizaciju znanja variraju u složenosti, strukturi i funkciji. Ontologiju je pridružio kategoriji

³¹⁶ Hodge, Gail. Systems of Knowledge Organization for Digital libraries: beyond traditional authority files. Washington, DC: *the Council on Library and Information Resources*. Dostupno na : <http://www.clir.org/pubs/reports/pub91/contents.html> (14.05.2015.).

liste odnosa.³¹⁷ Stoga je za elektroničke izvore informacija (enciklopedije, zbirke) potrebno razviti ontologiju, koja će omogućiti računalnu obradu i upravljanje informacijskim izvorima na temelju semantike kao njihovog sastavnog dijela.

Namjera ovog rada je istražiti problematiku razvoja odgovarajuće ontologije za mrežno izdanje Hrvatske enciklopedije (HE) Leksikografskog zavoda Miroslav Krleža na temelju strukturalne organizacije njezinih članaka tj. dosadašnje tradicionalne enciklopedičke organizacije znanja, kako bi se dale smjernice za poboljšanje i daljnji razvoj organizacije enciklopedičkog znanja općenito, a napose onoga u web okruženju imajući u vidu njegove specifičnosti i daljnji razvoj prema semantičkom webu. Kako bi se istražila prethodno navedena problematika i ostvario postavljeni cilj razvoja ontologije ovo poglavlje će ukratko prikazati strukturalnu organizaciju enciklopedijskog članka, osvrnut će se na osnovne vrste enciklopedijskih članaka, odlike biografskog enciklopedijskog članka i donijeti će analizu vrste podataka koje sadrže enciklopedijski biografski članci.

6.2 Enciklopedijski članak

Osnovna sadržajna jedinica enciklopedičkoga djela je enciklopedički članak. Svaki članak obrađuje jedan pojam ili temu te predstavlja cjelovit i samostalan dio koji se može usvajati neovisno od drugih članaka ili cjeline djela. Sveobuhvatno obuhvaća obrađivani pojam ili temu odgovarajući na pitanja *tko, što, kada, gdje, kako i zašto*. Može obuhvaćati tekstualni dio, ilustracije, multimedijske prezentacije sadržajno vezane uz temu članka; a duljina enciklopedijskog članka ovisi o tome obrađuje li se u njemu samo jedan pojam (*mikropedijski* kraći članak) ili jedna tema s objašnjenima više pojmova (*makropedijski* dulji članak) zahvaćajući neku temu veoma opsežno i uvodeći naslovljena poglavlja zbog lakšeg snalaženja u samom članku. Samo oblikovanje enciklopedičkog članka predstavlja središnji dio enciklopedičkog rada. Struktura članka treba biti jednako provedena kod svih članaka nekog enciklopedičkog djela, što znači da su članci koji obrađuje pojedinu struku ili tematsko područje organizirani na isti način tj. svaka vrsta enciklopedijskog članka ima ujednačenu strukturu s obzirom na nizanje pojedinih elemenata članka (npr. biografski članci, članci kemijskih elemenata i sl.).

³¹⁷ Ibid.

Strukturu enciklopedijskog članka čini glava i tijelo članka. *Glava članka* obuhvaća natuknicu (naslov članka) i dodatne obavijesti o natuknici, uglavnom lingvističke naravi kao što su istoznačnice, istovrijednice, prevukovljivanje (transkripcija), preslovljavanje (transliteracija), podrijetlo riječi (etimologija) i oznake struke (terminološke odrednice). *Natuknica* ili *naslov članka* je riječ ili skupina riječi (sintagma) koja opisuje o čemu se u članku govori te omogućava korisniku pronalaženje članka koji najbolje odgovara njegovom interesu. Njezina funkcija odgovara predmetnoj odrednici ili deskriptoru, tj. normiranoj riječi ili skupini riječi koja se u informacijskim sustavima koristi za indeksiranje i pretraživanje dokumenata. Stoga je za natuknicu potrebno izabrati riječ ili sintagmu na koju će prosječan korisnik najprije pomisliti kada razmišlja o traženom pojmu tj. izabrati riječi ili sintagme kojima se u praksi najčešće označava pojedini pojam. Potrebno je provesti usklađivanje nazivlja te se odabranog normiranog naziva treba pridržavati u tekstovima, a prilikom izbora natuknica se treba rukovoditi i odabirom najprikladnijih naziva³¹⁸ tj. domaće riječi imaju prednost pred stranim te je dobro izbjegavati tuđica i druge riječi koje nisu uvijek najfrekventnije u uporabi. Natuknica može biti jednočlano vlastito ime (npr. ime osobe, države ili pokrajine, grada itd.), može biti sastavljena i od prezimena i imena (Shakespeare, William), može biti tročlano rimsko ime (*Ciceron, Marko Tulije*) ili pak višečlano zemljopisno ime (*Sveti Petar u Šumi; Usti nad Labem; Velika Britanija*) itd. Natuknica jednako tako može biti samo jedna riječ (*astma, cvijet, sisavci*) ili skup riječi (*aramejski jezik, socijalno osiguranje*), ali i cijela izreka (*homo homini lupus, mens sana in corpore sano*;

³¹⁸ Mihaljević donosi načela za vrednovanje istoznačnih naziva:

1. domaće riječi imaju prednost pred stranim (npr. nazive kontrola, audit, evaluacija, verifikacija, validacija, inspekcija itd., bolje je zamijeniti nazivima paćenje, nezavisna ocjena, vrednovanje, potvrda, ovjera, pregled itd.),
2. nazivi latinskog i grčkog podrijetla imaju prednost pred nazivima preuzetim iz engleskoga, francuskoga, njemačkoga itd., npr. licencija je bolje nego licenca (hrvatski je naziv ovlasnica),
3. prošireniji naziv ima prednost pred manje proširenim (npr. naziv računalo prošireniji je od naziva rednik),
4. naziv koji je korisnicima prihvatljiviji ima prednost pred nazivom koji je manje prihvatljiv,
5. naziv mora biti usklađen s (fonološkim, morfološkim, tvorbenim, sintaktičkim) sustavom hrvatskoga književnog jezika (npr. tenis-klub treba zamijeniti nazivom teniski klub, brojčano-slovčane oznake s brojčano-slovne oznake),
6. kraći nazivi imaju prednost pred duljim (npr. jedinica za obradu je obradnik),
7. naziv od kojega se lakše tvore tvorenice ima prednost pred onim od kojega se ne mogu tvoriti tvorenice,
8. naziv ne smije unutar istoga terminološkog sustava imati više značenja,
9. značenje naziva ne smije se bez valjana razloga mijenjati – ako jedan naziv već ima određeno značenje, ne treba i tom nazivu davati nova značenja,
10. naziv ima prednost ako odgovara pojmu kojem je pridružen i odražava svoje mjesto u pojmovnom sustavu, a o tome može odlučiti predmetni stručnjak // Mihaljević, Milica. Terminološki priručnik. Zagreb: Hrvatska sveučilišna naklada, 1998. Str. 36-37.

panta rhei). Napokon, kao natuknica može dolaziti i jedno slovo ili jednoslovčani simbol, odnosno kratica (*H, O, m*), višeslovčani simbol (*Au, kg, pH*) ili višeslovčana kratica (*a. a., p. m.*) ili pak sigla (*HV, INRI, SAD, UK, ZND* i dr.)... «³¹⁹. Osobna imena kao natuknice mogu biti jednosložna (*Abel, Atila*), ili su najčešće višesložna i sadrže ime i prezime u inverznom redosljedju (*Bošković, Ruđer; Petrović, Dražen*), ili se sastoje od osnovnoga imena i kakva atributa (pridjeva, rednog broja i sl.), pa se donose u prirodnom redosljedju, npr. *Karlo Veliki, Leonardo da Vinci, Grgur Ninski, Petar Krešimir IV*. Strana osobna imena donose se u latiničnom pismu. Imena iz onih jezika koji se izvorno pišu latiničnim pismom u osnovi se donose u izvornoj grafiji, tj. kako se pišu na izvornom jeziku, uz što se kadšto bilježi i njihov izgovor, tj. transkripcija. Imena iz jezika koji se pišu drugim pismima (ćirilicom, modernim grčkim pismom, arapskim hebrejskim, japanskim pismom itd.) donose se u kroatiziranom liku (najčešće ruska, ukrajinska, bugarska, srpska imena), ili u uobičajenom latiničnom zapisu u dotičnoj zemlji (arapska, kineska, japanska imena), kadkad uz navođenje njihova izvornog pisanja transliteracija). Natuknica je u grafički oblikovanom tekstu istaknuta masnom i većom veličinom pisma od ostatka teksta te uvlakom. Natuknica bi trebala početi malim slovom, a ako je riječ o vlastitom imenu onda velikim slovom. Enciklopedije se bave isključivo pojmovima (ne riječima), stoga je natuknica najčešće imenica (ili nosiva riječ u višočlanoj natuknici). Navodi se u kanonskom obliku³²⁰ u nominativu jednine. Katkada se, u iznimnim slučajevima, natuknica donosi u množini. Višočlane natuknice mogu biti dvočlane, tročlane, četveročlane i dr., a najčešće se javljaju kao kombinacija pridjeva i imenice, te imenice u nominativu i imenice u genitivu. Međutim, ponekad postoji prihvatljiv jednočlani ili kraći višočlani naziv kojim bi se oni mogli zamijeniti, pa se tom rješenju svakako treba prikloniti. Ponekad se provodi inverzija njihova uobičajena poretka u sintagmi. To se kod većine enciklopedija primjenjuje na osobna imena i prezimena (*Tesla, Nikola*, a ne *Nikola Tesla*), što dakako pridonosi lakšem pronalaženju tražene osobe. Kod nekih enciklopedičkih djela, napose strukovnih, sustavno se provodi inverzija natuknica sastavljenih od pridjeva i imenice, (*jezik, programski; jezik, prirodni*), što se opravdava namjerom da se čak i u abecednom redosljedju članaka, članci slične tematike, koji opisuju različite izvedbene ili funkcionalne inačice nekog pojma, nađu jedan blizu drugoga, pa bi se zahvaljujući takvom principu sve

³¹⁹ Organizacija natuknice. // Hrvatska enciklopedija: 1 : A – Bd / < glavni urednik Dalibor Brozović >. Zagreb : Leksikografski zavod Miroslav Krleža, 1999- . Str. XIV.

³²⁰ kanonski oblik, *lingv.* osnovni oblik u kojemu su natuknice navedene u rječniku: imenice u nominativu, glagoli u infinitivu itd.; rječnički oblik. // Hrvatski jezični portal. Dostupno na URL: http://hjp.znanje.hr/index.php?show=search_by_id&id=e151Xhk%3D (14.07.2015.).

vrste strojeva našle na jednom mjestu (*stroj, električni; stroj, parni; stroj, šivaći...*), ali je korisnost takvog principa upitna, jer korisnik u mnogim slučajevima, kada traženi pojam ne nađe pod višečlanim nazivom prirodnoga slijeda, odustaje od daljnje potrage. Ponekad se u nekim enciklopedičkim djelima takva inverzija iznimno provodi radi veće vjerojatnosti da će se traženi pojam naći pod invertiranim nazivom, npr. *crv, računalni*. Akronime tj. složene kratice, koji nastaju od početnih glasova ili slogova višečlanih naziva, trebalo bi izbjegavati kao natuknice jer bi se ti pojmovi u pravilu trebali obraditi pod njihovim punim nazivima. Međutim, postoje mnogi akronimi za koje prosječan čitatelj ne zna kako glasi njihov puni naziv, kod tih se slučajeva korisnika treba uputiti na članak tj. natuknicu pod kojom će pojam biti obrađen pomoću *slijepih natuknica*. Za akronime koji predstavljaju uvriježen naziv, dok se njihov puni naziv rijetko koristi, može se razmisliti o njihovom donošenju u vidu natuknica članaka pod kojim će taj pojam biti obrađen. Postoje tuđice i posuđenice za koje se u potpunosti udomaćila odgovarajuća hrvatska riječ (npr. kompjutor – računalo, tastatura – tipkovnica), pa se one mogu samo navesti kao drugi nazivi nekog pojma. Za tuđice i posuđenice za koje postoji zamjenska hrvatska riječ, koja se međutim manje koristi, hrvatska riječ se prihvaća kao natuknica članka u kojem se taj pojam opisuje, ali se korisnika slijepom natuknicom upućuje na njega. Tuđice i posuđenice za koje se nije uvriježila njihova domaća zamjena navode se takve. Za imena geografskih pojmova, međunarodnih tvrtki i institucija vrijede ista pravila kao i za osobna imena. Imena naselja, zemalja, pokrajina, rijeka, jezera, planina, otoka i dr. najčešće su udomaćena u hrvatskom jeziku, pa im se izgovor i ne treba bilježiti. Kada za isti pojam postoje dva ili više ravnopravnih naziva (istoznačnice kao dio natuknice), ponekad se u pojedinim slučajevima donose oba naziva u sklopu natuknice variranjem oblika natuknice tj. odvajanjem naziva veznikom *ili* uz istodobno variranje veličine pisma (biografski leksikon ili rječnik), ili se istoznačnica stavlja u zagradu, npr. apsolutna (nadmorska) visina. Takvu praksu, međutim, valja u potpunosti izbjeći.

Tijelo članka čini definicija i detaljna objašnjenja kako bi se objasnilo značenje i opisao pojam kojega članak obrađuje. Dakle, započinje definicijom pojma, iza kojega slijedi tekst s opisom i uputnicama (poveznicama) na druge članke. U tijelo članka se također mogu ubrojiti ilustracije ili multimedijски sadržaji vezani uz članak. Na kraju katkada dolazi i potpis autora, preporučljiva literatura, poveznice na internetske izvore i sl. Definicija nedvosmisleno i jednoznačno opisuje značenje nekog pojma. Enciklopedička definicija je znanstvena (logička) jer enciklopedije obrađuju pojmove koji su stručni ili znanstveni. Logička definicija

identificira objekt te ga nedvosmisleno okarakterizira kao člana klase.^{321,322} Sastoji se od višeg rodnog pojma (nadređeni pojam, genus proximum) i specifične razlike (ograničavajuće svojstvo, differentia specifica). Pojam kojemu se sadržaj definicijom određuje naziva se definendum, a pojmove s pomoću kojih se određuje definiens. Pojmove koji se ne mogu definirati jer nemaju nadređenih pojmova naziva se kategorija. Tako se neki pojam definira tako da se najprije odredi najbliža (najuža) skupina pojmova kojoj on pripada (viši rodni pojam, odnosno hiperonim), a potom se odredi specifična razlika po kojoj se razlikuje od drugih pojmova iz skupine. Tako koncipirane definicije nazivaju se *sadržajnim (intenzijskim) definicijama*, koje se razlikuju od *opsegovnih (ekstenzijskih) definicija* tj. vrste definicije koja se rabi u slučajevima kada nije moguće odrediti specifičnu razliku drukčije nego da se navedu svi pojmovi koji spadaju pod neki viši rodni pojam, a kadšto nije moguće niti odrediti viši rodni pojam, pa se neki pojam definira nabrojanjem svih slučajeva koje on obuhvaća. Postoje još i *odnosne definicije (kontekstne, relacijske)* koje definiraju pojam pokazujući u kakvom je on odnosu s drugim poznatim pojmovima, što je često rabi kao rječnička definicija, ali je za enciklopedičku definiciju uglavnom neprikladno. Definicija se nastavlja na glavu članka, a od nje se uobičajeno odvaja zarezom, ali i divizom (–), dvotočkom, prijelazom u novi redak, ili se posebno ne odvaja. Uglavnom počinje malim slovom, ali može početi i kao nova rečenica velikim slovom (napose kada od glave članka nije odvojena interpunkcijskim znakom). Iz definicija se u pravilu izostavlja pomoćni glagol. Osim toga, u definicijama nije potrebno navoditi naziv za..., ili vrsta..., jer se podrazumijeva da je natuknica naziv za pojam koji se u definiciji obrađuje, dok je velika većina pojmova koje se definira vrsta nečega, što je obuhvaćeno višim rodnim pojmom. Kod većih članaka poželjno je nakon definicije dati kratak uvod koji će korisniku dati kratko objašnjenje pojma izlazeći u susret onim korisnicima koji traže samo kratak opis pojma, ne ograničavajući ulogu djela kao izvora podataka onima koji žele detaljnije istražiti temu jer će opsežno i detaljno objašnjenje biti dano u daljnjem tekstu članka. Uvod Hrvatske enciklopedije navodi kako fizički opseg jedinice nije mehanički izraz njezine važnosti jer broj redaka posvećenih jednoj jedinici nije ni nužno, ni izravno u svezi s njezinom važnošću. Međutim, jedinice (pojmovi) koje imaju osobito kulturno-povijesno, društveno, gospodarsko i sl. značenje (nazivi umjetničkih smjerova, formacija i škola, polit. stranaka, pokreta i razdoblja, društva. Ili umjetničkih skupina, te kult. i polit. događaja i sl.) popraćene su, uz osnovnu definiciju, širim tumačenjem enciklopedijskog

³²¹ Zgusta, L. Priručnik leksikografije. Sarajevo: Svjetlost, 1991. Str. 238-240.

³²² Bogutovac, Antonijela. Razvoj definicija pojmova astronomija, fizika i matematika u hrvatskim jednojezičnim rječnicima. // Studia lexicographica 2, 1(2), 2008., str.52.

značaja, osim ako se izrijekom ne upućuje na neku drugu natuknicu.³²³ Nakon definicije, odnosno kratkog uvoda kod većih članaka, slijedi tekst u kojemu se detaljno objašnjava pojam ili tema obuhvaćena natuknicom sagledavajući ga iz svih mogućih aspekata. Tekst članka zauzima najveći dio članka te varira od nekoliko redaka do nekoliko stranica. Počinje kao nova rečenica koja se nastavlja na definiciju, ali od koje može biti odijeljena točka zarezom. Struktura tekstovnog objašnjenja natuknice podrazumijeva ujednačen slijed kojim se iznose činjenice o datom pojmu ili temi jednog stručnog područja ili istovjetne obavijesne razine, kao i način isticanja pojedinih dijelova objašnjenja. Tako će se primjerice istim redom nizati pojedine činjenice ili aspekti obrade nekog tehničkog pojma, nekog geografskog pojma, botaničke ili zoološke vrste, kemijskog elementa ili spoja, biografije neke osobe. Pojedini dijelovi objašnjenja najčešće se posebno ne ističu, tj. tekst slijedi u istom odlomku. Kod opsežnijih enciklopedijskih članaka pojedine smislene cjeline objašnjenja odvajaju se u zasebne odlomke, označene prelaskom u novi red i uvlakom, što često predstavlja više iznimku no pravilo. Kod najopsežnijih članaka ostavlja se mogućnost razdvajanja tekstovnog objašnjenja u poglavlja, koja se i naslovljavaju, a u nekim se slučajevima strukovno orijentiranih makropedija i obrojčavaju. Kod tako strukturiranih objašnjenja, otvara se mogućnost i uvođenja njihova sadržaja, koji slijedi iza uvodnog dijela. Isticanjem pojedinih riječi u tekstu drukčijim pismom od osnovnog korisniku se olakšava snalaženje u njemu jer su pojedine riječi uočljivije već pri letimičnom pregledavanju teksta što doprinosi olakšanja čitanja i pronalaženja informacija. Prilikom pisanja tekstovnog objašnjenja pojedinog pojma upotrebljavaju se oni pojmovi za koje se može pretpostaviti da ih korisnik poznaje, a nepoznati pojmovi trebaju biti obrađeni kao zasebni članci ili ih u dotičnom članku treba istaknuti drukčijim stilom pisma (masnim, kurzivnim), te ih u zagradi ili umetnutoj rečenici objasniti na onom mjestu na kojem se prvi puta spominju. Ukoliko su ti pojmovi objašnjeni u zasebnim člancima, pa u predmetnom članku niti nema njihova objašnjenja, na te članke nije potrebno dodatno upozoravati uputnicom, jer se njihovo postojanje podrazumijeva.

Dakle, uputnice mogu biti sastavni dio teksta u kojemu se objašnjava značenje natuknice u tradicionalnim enciklopedičkim djelima. U novije vrijeme pojedine uputnice se sve češće grupiraju u zasebne dijelove članaka sa značenjem »vidi još«. Uputnice se u enciklopedičkim djelima javljaju u tri osnovna oblika: uputnice kao samostalni članci, uputnice u tekstu i uputnice na kraju teksta.

³²³ Tumačenja značenja i sadržaja definicije riječi. // Hrvatska enciklopedija: 1 : A – Bd / < glavni urednik Dalibor Brozović>. Zagreb : Leksikografski zavod Miroslav Krleža, 1999- . Str. XVI.

Uputnice kao samostalni članci se javljaju u enciklopedičkim djelima kao članci u kojima se za objašnjenje natuknice, odnosno naziva iz naslova članka, upućuje na neki drugi članak te se takvi članci nazivaju *slijepim natuknicama* jer se sastoje samo od natuknice i uputnice. Javljaju se kada je neki pojam obrađen pod prikladnijim nazivom ili u sklopu većeg, preglednog članka. Slijepa natuknica ima ulogu indeksne jedinice kod kazala (jedinica kazala), pri čemu naslov (natuknica) članka ima ulogu indeksne odrednice, a uputnica ulogu oznake mjesta. Ukoliko neko enciklopedičko djelo ima kazalo, tada slijepa natuknica zapravo i nisu prijeko potrebne te ih je moguće svesti na najmanju moguću mjeru. No, ukoliko djelo nema kazala, dobro je imati što više slijepih natuknica tj. bilo bi dobro sve nazive svih pojmova koji su obrađeni u djelu obraditi kao slijepa natuknica, ukoliko ne postoje zasebni članci pod tim nazivima. U nekim se slučajevima, kod složenih članaka mjesto objašnjenja jednoga ili više značenja naziva može se javiti samostalna uputnica.

Uputnice u tekstu vezane su uz neki pojam koji se spominje, a treba ih biti što manje jer prekidaju tijek usvajanja teksta prilikom čitanja članka. Objašnjenje pojedinog pojma može se naći u članku s nekim drugim nazivom, pa je na njega potrebno uputiti npr. uputnicom u zagradi, koja slijedi iza pojma. Na kraju teksta pojedinog članka, ponekad i na kraju nekog poglavlja u članku, dolaze uputnice sa značenjem »vidi još« tj. tzv. uputnice na kraju teksta. One upućuju na članke koje bi bilo dobro proučiti kao nadopunu osnovnom članku. Obično se stavljaju iza točke, mogu biti i u zagradi ili se mogu i odvojiti kao zaseban dio članka naslovljen npr. *Vidi još, Povezani članci, Za one koji žele znati više* i sl.

Iza teksta objašnjenja, na kraju članka, u nekim enciklopedičkim djelima dolaze dopunski podaci koji, osim već spomenute uputnice na kraju članka, obuhvaćaju podatke o autoru članka, preporučenu literaturu i poveznice na vanjske izvore informacija vezane uz članak.

6.3 Vrste enciklopedijskih članaka

Enciklopedijski članci pojedine struke ili tematskog područja iste obavijesne razine organizirani su na sličan način. Tako će isto biti koncipirani članci koji obrađuju pojedine države, gradove, rijeke, pojedine biljne ili zoološke vrste, kemijske elemente ili spojeve, svemirske letjelice ili rakete, čovječje ili životinjske bolesti.

Međutim, postoje i članci koji se izdvajaju od ostalih po svojoj organizaciji, bez obzira na struku kojoj pripadaju. Takvi članci su slijepa natuknica, pregledni članak s uputnicama,

složeni članak i biografski članak. Sličan pristup razvrstavanja enciklopedijskih članaka Hrvatske enciklopedije prikazan je u internoj brošuri³²⁴, prema kojoj su članci Hrvatske enciklopedije³²⁵ zasnovani na tri vrste članaka:

- *veliki pregledni članci* („kompleksni“ ili složeni članci) koji obrađuju više struka (npr. članci o državama i narodima, književnost, gospodarstvo itd.),
- *makropedijski članci* koji obrađuju sadržaj vezan uz jednu struku, iako mogu sadržavati mnogobrojne pojmove i nazivke (termine),
- *mikropedijski članci* koji obrađuju jedan pojam ili jednu osobu, bez obzira na broj redaka.

Enciklopedijski članci, osim po veličini, razlikuju se i po sadržaju. Tako postoje šest skupina članaka koji tvore pojmovnik (*»nominarij«* i *»terminarij«*) i četiri skupine koje tvore personarij (*biografije djelatnih osoba*) tj. ukupno postoji deset osnovnih tipova članka:

- *Veliki pregledni članci (»kompleksni« ili složeni članci)* samo u iznimnim slučajevima (*Hrvatska, Republika; Hrvati; Sjedinjene Američke Države; Njemačka i sl.*) i mogu imati do 5000 redaka,
- *Veliki članci* koji prikazuju pojedine znanosti, područja umjetnosti, fizičkih, kulturnih i drugih cjelina (filozofija, fizika, medicina, kemija; glazba, slikarstvo, književnost; antika, renesansa; svijet: brodogradnja: narodni preporod; gospodarstvo itd.); takvi članci imaju 100 do 500 redaka,
- *Veliki biografski članci* - objašnjenje Odjeljku 6.4.,
- *Značajni pregledni članci* obrađuju uža područja znanosti, umjetnosti, povijesti, (leksikografija, geofizika, hrvatsko-ugarska nagodba itd.); imaju 40 do 100 redaka,
- *Značajni biografski članci* - objašnjenje u Odjeljku 6.4.,
- *Pregledni članci* prikazuju pojedine događaje, institucije i sl.: imaju 20 do 40 redaka,
- *Standardni biografski članci* - objašnjenje u Odjeljku 6.4.,
- *članci niže obavijesne razine o uzgrednim pojmovima i pojavama (Narodni dom, dvorana i dr.)*, 10 do 20 redaka;
- *Manji biografski članci* - objašnjenje u sljedećem odjeljku,
- *Mali članci*, do 10 redaka.

³²⁴ Hrvatska Enciklopedija: uredništvo, organizacija, upute. Zagreb: Leksikografski zavod Miroslav Krleža, 1998. Str. 6.

³²⁵ Hrvatska enciklopedija. Zagreb: Leksikografski zavod Miroslav Krleža, 1999- 2009.

6.4 Struktura i analiza vrste podataka biografskog enciklopedijskog članka

Dakle, prethodne podjele upućuju na to da se biografski članci izdvajaju od ostalih bez obzira na struku kojoj pripadaju te da ih tvore četiri vrste članaka s obzirom na veličinu (veliki biografski članci, značajni pregledni članci, standardni biografski članci, manji biografski članci), a po sadržaju tvore personarij Hrvatske enciklopedije.

Biografski članci su u enciklopedičkom djelu posvećeni pojedinim osobama. Ujednačeni su u pogledu strukture i slijeda podataka o životu i djelovanja osobe, opsega kojega zauzimaju imajući u vidu značaj osobe koja se u njima obrađuje. Tako se primjerice, suradnicima Hrvatske enciklopedije³²⁶ biografski enciklopedijski članci prema značaju osoba razvrstavaju na:

- *velike biografske članke* koji donose velike biografije malog broja stvaralaca „izvan kategorija“, kao što su npr. *Dante, Shakespeare, Marulić, Bethoven, Zajc* i sl., obavezno prikazujući recepciju u Hrvatskoj za strane stvaraoce i u svijetu za hrvatske velikane; članci sadrže 100 do 200 redaka,
- *značajne biografske članke* koji donose životopise značajnih stvaralaca, kao što su npr. *Dickens, Churchill, Trumbić, Gjalski, Jagić, Radić* i sl., a sadrže 30 do 100 redaka,
- *standardne biografske članke* koji donose biografije značajnih stvaraoaca, kao što su *Johnson, Samuel; Belostenec, Ivan; Jambrešić, Andrija* i dr., a sadrže do 15 redaka,
- *manje biografske članke* koji pretežno donose životopise suvremenika, a sadrže do 15 redaka.

Prema istim uputama navodi se: „Članak počinje imenom (prezimenom i „danim“ imenom, imenom i nadimkom, imenom i brojem itd.) u poludebelom kurentu. Iza stranog prezimena (ili imena), ako glasovna vrijednost latiničke grafije bitno odudara od hrvatskog izgovora, bilježi se transkripcija u uglatim zagradama... Slijedi zatim u oblim zagradama mjesto i datum rođenja te mjesto i datum smrti. Od činjenica iz životopisa donosi se ono što je relevantno za djelovanje i opus; navodi se školska formacija i profesija; doktorat se navodi, u načelu, za domaće osobe (za strane iznimno, ako je relevantno). Za domaće i strane osobe navodi se članstvo (djelovanje) u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama, prestižne nagrade i priznanja. Djelatnost se – ovisno o važnosti osobe – iznosi u svim strukovnim razdjelima ili pak samo u onoj struci u kojoj je osoba ostavila vidan trag.

³²⁶ Hrvatska Enciklopedija: uredništvo, organizacija, upute, op. cit. (bilj. 324), Str. 6-7.

Djela se navode u reprezentativnom izboru... Na kraju članka (vrlo) selektivno navodi se literatura.³²⁷

Nakon opisa strukture i redoslijeda elemenata enciklopedijskog biografskog članka, cilj ovog odjeljka je analizirati vrstu i oblik podataka koji se pojavljuju u biografskom enciklopedijskom članku Hrvatske enciklopediji za područje hrvatska književnost. Navedena analiza vrste podataka biografskog enciklopedijskog članka doprinijet će izgradnji ontologije književnosti jer ukazuje na vrste podataka koji će biti uključeni u njezinu izgradnju. Na temelju prethodno iznesene strukture enciklopedijskog članka s posebno istaknutim odlikama biografskog enciklopedijskog članka slijedi pregled vrste podataka biografskog enciklopedijskog članka hrvatske književnosti u Hrvatskoj enciklopediji u nastavku teksta i Tablici 8.

Glava članka:

- Natuknica (naslov članka) – jednočlano ili višečlano ime (jednočlano ime ili višečlano ime tj. prezime s »danim« imenom u inverznom redoslijedu, ime i nadimak, ime i broj itd.). Kako se radi o hrvatskim imenima i prezimenima izostavljene su dodatne obavijesti o natuknici (transkripcija, transliteracija, etimologija, terminološke odrednice).

Tijelo članka

- Definicija biografskih članaka – analizira se viši rodni pojam (npr. književnik, crkveni pisac), te specifična razlika tj. nacionalna pripadnosti (narodnost), mjesta i datumi rođenja i smrti u obliku zagrada.
- Tekst članka – bez obzira na struku kojoj pripada biografski enciklopedijski članci se izdvajaju po svojoj organizaciji jer su ujednačeni u pogledu strukture i slijeda podataka donoseći činjenice iz životopisa (informacije koje se odnose na školsku formaciju i profesiju, doktorat, zvanje, članstvo u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama, prestižne nagrade i priznanja, djelatnost) i činjenice o književnikovom opusu (djela navedena u reprezentativnom izboru i godine njihovog objavljivanja).

³²⁷ Ibid.

Tabela 8: Opis vrste i oblika podataka biografskih enciklopedijskih članaka hrvatske književnosti

BIOGRAFSKI ČLANAK HE	PODATAK	OBLIK PODATKA
<i>Glava članka</i>		
<i>Natuknica</i>	- ime osobe	- jednosložno ime (Aristotel), - višesložno sadržeći ime s prezimenom u inverznom redosljedu (Shakespeare, William), - tročlano rimsko ime (Ciceron, Marko Tulije), - osnovno ime i kakav atribut (pridjev, redni broj i sl.) pa se donose u prirodnom redosljedu (Karlo Veliki, Leonardo da Vinci, Petar Krešimir IV)
		- plemićke oznake i naslovi i dr. srodni pridjevci stavljaju se iza imena (Adamović Čepinski, Bela pl.), - osobe koje su proglašene svetima, oznaka sv. stavlja se iza imena (Franjo Asiški, sv.)
		- ukoliko je nepoznat dio imena, navode se inicijali osobnog imena.
<i>Tijelo članka</i>		
<i>Definicija</i>		
viši rodni pojam	pojam	- npr. književnik, esejist, pjesnik i sl.
	nacionalna pripadnost	- hrvatski
	stručna pripadnost, zvanje, zanimanje	- pjesnik, pravnik itd
Datum u okruglim zagrada	datum rođenja	datum.mjesec.godina
	datum smrti	datum.mjesec.godina
mjesto rođenja slijedi u okruglim zagrada iza imena i podataka o nacionalnoj pripadnosti i stručnoj djelatnosti (npr., njem. pjesnik)	mjesto rođenja	ime grada

mjesto smrti slijedi u okruglim zagradama iza imena i podataka o nacionalnoj pripadnosti i stručnoj djelatnosti (npr., njem. pjesnik)	mjesto smrti	ime grada
<i>Tekst članka</i>		
Prvi odjeljak članka	životopis osobe	tekst – dio teksta
	školovanje osobe	tekst – 1-2 rečenice: naziv škole, fakulteta i mjesta školovanja, godina školovanja
	poslovi koje je osoba obavljala	tekst – 1-2 rečenice: naziv poslova, ustanova i mjesta obavljanja, godina rada
Prvi odjeljak članka – dio koji se odnosi na životopis osobe.	prebivalište (zemlja, država/pokrajina, grad itd. u kojem osoba pribiva ili je pribivala)	tekst: ime mjesta, države, grada itd.
Prvi odjeljak članka – dio koji se odnosi na životopis osobe (članstvo ponekad i na samom kraju članka).	pripadnost (skupina kojoj osoba pripada ili je pripadala kroz zaposlenje, članstvo, kulturni identitet itd.)	tekst: nazivi ustanova, organizacija, društava te godina ili razdoblje djelovanja
Drugi odjeljak članka	naslovi književnih djela	tekst: <i>naslov djela</i>
	podatak o odgovornosti	tekst: napisao, uredio, preradio itd.
	godina objavljivanja djela	godina u zagradi iza naslova djela
	vrsta književnih djela	tekst: riječi, npr. roman, pjesme, eseji i sl.
Drugi odjeljak članka koji daje podatke o stvaralaštvu osobe (obično na samom početku, a ponekad i na samom kraju ovog dijela članka)	jezik osobe (jezik koji osoba koristi pri stvaranju izraza za objavljivanje, emitiranje itd.), vrsta stiha, strofe itd.	tekst: jezik kojim je osoba napisala književno djelo
Drugi odjeljak članka koji daje podatke o stvaralaštvu osobe	nagrade	tekst: uvijek se spominje riječ nagrada s imenovanjem dodijeljene nagrade

Analizom dobivenog opisa vrste i oblika podataka dobiva se uvid u vrste informacija koje se nalaze u biografskim enciklopedijskim člancima iz područja hrvatske književnosti u Hrvatskoj enciklopediji. Ovako dobiveni uvid u vrstu informacija omogućuje postavljanje temeljnih aspekata informacija (*životopis, osobne informacije, školovanje, posao, djelo, autorstvo stvaralaštvo*), koji će zapravo omogućiti postavljanje temeljnih faceta ontologijskih slojeva što omogućuje facetno pretraživanje tj. sužavanje skupa rezultata koji se dobiju prvotnim pretraživanjem/pregledavanjem s obzirom na kriterije koji su omogućeni facetnim pretraživanjem. Dobiveni oblik podataka za određenu vrstu podataka će se primijeniti u sljedećem poglavlju prilikom razvoja ontologije književnosti.

6.5 Zaključak

Poglavlje je teoretskim opisom strukture i vrsta enciklopedijskih članaka prikazalo postojanost i definiranost ujednačene strukture enciklopedijskih članaka kojih se pridržavaju enciklopedisti Leksikografskog zavoda Miroslav Krleža prilikom stvaranje Hrvatske enciklopedije. Prikazom odlika enciklopedijskih članaka, a posebno biografskih enciklopedijskih članaka, ukazalo se na vrstu i oblik podataka koje sadrže biografski članci te su time uspostavljeni temeljni ontologijski slojevi tj. facete prema kojima će se razvrstavati ontologijski odnosi u ontologiji književnosti, kao što je utvrđen i oblik podataka za pojedine vrste informacija, što će se primijeniti prilikom razvoja ontologije književnosti u sljedećem poglavlju.

7 RAZVOJ ONTOLOGIJE KNJIŽEVNOSTI

7.1 Uvod

Ontologija je formalni opis koncepata neke domene, njihovih međusobnih odnosa i svojstava (atributa) te ograničenja nad njima. Jedna od prvih definicija ontologije kaže da ontologija obuhvaća osnovne pojmove i odnose koji čine rječnik područja istraživanja kao i pravila za kombiniranje pojmova i odnosa u cilju definiranja proširenja rječnika³²⁸. Kako bi ontologija mogla valjano predstaviti domenu na koju se odnosi potrebno je u njenom istraživanju obuhvatiti koncepte koji je opisuju, odnose među njima, kao i attribute koji opisuju koncepte. Taj cilj će se ovim istraživanjem ostvariti za enciklopedijske članke *Hrvatske enciklopedije* iz područja hrvatska književnost.

Ontologija hrvatske književnosti biti će rezultat istraživanja korpusa biografskih enciklopedijskih članka *Hrvatske enciklopedije* iz područja hrvatske književnosti. Ekstrahirat će se književni pojmovi koji se pojavljuju unutar enciklopedijskih članka i odnosi kako bi se obuhvatilo znanje koje se odnosi na hrvatske književnike (životopisne podatke o osobama, obilježjima njihovog stvaralaštva) i djela hrvatske književnosti. Ontologija će se dobiti svrstavanjem pojmova u hijerarhiju, definiranjem odnosa među njima i atributa koji ih opisuju.

Ovako dobivena ontologija je deskriptivna ontologija jer predstavlja znanje iz domene hrvatske književnosti te će se upotrijebiti za automatsko zaključivanje o tako predstavljenom području.³²⁹ Uzorak će uključiti korpus Hrvatske enciklopedije tj. stručno područje k10 - hrvatska književnost. Korpus sadrži 1170 enciklopedijskih članaka, koji su ujednačeni u pogledu slijeda podataka o životu osobe i opsega koji zauzimaju imajući u vidu značaj osobe koja se u njima obrađuje. Tip enciklopedijskih članaka je biografski. Razvijena ontologija poslužiti će za organizaciju, pretraživanje i pregledavanje znanja pohranjenog u mrežnoj

³²⁸ Neches, Robert...et al. Enabling Technology For Knowledge Sharing. AI Magazine 12 , 3 (1991), 40.

³²⁹ Deskriptivne ontologije se nazivaju sheme izgrađene za opis domene, gdje termini na čvorovima predstavljaju individue ili klase entiteta stvarnog svijeta. Objašnjenje preuzeto iz Classification & ontology: formal approaches and access to knowledge : proceedings of the International UDC Seminar, 19 - 20 September 2011, The Hague, The Netherlands / edited by Aida Slavic and Edgardo Civallero. Wuerzburg : Ergon Verlag, 2011. Str. 78.

online Hrvatskoj enciklopediji, kao i za dobivanje preciznih odgovora na postavljena pitanja, što sada nije slučaj.

Poglavlje razvoj ontologije književnosti utvrdit će gradivne ontologijske elemente iz enciklopedijskog biografskog članka, prikazati će metodologiju primijenjenu u razvoju ontologije, kao i dobivenu pojmovnu taksonomiju i ontologijske odnose. Tako dobiveni rezultati istraživanja će se predstaviti kroz prikaz ontologijskih modula i pripadajućih im ontologijskih odnosa. Također će se dobiti svi ontologijski odnosi koji će opisivati biografski enciklopedijski članak iz područja hrvatske književnosti. Konačni rezultati će se prikazati kroz primjenu tako dobivene ontologije u opisu individue, kao i kroz mogućnost postavljanja složenih semantičkih pitanja nestrukturiranom enciklopedijskom biografskom članku, što do sada nije bilo moguće.

7.2 Gradivni ontologijski elementi enciklopedijskog biografskog članka iz područja književnosti

Izgradnja ontologije iz enciklopedijskih biografskih članaka područja hrvatske književnosti izvest će se iz strukturalnih elemenata enciklopedijskog članka opisanih u prethodnom odjeljku. U cilju izgradnje ontologije koja će predstaviti domenu književnosti dobivenu iz područja hrvatske književnosti predstavljene u Hrvatskoj enciklopediji obuhvatit će se pojmovi koji pripadaju domeni književnosti te će se razviti njihova taksonomija, atributi pojmova i odnosi među njima kako bi isti predstavljali gradivne elemente ontologije književnosti. Navedeni gradivni elementi ontologije će se ekstrahirati iz odabranog korpusa enciklopedijskih članaka. Korpus predstavlja 1170 biografskih enciklopedijskih članaka hrvatske književnosti iz Hrvatske enciklopedije. Izvršila se poluautomatska i ručna ekstrakcija termina potrebnih za izgradnju ontologije.

Cilj ovog odjeljka je prikazati ulogu već prethodno predstavljenih strukturalnih elemenata enciklopedijskih biografskih članaka Hrvatske enciklopedije u izgradnji ontologije kao njezinih gradivnih elemenata. Dakle, ekstrahirani elementi enciklopedijskih biografskih članaka hrvatske književnosti iz Hrvatske enciklopedije su gradivni ontologijski elementi sa sljedećim ulogama:

Glava članka

- natuknica (naslov članka) - strojna ekstrakcija (uglavnom jednočlano ili višečlano ime tj. prezime, s »danim« imenom u inverznom redoslijedu, imenom i nadimkom, imenom i brojem itd.): Ekstrahirane natuknice predstavljaju *individue* ontologije s opisnim ontologijskim odnosom *ime*. Ukoliko određena osoba npr. posjeduje pseudonim ili nadimak to će se naznačiti u opisu *individue* kao *altLabel* opisni odnos. *Individue* generirane od natuknica predstavljat će članove klase *Književnik*, koja je u biti viši rodni pojam enciklopedičke definicije, kao i *individue* klasa pojedinih zanimanja npr. Učitelj, Pedagog itd.
- dodatne obavijesti o natuknici (uglavnom lingvističke naravi, ukoliko su prisutni u članku) – ručno se ekstrahiraju istoznačnice i istovrijednice te dobiveni podaci predstavljaju opisna svojstva *individua* ontologije.

Tijelo članka

- enciklopedička definicija- ručno se ekstrahira viši rodni pojmovi (nadređeni pojam, genus proximum) koji će predstavljat *ontologijske klase* kojima pripadaju *individue* tj. natuknice članka (npr. književnik, leksikograf, pedagog i sl.) i specifične razlike tj. ograničavajuće svojstvo (*differentia specifica*) kao što su nacionalna pripadnost (hrvatski – *individue* ontologijske klase Nacionalnost, kao što su hrvat, poljak i sl. te *objektno ontologijskog svojstvo* o narodnosti); podaci o mjestu i datumu rođenja te mjestu i datumu smrti (automatska ekstrakcija *objektnih, podatkovnih i opisnih odnosa ontologije*). Podaci o godini rođenja i smrti također će se izdvojiti zasebno kao *individue* klase *Godina* (podklasa klase *Vrijeme*) te će iz istih proizaći *podatkovni odnosi* za godine rođenja i smrti.
- tekst članka koji donosi činjenice iz životopisa relevantne za djelovanje (informacije koje se odnose na školsku formaciju i profesiju, doktorat, zvanje, članstvo u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama, prestižne nagrade i priznanja, djelatnost) i činjenice o književnikovom opusu (djela navedena u reprezentativnom izboru i godine njihovog objavljivanja) – ručno ekstrahirani podaci predstavljaju *klase* te *objektna, podatkovna i opisna ontologijska svojstva*. Podaci poput naslova djela i godine njihovog objavljivanja se automatski ekstrahiraju te predstavljaju *individue*

klase *Književno djelo* i *Godina*, dok će se ontologijski odnosi koji opisuju književno djelo i njegov odnos s književnikom načiniti ručno.

7.3 Metodologija

Veliki je broj metodologija razvoja ontologija jer autori raznih alata ujedno preporučuju metode koje su najprimjerenije za te alate. Ne postoji samo jedan ispravan način ili metodologija za razvijanje ontologija. Odluka o izboru određene metodologije ovisi o tome u kojem se trenutku razvoja ontologije koriste, koji su dostupni izvori znanja te koje su mogućnosti odabranog alata.

Osnovne metode razvoja ontologija, kako im i sami naziv upućuje, uključuju cijeli postupak razvoja ontologije. U njih se može ubrojiti Cyc metoda, metoda Uschold i King³³⁰, metoda Grüninger i Fox³³¹, KACTUS ili SENSUS, METHONTOLOGY³³² i On-To-Knowledge. Uschold i King metodologija koja se temelji na prikupljenom iskustvu razvijanja *Enterprise Ontology*³³³ te Grüninger i Fox metodologije koja se temelji na iskustvu razvijanja TOVE (*TOronto Virtual Enterprise*) ontologijskog projekta³³⁴, obje u oblikovanju domene poduzetništva, 1995 godine predloženi su prvi metodološki obrisi koji su kasnije i pročišćeni u Uschold³³⁵ i Uschold & Grüninger³³⁶ metodologijama. Bernaras et al.³³⁷ predstavili su metodu upotrijebljenu u izgradnji električnih mreža kao dio Esprit KACTUS projekta.

³³⁰ Uschold, Mike. *Building Ontologies: Towards a Unified Methodology*. Edinburgh : Artificial Intelligence Applications Institute, University of Edinburgh, 1996. Dostupno na URL: <http://www.aii.ed.ac.uk/publications/documents/1996/96-es96-unified-method.pdf> (29.01.2016.)

³³¹ Gruninger, Michael; Fox, Mark S. *Methodology for the Design and Evaluation of Ontologies* // International Joint Conference on Artificial Intelligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal, Canada, 1995. Dostupno na URL: <http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf> (29.01.2016.)

³³² Gómez-Pérez A.; Fernández López, M.; Corcho, O. *Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London ; New York: Springer, 2004. Str. 105.

³³³ Uschold, Mike; loc cit. (bilj. 330).

³³⁴ Gruninger, Michael; Fox, Mark S., loc cit. (bilj. 331).

³³⁵ Uschold, Mike, loc cit. (bilj. 330).

³³⁶ Uschold, M.; Grüninger, M. *Ontologies: Principles, Methods and Applications*. Knowledge Sharing and Review 11, 2 (1996). Dostupno na URL: <http://www.upv.es/sma/teoria/sma/onto/96-ker-intro-ontologies.pdf> (29.01.2016.).

³³⁷ Bernaras, A.; Laresgoiti, I.; Corera, J. *Building and Reusing Ontologies for Electrical Network Applications* // ECAI 96 : 12th European Conference on Artificial Intelligence, August 11-16, 1996, Budapest, Hungary : proceedings / edited by Wolfgang Wahlster. Chichester ; New York : Wiley, 1996. Str. 298-302.

METHONTOLOGY³³⁸ se pojavljuje u isto vrijeme te se proširuje u kasnijim radovima^{339,340}. Godinu dana kasnije, 1997, predložena je metodologija za izgradnju ontologija temeljenih na SENSUS ontologiji³⁴¹. Pristup Amaya Berneras et al. je povezan s KACTUS projektom³⁴², kojemu je jedan od ciljeva istražiti izvedivost ponovne upotrebe znanja u složenim tehničkim sustavima i ulozi ontologija koje će ih podržati³⁴³. Metodologija temeljena na razvoju SENSUS ontologije je namijenjena upotrebi u procesima obrade prirodnog jezika (engl. natural language processing), razvijena od strane ISI radne grupe za prirodni jezik (Information Sciences Institute) kako bi osigurala široko usmjerenu pojmovnu strukturu za razvijanje strojnog prevođenja^{344,345}. Njegov trenutni sadržaj je dobiven ekstrakcijom i sjedinjavanjem (engl. merging) informacija iz različitih elektroničkih izvora znanja. Ovaj proces je započeo sjedinjavanjem PENMAN Upper Model³⁴⁶ i ONTOS (dvije vrlo visoke razine jezično-temeljene ontologije) te semantičkih kategorija iz rječnika kao ručno proizvedenog ontologijskog temelja. Zatim je upotrijebljen alat za spajanje kako bi se spojio WordNet s engleskim rječnikom. U svrhu podrške strojnom prevođenju, rezultati ovog spajanja su zatim povećani španjolskim i japanskim leksičkim entitetima iz Collins španjolsko/engleskog rječnika i Kenkyusha japansko/engleski rječnik³⁴⁷. SENSUS ima više

³³⁸ Gómez-Pérez, A. A Framework to Verify Knowledge Sharing Technology. *Expert Systems with Application* 11, 4(1996), 519-529.

³³⁹ Fernández-Lopez, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. // *AAAI-97 Spring Symposium Series SS-97-06*, 1997, 33-40. Dostupno na URL: http://oa.upm.es/5484/1/METHONTOLOGY_.pdf (8.2.2016.).

³⁴⁰ Fernández-López, M. ...et al. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *IEEE Intelligent Systems and Their Applications* 14, 1(1999), 37-46. Dostupno na URL: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=747904> (8.2.2016.).

³⁴¹ Swartout, Bill...et al. Toward Distributed Use of Large-Scale Ontologies. // *AAAI Technical Report SS-97-06*. Dostupno na URL: <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-018.pdf> (8.2.2016.).

³⁴² The KACTUS Booklet version 1.0. Esprit Project 8145. September, 1996. Dostupno na <http://www.citeulike.org/user/eduardocruz/article/3926072> (8.2.2016.).

³⁴³ Schreiber, G.; Wielinga, B.; Jansweijer W. The KACTUS View on the 'O' World. // In *Proceedings of the National Dutch AI Conference. NAIC'95*. 1995. Dostupno na URL: <http://www.math.vu.nl/~guus/papers/Schreiber95a.pdf> (8.2.2016.).

³⁴⁴ Knight, Kevin; Luk, Steve K. Building an Large Knowledge Base for Machine Translation. // *Proceedings of the American Association of Artificial Intelligence Conference (AAAI-94)*. Seattle (USA). 1994. Dostupno na URL: <https://www.aaai.org/Papers/AAAI/1994/AAAI94-118.pdf> (8.2.2016.).

³⁴⁵ Knight, Kevin...et al. Filling Knowledge Gaps in a Broad-Coverage MT System. *Proceedings of the 14th IJCAI Conference*. Montreal (Canada). 1995. Dostupno na URL: <http://arxiv.org/pdf/cmp-lg/9506009.pdf> (8.2.2016.).

³⁴⁶ Bateman, John A...et al. A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model. // USC/Information Sciences Institute. Marina del Rey, 1990. Dostupno na URL: <http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/um89/UM89-penman.pdf> (8.2.2016.).

³⁴⁷ Swartout, Bill...et al., loc cit. (bilj. 341).

od 50,000 pojmova organiziranih u hijerarhiju, uključujući termine s visokom i srednjom razinom abstrakcije, ali općenito govoreći ne pokriva termine iz specifične domene. Termini domene su povezani sa SENSUS-om u svrhu izgradnje ontologija za određene domene, a bilo koji nerelevantni pojmovi su pročišćeni u SENSUS-u.

Metode za reinženjerstvo ontologija se primjenjuju prilikom postupka stvaranja nove ontologije iz postojećih izmjenjujući ih prema zahtjevima nekog korisnika. Metode za spajanje ontologija se primjenjuju postupkom razvijanja nove ontologije spajajući već postojeće ontologije. Metodama za reinženjerstvo i metodama za spajanje ontologija se skraćuje cijeli postupak razvoja ontologija te se naglašava distribuiranost postojećih ontologija. Metode tzv. učenja ontologija nisu izričite metode razvoja ontologija jer se uglavnom radi o djelomičnom automatiziranom procesu prikupljanja znanja za razvoj ontologije koji se temelji na analizi prirodnog jezika i tehnikama strojnog učenja.

Proučavanjem različitih razvojnih metodologija se također mogu identificirati dvije glavne skupine³⁴⁸: (a) metodologije temeljene na iskustvu, kao što je metodologija koju predlažu Grüninger i Fox³⁴⁹ na temelju TOVE³⁵⁰ projekta ili druge kao npr. Uschold i King³⁵¹, Uschold & Gruninger³⁵² iz domene poduzetništva izdane 1995. godine; (b) metodologije koje predlažu razvojne uzorke modela, kao što je METHONTOLOGY³⁵³, koja predlaže niz aktivnosti za razvoj ontologije na temelju njezinog životnog ciklusa i doradivanom uzorku^{354,355}, metodom iterativnog pristupa razvoju ontologije. METHONTOLOGY metodologija je razvijena u laboratoriju umjetne inteligencije na Politehničkom sveučilištu u Madridu. METHONTOLOGY okvir^{356,357,358} omogućava razvijanje ontologije na razini

³⁴⁸ Wache H...et al. Ontology-Based Integration of Information: A Survey of Existing Approaches. // Proc. International Joint Conference on Artificial Intelligence; Workshop: Ontologies and Information Sharing Workshop: Ontologies and Information Sharing, Seattle, WA, 108-117.

³⁴⁹ Grüninger, Michael; Fox, Mark S., loc cit. (bilj. 331).

³⁵⁰ Enterprise Modelling Methodology. Dostupno na URL: <http://www.eil.utoronto.ca/enterprise-modelling/entmethod/index.html> (8.2.2016.).

³⁵¹ Uschold, Mike, loc cit. (bilj. 330).

³⁵² Uschold, M.; Grüninger, M. loc cit. (bilj. 336).

³⁵³ Gómez-Pérez, A.; Fernández López, M.; Corcho, O., op. cit. (bilj. 332), str. 67.

³⁵⁴ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

³⁵⁵ Ibid.

³⁵⁶ Fernández-Lopez, M.; Gómez-Pérez, A.; Juristo, N., loc cit. (bilj. 339).

³⁵⁷ Gomez-Perez, A. Knowledge Sharing and Reuse: ontologies and applications / A Tutorial on Ontologies Engineering, IJCAI, 1999. Str. 11.

³⁵⁸ Fernández-López, M. ...et al., loc cit. (bilj. 340).

znanja i uključuje³⁵⁹: utvrđivanje ontologijskog razvojnog procesa, temeljeći životni ciklus ontologije na razvoju uzoraka i određene tehnike za provođenje svake aktivnosti. METHONTOLOGY okvir je podržan od strane ODE^{360,361}. Obično, prvi način (a) se primjenjuje kada su zahtjevi jasno poznati na početku, a drugi (b) kada ciljevi nisu jasni od početka. Štoviše, uobičajeno je spajanje različitih metodologija jer svaka od njih pruža stvaranje zamisli koje ga razlikuju od drugih. Ovo sjedinjavanje ovisi o ontologijskim korisnicima i ciljevima. S druge strane, kao i bilo koja druga aktivnost konceptualnog modeliranja, izgradnja ontologije mora biti podržana tehnikama softverskog inženjerstva (engl. software engineering techniques)³⁶². Dakle, koriste se metode i alati iz software-skog inženjerstva (engl. software engineering) kako bi se podržale ontologijske inženjerske aktivnosti (engl. engineering activities). Općenito, razvoj ontologije se može podijeliti u dvije glavne faze: specifikaciju i konceptualizaciju. Cilj faze specifikacije je postići neformalno znanje o domeni. Cilj faze konceptualizacije je organizirati i strukturirati to znanje koristeći predstavljanje (opisivanje) neovisno o implementacijskom jeziku i okruženjima. Cilj je prikazati kako se prilagodila metodologija u razvoju ontologije iz područja književnosti za enciklopedijske biografske članke te kako se upotrijebila softverska inženjerska tehnika za njezino definiranje.

U razvoju ontologije biografskih enciklopedijskih članaka (područje hrvatska književnost) primijenjena je metodologija proizašla je iz općenitih metodologija razvoja ontologije koje su prethodno prikazane u ovom odjeljku, a napose METHONTOLOGY metodom.³⁶³ U skladu s ciljem disertacije i prikazanim metodologijama, ontologija biografskih enciklopedijskih članaka (područje hrvatska književnost) će se razviti njihovom analizom, identifikacijom termina i veza među njima te definiranjem vrsta entiteta koji čine ontologiju. Metodologijski koraci primijenjeni u razvoju ontologije su sljedeći³⁶⁴:

³⁵⁹ Blázquez, M...et al. Building Ontologies at the Knowledge Level using the Ontology Design Environment. // Proceedings of the 11th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'98), 2, 1998. Dostupno na URL: http://oa.upm.es/6457/1/Building_Ontologies_at_the_K.pdf (8.2.2016.).

³⁶⁰ Ibid.

³⁶¹ Fernández-López, M. ...et al., loc cit. (bilj. 340).

³⁶² Falbo, Ricardo De Almeida. Experiences in Using a Method for Building Domain Ontologies. // SEKE 2004 : the 16th International Conference on Software Engineering & Knowledge Engineering : proceedings, technical program, July 20-24, 2004, Banff Alberta, Canada. Str. 474-477. Skokie, Il. : Knowledge Systems Institute, 2004.

³⁶³ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

³⁶⁴ Ahmed, S.; Kim, S.; Wallace, K. A methodology for creating ontologies for engineering design / Journal of Computing and Information Science in Engineering 7, 2(2007), 132-140. Dostupno na URL: <http://computingengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1400756> (8.2.2016).

1. Identifikacija postojeće ontologije najviše razine apstrakcije je provedena analizom postojećih istraživačkih projekata te je kao najprikladnija odabrana PROTON ontologija,
2. Identificiranje ključnih pojmova u promatranom području – izvršit će se analizom odabranog korpusa tj. biografskih enciklopedijskih članka,
3. Stvaranje taksonomije za pojmovne rječnike izvršeno je analizom biografskih enciklopedijskih članka,
1. Testiranje za opis znanja izvršeno je preslikavanjem instanci u taksonomiju rječnika, što je izvršeno stvaranjem individue koje se opisuju stvorenim ontologijskim rječnikom.

U cilju uspješnog izvršenja svih navedenih koraka provedeni su postupci kojima se određuje ontologijska domena i doseg, razmatra se ponovna upotreba neke već postojeće ontologije, određuju se važni termini tj. stvara se terminološki rječnik, definiraju se ontologijske klase i klasna hijerarhija, određuju se ontologijski odnosi i stvaraju se individue. U nastavku slijedi prikaz navedenih postupaka.

7.3.1 Određivanje ontologijske domene i dosega

Definicija cilja i opsega ontologije smatra se prva aktivnosti u METHONTOLOGY metodi.³⁶⁵ Opseg ograničava ontologiju, navodeći što mora biti uključeno, a što ne smije. To je važan korak za umanjivanje količine podataka i pojmova koji će se analizirati. U tu svrhu je potrebno napraviti prethodnu analizu domene, stoga Noy i McGuinness³⁶⁶ predlažu nabranje važnih termina. Početak razvoja ontologije započinje definiranjem njezine domene i dosega, što podrazumijeva odgovor na nekoliko osnovnih pitanja:

- Koji će domenu ontologija pokrivati?
- Za što će se koristiti ontologija?
- Za koju vrstu pitanja trebaju osigurati odgovore informacije u ontologiji?
- Tko će koristiti i upravljati ontologijom?

Ovim radom se razvija *ontologija domene*, vrsta ontologije koja opisuje određeno područje na općem nivou, u ovom slučaju opisuje se područje književnosti u okviru nacionalnih književnosti tj. konkretno hrvatske književnosti. Stoga se može utvrditi kako je

³⁶⁵ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

³⁶⁶ Ibid.

predstavljanje enciklopedijskog znanja hrvatske književnosti domena ontologije ovog rada. Ontologija književnosti dobivena ovim radom obuhvatit će enciklopedijsko znanje vezano uz pojedinu nacionalnu književnost tj. u ovom slučaju hrvatsku književnost. Dobiveni ontologijski pojmovi i odnosi će biti upotrebljivi za opis bilo koje nacionalne književnosti. Dakle, konkretno se ova ontologija planira upotrijebiti za opis enciklopedičkog znanja hrvatske književnosti, njegovu bolju organizaciju te dobivanje preciznih odgovora na pojedina pitanja o hrvatskim književnicima i hrvatskoj književnosti.

Ontologijski pojmovi opisivat će različite vrste književnih djela i književno područje prema nacionalnoj pripadnosti tj. u ovom slučaju hrvatske književnosti, što znači kako će se moći opisati znanje o književnim djelima i književnicima specifično za određeni narod te će tako dobivena ontologija omogućiti proučavanje nacionalnih književnosti jer će sadržati sve potrebne odnose za opis književnosti bilo kojeg naroda. Obuhvaćen je veliki broj odnosa potreban za opis produkcije književnika pojedinih nacionalnih književnosti, njihovih životopisa, međusobnih odnosa i cjelokupnog znanja koje se nalazi u biografskim enciklopedijskim člancima iz područja književnosti čime će se pružiti mogućnost opisa književnosti bilo kojeg naroda. Budući kako ontologija osim književnih pojmova obuhvaća i opis znanja vezan uz konkretne osobe tj. u ovom slučaju književnike, sadržat će pojmove i odnose koji nisu usko vezani uz književnost (npr. lokalitete, ustanove, zanimanja, životopis itd.).

Opseg ontologije će se skicirati popisom tzv. kompetencijskih tj. mjerodavnih pitanja (engl. *competency questions*)³⁶⁷ na koja će baza znanja temeljena na razvijenoj ontologiji biti u mogućnosti odgovoriti s tim da ona predstavljaju samo skicu i ne moraju biti iscrpna. Prema znanju koje će razvijena ontologija opisivati moguća su sljedeća kompetencijska pitanja:

- Koji su književnici rođeni u određenom gradu (npr. Splitu, Zagrebu itd.)?
- Tko je preveo djelo određenog književnika?
- Kojeg je djela preradba određeno književno djelo?
- Koji su književnici studirali na određenom fakultetu?
- Koji su književnici pisali novele?
- Koji su književnici pisali prozu?
- Koja su književna djela određenog naroda romani (drame, romani, novele itd.)?

Dakle, ontologija će omogućiti uvid u detaljne informacije o književnim djelima i književnicima određenog naroda. Prema navedenom primjeru pitanja na popisu ontologija će

³⁶⁷ Gruninger, Michael; Fox, Mark S., loc cit. (bilj. 331).

uključivati informacije o životopisu književnika, produkcijskoj povezanosti pojedinih književnika, odnose između pojedinih književnih djela, informacije s aspekta teorije i povijesti književnosti o pojedinim književnicima i djelima, kao i informacije o cjelokupnoj književnoj produkciji određenog naroda.

7.3.2 Upotreba postojeće ontologije

Ne postoje ontologije koje pokrivaju predmetnu domenu. Određeni pojmovi iz sličnih domena se mogu preuzeti, no njihovo preuzimanje ne može biti doslovno. DBpedia sadrži određene književne pojmove, ali to nije ontologija koja bi mogla opisati znanje književnosti određenog naroda. Kao polazna ontologija višeg reda odabrana je PROTON (PROTo ONtology) ontologija gornje razine koja služi kao modelirajuća osnova za brojne zadatke u različitim domenama. Sadrži oko 300 klasa i 100 svojstava, osiguravajući pokrivenost općih pojmova. Principi dizajniranja ontologije su: (1) domenski neovisna, (2) lagane logičke definicije, (3) poravnata je i kompatibilna s popularnim standardima i ontologijama, (4) dobra pokrivenost imenovanih entiteta i konkretnih domena (npr. ljudi, organizacije, lokacije, brojevi, datumi, adrese). Sve su to razlozi da je za polaznu ontologiju višeg reda odabrana PROTON ontologija. Stoga je prvi korak u izgradnji ontologije iz enciklopedički organiziranog znanja određivanje najopćenitijih ontologijskih pojmova, koji su klasificirani u skladu s podijelom proizašlom iz najviše razine temeljne PROTON ontologije. Preuzeti pojmovi iz najviše razine temeljne PROTON ontologije su slijedeći: temeljni pojmom Entity, a isti se dalje dijeli na hijerarhijske podpojmove tj. ontologijske klase *Object*, *Abstract*, *Happening*. Podklase za klasu *Object* preuzete iz PROTON ontologije su: *Agent (Group, Person)*, *Location*, *PieceOfArt*, *Statement (InformationResource)*. Podklase za klasu *Abstract* preuzete iz PROTON ontologije su: *SocialAbstraction (ResearchArea, Science)*. Podklase za klasu *Happening* preuzete iz PROTON ontologije su: *TimeInterval (CalendarYear)*, *Situation*, *Event*.

7.3.3 Označavanje važnih termina u ontologiji

U ovom postupku sastavlja se popis svih termina za koje se želi načiniti tvrdnje ili objasniti korisniku. Koji su to termini za koje se želi načiniti tvrdnje? Koja svojstva ovi termini imaju? Što se želi kazati o ovim terminima? Dakle, bitno je dobiti opsežan popis

termina ne vodeći računa o međusobnom preklapanju pojmova koje predstavljaju, odnose između termina, opisne odnose koje pojmovi mogu imati te utvrditi jesu li pojmovi klase ili svojstva.

Označavanje važnih termina je u zavisnoj vezi s postavljanim dosegom (granicom) domene. Kako bi se navedeno ostvarilo, provedeno je izgrađivanje rječnika termina na biografskim enciklopedijskim člancima Hrvatske enciklopedije iz područja hrvatska književnost. Ovim postupkom je utvrđen skup termina koji će biti uključeni u ontologiju tj. skup termina uključuje sve relevantne termine domene (pojmove, instance, attribute koji predstavljaju svojstva pojmova, odnose između pojmova itd.), njihov opis prirodnim jezikom te njihove sinonime itd. Tablica 9 ilustrira dio rječnika termina ontologije književnosti. Važno je napomenuti kako početni korak u stvaranju rječnika termina može sadržati nekoliko termina koji se odnose na isti pojam, što je zapravo utvrđivanje pojavljivanja sinonima.

Tabela 9: Prikaz dijela rječnika termina ontologije književnosti

Natuknica ili dio članka	Termini	Sinonim	Opis	Vrsta
Gundulić, Ivan	Gundulić, Ivan	Dživo, Mačica	Ime hrvatskog književnika	Individua
Matoš, Antun Gustav	Matoš, Antun Gustav		Ime hrvatskog književnika	Individua
hrvatski	Hrvat (osoba, nacionalnost)		Podatak o nacionalnoj pripadnosti	Individua: Hrvat, Relacija: jePoNarodnosti
pjesnik	pjesnik		Stručna djelatnost	pojam
dramatičar	dramatičar	Dramatik, Dramski pisac	Stručna djelatnost pisac kazališnih komada, drama	pojam
književnik	književnik	Pisac Spisatelj Spisateljica	Stručna djelatnost	pojam
Dubrovnik, 9. I. 1589	Dubrovnik (osoba, mjesto)		Mjesto i datum rođenja	Individua: Dubrovnik, Relacija: jeMjestoRođenja, jeGodinaRođenja atribut: datum rođenja
	9. I. 1589 (osoba, datum)			
Dubrovnik, 8. XII. 1638	Dubrovnik (osoba, mjesto)		Mjesto i datum smrti	Individua: Dubrovnik, Relacija: jeMjestoSmrti, jeGodinaSmrti, atribut: datum smrti
	8. XII. 1638 (osoba, datum)			
Tovarnik, 13. VI. 1873	Tovarnik (osoba, mjesto)		Mjesto i datum rođenja	Individua: Tovarnik, Relacija: jeMjestoRođenja, jeGodinaRođenja, atribut: datum rođenja
	13. VI. 1873 (osoba, datum)			
Zagreb, 17. III. 1914	Zagreb (osoba, mjesto)		Mjesto i datum smrti	Individua: Zagreb, Relacija: jeMjestoRođenja, jeGodinaRođenja, atribut: datum smrti
	17. III. 1914 (osoba, datum)			

Školovao se u Dubrovniku	Dubrovnik (osoba, mjesto)		Mjesto u kojem se osoba školovala	Individua: Dubrovnik , atribut: mjesto školovanja
U Zagrebu, gdje mu se obitelj naselila ubrzo nakon njegova rođenja	Zagreb (osoba, mjesto)		Prebivalište (zemlja, država/ pokrajina, grad itd. u kojem osoba pribiva ili je pribivala)	Individua: Zagreb, Relacija: prebivalište
God. 1608 postao je članom Velikoga vijeća	1608. (godina članstva)		Pripadnost (skupina kojoj osoba pripada ili je pripadala kroz zaposlenje, članstvo, kulturni identitet itd.)*, godina članstva	Individua: 1608 , Relacija: godina učlanjivanja
	Veliko vijeće (osoba, ustanova/tijelo i sl.)...			
pripovijetka	pripovijetka	Kratka priča Pripovijest Priča		Pojam
Na Sveučilištu u Zagrebu studirao prirodne znanosti, u Beču i Italiji romanistiku i slavistiku. U Beču je i diplomirao 1903. Od 1920. u Zagrebu predavao na Glumačkoj školi i pisao kazališne kritike (Novosti, Večer). Od 1927. ravnatelj Drame HNK u Zagrebu...	Sveučilište u Zagrebu (osoba, obrazovna ustanova), (osoba, mjesto studiranja)		Životopis, Prebivalište (zemlja, država/ pokrajina, grad itd. u kojem osoba pribiva ili je pribivala). Školovanje (u kojem mjestu se školovao?, Koju školu je osoba završila? Koji fakultet je osoba završil	Individua: Sveučilište u Zagrebu, Beč, Italija, romanistika, slavistika, 1903, 1920, Zagreb, Glumačka škola , Novosti Večer Relacije: jeStudiraoNaUstanovi, mjesto studiranja, jePodručjeStudiranja, godina diplomiranja, vrijeme rada, mjesto rada, jeRadio, jePisaoKnjiževneVrste, jeObjavljivaoUČasopisu; Pojam: prirodne znanosti, kazališna kritika
	prirodne znanosti (osoba, područje studiranja)			
	Beču (osoba, mjesto studiranja)			
	Italija (osoba, mjesto studiranja)			
	romanistika (osoba, područje studiranja)			
	slavistika (osoba, područje studiranja)			
	1903 (osoba, godina diplomiranja)			
	1920. (osoba, godina rada u ustanovi)			
	Zagreb (osoba, mjesto rada)			
	Glumačka škola (osoba, ustanova rada)			
	kazališna kritika (osoba , vrsta knjiž. djela)			
	kazališna kritika (pojam)			
	Novosti (individua), (osoba, serijska			

	<i>publikacija</i>			
	<i>Večer</i> (individua), (osoba, serijska publikacija)			
	1927. (godina rada u ustanovi)			Individua
	ravnatelj (osoba, ustanova)			Relacija: jeRadio, jeRavnatelj
	Drama HNK u Zagrebu			Individua
Begovićeve dramatičarske odlike najizrazitije su u dramama <i>Pustolov pred vratima</i> (1926) i <i>Bez trećega</i> (1931).	drama (osoba, književni rod/ vrsta)		jedan od triju glavnih književnih rodova (uz liriku i epiku), književno djelo koje se odlikuje ozbiljnošću zapleta i dubinom proživljavanja, <i>opr.</i> komedija	Pojam: drama, Relacija: jePisaoKnjiževneVrste, Individua: Pustolov pred vratima, Bez trećega,
	Pustolov pred vratima, Bez trećega		Vrsta djela koje je književnik pisao	Atribut: naslov djela
	1926 1931		Ime djela Godina kada je objavljeno djelo	Atribut: godina objavljivanja

Sljedeća dva postupka su razvoj klasne hijerarhije i definiranje odnosa. Ova dva postupka su usko isprepletena te je teško sasvim odvojeno učiniti jedan od njih. Uobičajeno se načini nekoliko definicija pojmova u hijerarhiji i zatim se nastavi s opisivanjem svojstava ovih pojmova itd. Ova dva postupka su najvažniji u procesu stvaranja ontologije, a njihov opis slijedi u nastavku.

7.3.4 Definiranje ontologijskih klasa i klasne hijerarhije (pojmovne taksonomije)

Nakon definiranja rječnika termina, izgrađuje se pojmovna taksonomija, odnosno ontologijske klase i klasna hijerarhija. Provodi se odabir termina iz rječnika koji će

predstavljati pojmove u pojmovnoj taksonomiji. Veoma je važno u pojmovnoj taksonomiji utvrditi skupove razdvojenih (engl. disjoint) pojmova tj. onih koji nemaju zajedničke instance.

Primijenio se Uschold i Gruninger³⁶⁸ pristup za razvoj klasne hijerarhije tj. metodologija razvoja ontologije književnosti primjenjuje pristup od *vrha prema dnu* (engl. top-down) koja počinje s definicijom najopćenitijih pojmova u domeni, a zatim slijedi specijalizacija pojmova. Na primjer, započelo se s izradom klasa za opće pojmove *Književnik* i *Književnost*. Zatim se klasa *Književnik* može specijalizirati stvaranjem nekih njezinih podpojмова, npr. *Publicist*, *Pjesnik*, *Prozaik*. Dalje se *Publicist* može kategorizirati kao *Feljtonist*, *Kritičar* itd.; *Pjesnik* se može kategorizirati kao *Epik*, *Prozaici* se kategoriziraju kao *Romanospisac*, *Pripovjedač* i sl.

Dakle, razvoj ontologije započinje definiranjem ontologijskih klasa iz popisa rječnika termina ontologije književnosti načinjenom u prethodnom koraku. Odabiru su termini koji opisuju objekte s neovisnom postojanošću umjesto termina koji opisuju te objekte. Ovi termini će biti ontologijske klase i postat će sigurno uporište u klasnoj hijerarhiji prilikom organizacije u pojmovnu taksonomiju. Prilikom izrade pojmovne taksonomije za ontologiju književnosti vodilo se načelom: „*Ako je klasa A nadklasa klase B, onda je svaka instanca od B također instanca od A*“. Drugim riječima, klasa B predstavlja pojam koji je „vrsta od“ A“. Primjerice, svaki Pjesnik je obavezno Književnik tj. pjesnik je „vrsta od“ književnik. Stoga je klasa Pjesnik podklasa od klase Književnik.

Iz navedenog je vidljivo kako su klase povezane na način da neke predstavljaju neki pojam općenitije, a neke specifičnije te ih se na temelju toga smješta u hijerarhiju. Prilikom razvoja hijerarhije potrebno je zadovoljiti sljedeće zahtjeve³⁶⁹:

- hijerarhija treba biti korektna, tj. svaka instanca podklase B je ujedno i instanca njene nadklase A;
- tranzitivnost hijerarhije, tj. ako je B podklasa od A, a C podklasa od B, onda C također mora biti podklasa od A;
- ažuriranje eventualnih promjena u hijerarhiji tijekom vremena;
- imati jasnu razliku između klase i njenog imena, tj. ukoliko za neko ime postoje sinonimi, oni ne smiju biti imena drugih klasa;
- izbjegavanje kružnih ciklusa, tj. ukoliko je B podklasa od A, onda A ne smije biti podklasa od B.

³⁶⁸ Uschold, M.; Gruninger, M., loc cit. (bilj. 336).

³⁶⁹ Noy, Natalya F.; McGuinness, Deborah L., loc cit. (bilj. 36).

Držeći se prethodno navedenih principa pristupilo se razvoju pojmovne taksonomije ontologije književnosti čiji prikaz slijedi u nastavku. Prvi postupak u izgradnji ontologije je određivanje najopćenitijih ontologijskih pojmova, koji su klasificirani u skladu s podjelom proizašlom iz najviše razine temeljne PROTON ontologije koja započinje s temeljnim pojmom *Entity*, a isti se dalje dijeli na hijerarhijske podpojmove tj. ontologijske klase *Object*, *Abstract*, *Happening*. Podklase za klasu *Object* preuzete iz PROTON ontologije su: *Agent (Group, Person)*, *Location*, *PieceOfArt*, *Statement (InformationResource)*. Podklase za klasu *Abstract* preuzete iz PROTON ontologije su: *SocialAbstraction (ResearchArea, Science)*. Podklase za klasu *Happening* preuzete iz PROTON ontologije su: *TimeInterval (CalendarYear)*, *Situation*, *Event*. Dakle u stvaranju ontologije hrvatske književnosti se krenulo s temeljnim klasama PROTON ontologije najviše razine koja je poslužila kao temelj prilikom razvoja ontologije za opis znanja pohranjenog u člancima Hrvatske enciklopedije. Temeljna taksonomija pojmova ontologije hrvatske književnosti je zbog preglednosti u nastavku rada prikazana kroz tzv. ontologijske module, a zbog brojnosti podpojmove i potrebne preglednosti pojedini moduli će se prikazati slikom iz Protege alata ili tablicama ukoliko se radi o većoj taksonomiji pojmova.

7.3.4.1 Agent modul

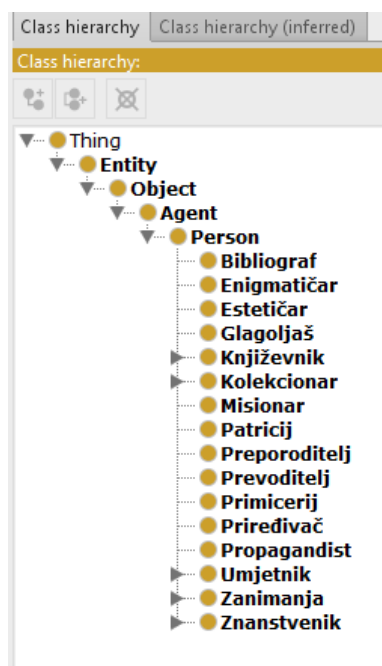
Modul *Agent* razvrstava pojmove o zanimanjima, profesiji i poslovima (književnik, arhitekt), ekstrahirane iz enciklopedijskih članaka Hrvatske enciklopedije, u taksonomiju pojmova. Ovim modulom su predstavljeni pojmovi zanimanja i profesija kojima se osoba bavila i obavljala tijekom života, a njegov cilj je omogućiti razvrstavanje književnika iz enciklopedijskih članaka Hrvatske enciklopedije prema profesiji, zanimanjima i sl. kako bi se npr. znalo koji su hrvatski književnici ujedno pravnici, arhitekti, lingvisti, slikari i sl.

Ontologijski *Agent* modul uključuje *foaf: Agent* klasu. Podklasa od *foaf: Agent* je *foaf: Person*, koja predstavlja osobe. Ostale vrste agenata uključuju *foaf: Organization* i *foaf: Group*. U PROTON ontologiji *protont: Object* je nadklasa od *protont: Agent*, *protont: Person*, *protont: Group*, *protont: Organization*, *protont: Location*, *protont: Statement*, *protont: InformationResource*, *protont: Document*, *protont: Product* i *protont: Service*. Ontologijski *Agent* modul uključuje *foaf: Agent* kao primarnu nadklasu te uvodi *protont: Person* i *protont: Group* kao podklase, kao što je prikazano na Slici 28.



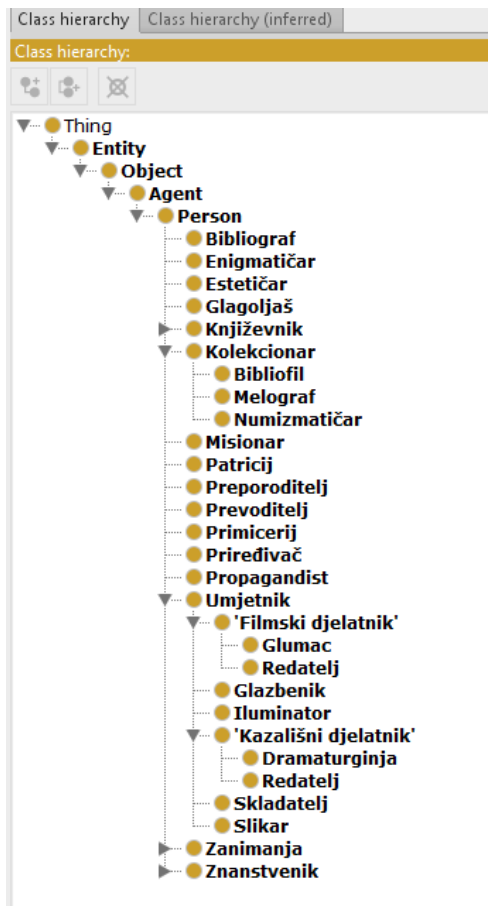
Slika 28: Agent modul

Nadalje, *protont*: *Person* je nadklasa definiranim podklasama prikazanim na slikama 29, 30 i 31. Navedene klase su dobivene iz biografskih enciklopedijskih članak Hrvatske enciklopedije. S obzirom na vrstu entiteta načinjene su podklase koje idu direktno pod klasu *Person* te one koje se mogu grupirati kao podklase određene nadklase, kao što su *Umjetnik*, *Zanimanja* i *Znanstvenik*.



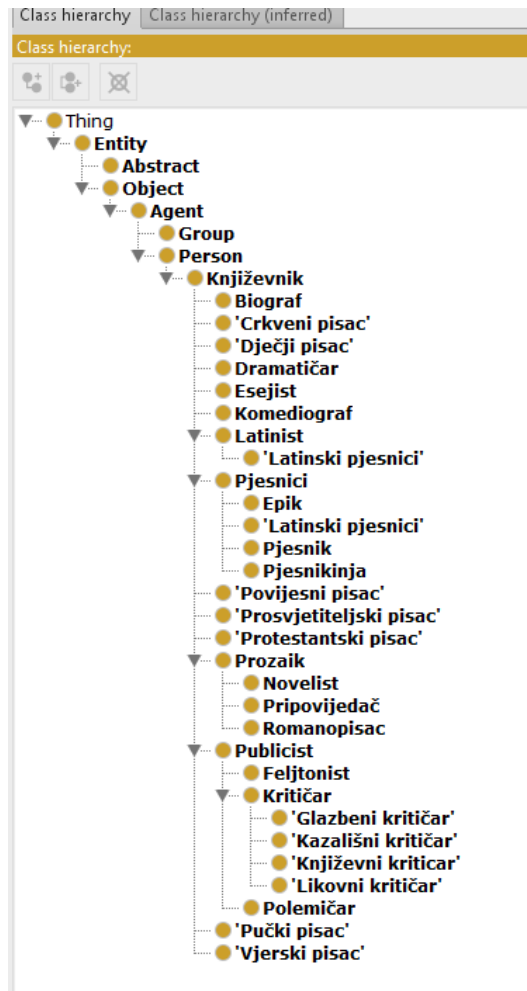
Slika 29: Podklase klase Person

Podklase klase *Umjetnik*, *Zanimanja* i *Znanstvenik* su prikazane na Slici 30, a sami njihovi nazivi govore kakvu vrstu pojmova svaka od tih nadklasa okuplja u svoje podklase.



Slika 30: Podklase klasa Umjetnik, Zanimanja i Znanstvenik

Budući je ontologijska domena hrvatska književnost, na Slici 31 će se prikazati sve podklase ontologijske klase *Književniki*. Ovaj dio pojmovne taksonomije ima za cilj razvrstavanje književnika iz enciklopedijskih članaka Hrvatske enciklopedije na pjesnike, esejiste, dramatičare, publiciste i sl.



Slika 31: Podklase ontologijske klase Književnici

7.3.4.2 Književno djelo modul

Modul Književno djelo uključuje klasu *protont: PieceOfArt* (podklasa *protont: Object*), koja kao podklasu sadrži *Književno djelo*, a ova pak okuplja sva književna djela hrvatske književnosti.

7.3.4.3 Književni rodovi i vrste modul

Modul *Književni rodovi i vrste* razvrstava pojmove o književnim rodovima i vrstama, ekstrahirane iz enciklopedijskih članaka Hrvatske enciklopedije, u hijerarhiju ontologije književnosti. Ovim modulom su predstavljeni pojmovi teorije književnosti, a njegov cilj je omogućiti razvrstavanje književnih djela iz enciklopedijskih članaka Hrvatske enciklopedije prema književnim rodovima i vrstama. Ontologijski modul *Književni rodovi i vrste* uključuje *protont*: *Science* klasu kao primarnu nadklasu te uvodi klasu *Znanost o književnosti*, koja ima podklasu *Teorija književnosti*, a ova je pak nadklasa od *Književni rodovi i vrste*.

Nadalje, klasa *Književni rodovi i vrste* sadrži podklase: *Didaktičke vrste*, *Diskurzivni književni oblici*, *Drama*, *Epika*, *Lirika*, *Poezija*, *Proza*. Navedene podklase sadrže također podklase, koje su navedene u Tablici 10.

Tabela 10: Hijerarhijski prikaz modula Književne vrste i rodovi

KLASA	PODKLASE
Didaktičke vrste	(1) Alegorija (1) Basna (1) Egzemplum (1) Epigram (2) 'Satirični epigrami' (1) Parabola (1) Pitalica (1) Satira
Diskurzivni književni oblici	(1) Biografija (2) Autobiografija (2) Hagiografija (1) Dnevnik (1) Esej (1) Govor (1) Kozerije (1) Memoari (1) Napisi (1) Pismo (2) Epistola (nadklasa: Lirika) (3) 'Pjesnička poslanica' (3) Apologija (1) Publicistika (2) Feljton (2) Kritika (3) 'Filmska kritika' (3) 'Glazbena kritika' (3) 'Kazališna kritika' (3) 'Likovna kritika' (3) Prikazi (4) Osvrt (4) Pregled (2) Polemika (2) Reportaža

	<ul style="list-style-type: none"> (3)Novinarska reportaža (2)Članak <ul style="list-style-type: none"> (3)Kazališni članci (3)Književni članci (3)Stručno pedagoški članci (3)Tekstološki radovi (1)Putopisi (1)Rasprava
Drama	<ul style="list-style-type: none"> (1) Crkvena prikazanja <ul style="list-style-type: none"> (2) Auto (2) Mirakul (2) Misterij (2) Mortalitet (2) Pasion (1) Drama apsurda (1) Drama u užem smislu <ul style="list-style-type: none"> (2) Hagiografska drama (2) Povijesne drame (2) Satirične drame (1) Epsko kazalište (1) Građanska drama (1) Povijesna drama <ul style="list-style-type: none"> (2) Dramska kronika (1) [Drama prema_mediju] <ul style="list-style-type: none"> (2) Televizijske drame (2) Radio-drama (2) Scenarij <ul style="list-style-type: none"> (3) Filmski scenarij (3) Televizijski scenarij (1) Burleska <ul style="list-style-type: none"> (2) Parodija (2) Travestija (1) Dramolet (1) Farsa (1) Igrokaz (1) Komedija <ul style="list-style-type: none"> (2) Eruditna komedija (2) Improvizirana komedija (2) Komedija intrige (2) Komedija karaktera (2) Komedija konverzacije (2) Komedija naravi (2) Komedija situacije (2) Ulična komedija (2) Lakrdija (1) Libreto (1) Maskareta (1) Melodrama (1) Mim (1) Pastoral (1)[Prema_organizaciji_jezicne_gradje] <ul style="list-style-type: none"> (2) Drama u stihu (1) Tragedija <ul style="list-style-type: none"> (2) Povijesna tragedija (1) Tragikomedija (1) Vodvilj
Epika	<ul style="list-style-type: none"> (1) Epska poezija <ul style="list-style-type: none"> (2) Epska pjesma (2) Balada

	<ul style="list-style-type: none"> (2) Bugarštica (2) Romanca (2) Spjev <ul style="list-style-type: none"> (3) Dramski spjev (3) Hagiografski spjev (3) Lirsko epski spjev (3) Prigodni spjev (3) Ep <ul style="list-style-type: none"> (4) Didaktički ep <ul style="list-style-type: none"> (4) Idilični ep (4) Komični ep (4) Narodni ep (4) Povijesni ep (4) Religiozni ep (4) Romantični ep (4) Životinjski ep (4) Epilij (5) Hagiografski epilij
<hr/> <p>Proza</p>	<ul style="list-style-type: none"> (3) Poema <ul style="list-style-type: none"> (4) Epopeja (4) Alegorijska poema <hr/> <ul style="list-style-type: none"> (1) Proza <ul style="list-style-type: none"> (2) Fikcijska proza (2) Jednostavni oblici <ul style="list-style-type: none"> (3) Anegdota (3) Bajka (3) Kronika <ul style="list-style-type: none"> (4) Stihovana kronika (3) Legenda <ul style="list-style-type: none"> (4) Stihovana legenda (3) Mit (3) Poslovica (3) Saga (3) Vic (3) Zagonetka (2) Proza u trapericama (2) Složeni oblici <ul style="list-style-type: none"> (3) Novela (3) Pripovijetka <ul style="list-style-type: none"> (4) Autobiografska pripovijest (4) Pjesnička pripovijest (4) Pripovijetka za djecu (4) Stihovana pripovijest (3) Roman <ul style="list-style-type: none"> (4) Dječji roman (4) Lirski roman (4) Panoramski roman (4) Roman esej (4) Roman struje svijesti (4) Roman za mladež (4) [proza prema stupnju integracije predliterarnih tekstova] <ul style="list-style-type: none"> (5) Biografski roman <ul style="list-style-type: none"> (5) Epistolarni roman (5) Esejistički roman (5) Roman-dnevnik

	<p>(5) Roman-feljton</p> <p>(4)[proza prema čimbenicima integracije elemenata]</p> <p>(5) Roman lika</p> <p>(5) Roman prostora</p> <p>(5) Roman vremena</p> <p>(5) Roman zbivanja</p> <p>(4) [proza prema izgradnji sižea]</p> <p>(5) Lančani roman</p> <p>(5) Paralelni roman</p> <p>(5) Prstenasti roman</p> <p>(4)[proza prema odnosu pripovjedača]</p> <p>(5) Autorski roman</p> <p>(5) Ja roman</p> <p>(5) Personalni roman'</p> <p>(4)[proza prema stavu autora]</p> <p>(5) Didaktički roman</p> <p>(5) Humoristični roman</p> <p>(5) Satirički roman</p> <p>(5) Sentimentalni roman</p> <p>(5) Tendenciozni roman</p> <p>(4) Tematska_klasifikacija</p> <p>(5) Društveni roman</p> <p>(5) Gotski roman</p> <p>(5) Kriminalističko humoristični roman</p> <p>(5) Ljubavni roman</p> <p>(5) Obiteljski roman</p> <p>(5) Odgojno-obrazovni roman</p> <p>(5) Pastoralni roman</p> <p>(5) Pikarski roman</p> <p>(5) Političko povijesni roman</p> <p>(5) Povijesni roman</p> <p>(5) Psihološki roman</p> <p>(5) Pustolovni roman</p> <p>(5) Viteški roman</p> <p>(5) Western roman</p> <p>(5) Znanstvenofantastični roman</p> <p>(5) Špijunski roman</p> <p>(5) Kriminalistički roman</p> <p>(2)[proza prema odnosu prema zbilji]</p> <p>(3) Dokumentarna proza</p> <p>(3) Poetska proza</p> <p>(2) Crtica</p> <p>(3) Putopisne crtice</p>
Lirika	<p>(1) Anacreontska pjesma</p> <p>(1) Korska lirika</p> <p>(1) Nabožne pjesme</p> <p>(1) Narodna pjesma</p> <p>(1) Pokladna pjesma</p> <p>(1) Refleksivna lirika</p> <p>(1) Ballata</p>

	<ul style="list-style-type: none"> (1) Bukolika (1) Ditiramb (1) Dramski monolog (1) Ekloga <ul style="list-style-type: none"> (2) Elegija <ul style="list-style-type: none"> (3) Heroide (2) Epicedij (2) Epistola <ul style="list-style-type: none"> (3) Pjesnička poslanica (3) Apologija (2) Epitaf (2) Epitalam (2) Espinela (2) Gazela (2) Glosa (2) Haiku (2) Himna (2) Idila (2) Kancona (2) Kantata (2) Kasida (2) Koleda (2) Madrigal (2) Oda (2) Popijevka (2)[proza prema tematici] <ul style="list-style-type: none"> (3) Dijektalna lirski pjesma (3) Ljubavna lirika (3) Misaone pjesme (3) Pejzažne pjesme (3) Pjesma u prozi (3) Religiozne pjesme (3) Rodoljubna lirika (3) Satirične pjesme (3) Socijalne pjesme (3) Šaljive pjesme (3) Bukolike (3) Rodoljubna_pjesma (2) Prepjev (2) Prigodnice <ul style="list-style-type: none"> (3) Pirna pjesma (3) Religiozne prigodnice (3) Satirične prigodnice (2) Psalam (2) Raspodija (2) Rispetto (2) Rondel (2) Rondo (2) Sestina (2) Sonet <ul style="list-style-type: none"> (3) Shakespearov sonet (3) Petrarkin_sonet (2) Strambotto (2) Tanka (2) Triolet (2) Tužaljka
Poezija	<ul style="list-style-type: none"> (1) Bukolička poezija (1) Prigodna poezija

7.3.4.4 Modul Periodizacija književnosti

Modul *Periodizacija književnosti* enciklopedijske članke Hrvatske enciklopedije tj. hrvatske književnike i književna djela razvrstava u pojedina razdoblja i pravce, književne pokrete i generacije te dodjeljuje književnim djelima stilsko obilježje pojedinog razdoblja. Ontologijski modul *Periodizacija književnosti* uključuje *protont*: *Science* klasu kao primarnu nadklasnu te uvodi njezinu podklasnu *Znanost o književnosti*, koja ima podklasnu *Povijest književnosti*, a ova je pak nadklasa od klase *Periodizacija književnosti*. Podklase *Periodizacija književnosti* su prikazane u Tablici 11.

Tabela 11: Periodizacija književnosti

Periodizacija književnosti
(1) Barokna književnost
(2)Eufizam
(2)Gongorizam
(2)Manirizam
(2)Precioznost
(2)Protureformacija
(2)Reformacija
(1) Književnost antike
(2)Grčka književnost
(3)Arhajsko razdoblje
(3)Helenizam
(3)Klasično doba
(3)Rimsko doba
(2)Rimska književnost
(3)Arhajsko doba
(3)Kasno carstvo
(3)Srebrno razdoblje
(3)Uvodno rimsko razdoblje
(3)Zlatno doba
(1) Književnost klasicizama
(2)Rokoko
(1) Književnost modernizama
(2) Avangarda
(3)Dadaizam
(3)Ekspresionizam
(3)Futurizam
(3)Imagizam
(3)Konstruktivizam
(3)Nadrealizam
(3)Zenitizam
(2) Esteticizam
(3) Simbolizam
(4)Neoromantizam
(2)Impresionizam
(2) Kasni modernizam
(3) Druga moderna
(4) Književnici u izbjeglištvu
(4) Krugovaši

	(4) Razlogovci
	(4) Stilski pluralizam
	(3) Egzistencijalizam
	(3) Književnost od 1929. do 1952.
	(4) Moderni objektivizam
	(4) Socijalistički realizam
(2) Moderna	
	(3) Ekspresionizam
	(3) Prva moderna
	(4) Secesija
	(4) Stilski pluralizam
(2) Postmodernizam	
(1) Književnost prosvjetiteljstva	
	(2) Predromantizam
	(3) Sentimentalizam
	(3) Sturm und Drang
(1) Književnost realizma	
	(2) Naturalizam
	(2) Parnasizam
	(2) Predrealizam
	(2) Verizam
(1) Književnost renesanse	
	(2) Elizabetansko doba
(1) Književnost romantizma	
	(2) Bidermajer
	(2) Ilirizam
(1) Književnost srednjovjekovlja	
	(2) Humanizam
	(2) Predrenesansa

7.3.4.5 Modul Dokument

Modul *Dokument* razvrstava književna djela iz enciklopedijskih članaka Hrvatske enciklopedije u vrste publikacija, dokumenata i članaka. Ekstrahirani su pojmovi koji se odnose na vrste publikacija i dokumenata iz enciklopedijskih članaka te su isti poredani u hijerarhiju modula *Dokument*.

Ontologijski modul *Dokument* uključuje *protont: Statement* klasu kao primarnu nadklasnu te uvodi njezinu podklasnu *protont:InformationResource*, koja ima podklasnu *Dokument*, a njezine podklase su prikazane u Tablici 12.

Tabela 12: Hijerarhijski prikaz modula Dokument

Dokument	
(1) Dokument dijelovi	(2) Kolumna (2) Prilozi (2) Rubrika
(1) Sabrana djela	(2) Antologije (3) Antologijske novele (2) Zbirke (3) Zbirka aforizama (3) Zbirka autobiografskih zapisa (3) Zbirka eseja (3) Zbirka novela (3) Zbirka pjesama (4) Zbirka crkvenih pjesama (4) Zbirka dječjih pjesama (4) Zbirka pjesama u prozi (4) Kanconijer (5) Ljubavni kanconijer (3) Zbirka pripovijedaka (3) Zbirka proza (3) Zbirka zapisa (3) Zbirka članaka (3) Zbirke anegdota (3) Zbirke crtica (3) Biblioteke
	(2) Web stranica (3) ExternalResource
(1) Bibliografija (1) Katalog (1) Manifest (1) Natpis	
	(2) Antički natpis
(1) Prijepisi (1) Publikacije	
	(2) Knjiga (3) Knjiga drama (3) Knjiga eseja (3) Knjiga feljtona (3) Knjiga govora (3) Knjiga književnokritičkih tekstova (3) Knjiga kritika (3) Knjiga makedonističkih studija (3) Knjiga novela (3) Knjiga ogleđa (3) Knjiga pjesama (3) Knjiga polemika (3) Knjiga pripovijetki (4) Knjiga pripovijetki za djecu (3) Knjiga proze (4) Knjiga kratkih proza (3) Knjiga putopisne proze (3) Knjiga rasprava (3) Knjiga studija (3) Knjiga zapisa (3) Knjige članaka (3) Poučne knjižice

	(3) Rijetke knjige	(4) Inkunabule
	(3) Biblija	(4) Biblija u stihu
	(3) Brošure	
	(3) Monografije	
	(3) Priručnik	(4) Teološki priručnici
		(5) Brevijari
		(5) Katekizam
		(5) Misali
		(5) Molitvenici
	(3) Udžbenici	(4) Gramatika
	(3) Zbornici	(4) Enciklopedijski zbornik
		(4) Pjesnički zbornici
	(2) Kodeks	
	(2) Referentna djela	
	(3) Enciklopedije	
	(3) Leksikon	(4) Biografski leksikon
	(3) Pojmovnik	
	(3) Rječnici	(4) Slikovni rječnik
		(4) Stručni rječnici
	(2) Serijske publikacije	
	(3) Dnevnik	
	(3) Glasilo	
	(3) Godišnjak	(4) Almanah
		(4) Kalendari
	(3) Listovi	
		(4) Novine
	(3) Tjednik	
	(3) Časopisi	(4) Časopisi za znanost o književnosti
		(4) Revije
(1) Rukopisi	(2) Autograf	
(1) Sastav		
(1) Spisi		
(1) Studija	(2) Informativna studija	
	(2) Traktakt	
(1) Trilogija		
(1) Zapis		

7.3.5 Određivanje ontologijskih odnosa

Same klase neće osigurati dovoljno informacija kako bi se odgovorilo na kompetencijska (mjerodavna) pitanja. Nakon definiranja klasa potrebno je opisati unutarnju strukturu pojmova. Ontologijski odnosi su preuzeti prilikom analize biografskih enciklopedijskih članaka Hrvatske enciklopedije te su ušli u terminološki rječnik ontologije književnosti. Nakon izbora klasa s popisa termina, većina preostalih termina će biti svojstva ovih klasa tj. činit će ontologijske odnose u ontologiji književnosti. Ovi termini uključuju primjerice *mjesto (rođenja, smrti)*, *datum (rođenja, smrti)*, *godina (rođenja, smrti, izdanja)*, *književna vrsta*, *prepjev*, *autor*, *urednik* itd. Oblik navedenih svojstava je preinačen, stoga se u konačnici od prethodno navedenih javljaju npr. sljedeća svojstva: *mjesto rođenja*, *mjesto smrti*, *datum rođenja*, *datum smrti*, *godina rođenja*, *godina smrti*, *godina izdanja*, *jeKnjiževnaVrsta*, *jePrepjev*, *jeAutor*, *jeUrednik*.

Potrebno je odrediti koju klasu opisuje svako svojstvo na popisu te će tako primjerice *Književnik* imati sljedeća svojstva koja će mu omogućiti uspostavu određene vrste odnosa: *mjesto rođenja*, *mjesto smrti*, *datum rođenja*, *datum smrti*, *godina rođenja*, *godina smrti*. Klasa *Književno djelo* ima npr. svojstva *godina izdanja*, *jeKnjiževnaVrsta*, *jePrepjev*. Klasa *Književnik* ima npr. svojstva *jeAutor*, *jeUrednik*. Prilikom izrade ontologije književnosti su prisutne tri vrste odnosa:

- opisni odnosi (engl. annotation property) kojima se opisuje npr. ime klase ili individue (label), stavlja se komentar za određenu klasu ili individuu, dodaje se alternativno ime ili sinonim ukoliko ga klasa ili individua ima (altLabel) itd.,
- objektni odnosi (engl. object property) uspostavljaju odnos između dva pojma tj. individua koje pripadaju različitim klasama,
- podatkovni odnosi (engl. data property) uspostavljaju odnos između individue i vrste podatka

U nastavku slijedi detaljan prikaz svih vrsta ontologijskih svojstava.

7.3.5.1 Objektni odnosi

Objektni odnosi uspostavljaju odnos između dvije individue. Analizom članaka uspostavljeni su objektni odnosi koji se mogu podijeliti u dvije veće skupine: prvu skupinu

čine odnosi ekstrahirani iz samog teksta enciklopedijskih članka (*svojstva izdavača, svojstva autora, svojstva djela i svojstva osobe*), a drugu skupinu čine odnosi potrebni da bi ontologija mogla izvesti automatsko zaključivanje nad unesenim činjenicama (*imaOblik, imaZnačajke, jePoVeličini, seIzvodi*)

7.3.5.2 Svojstva osobe

Svojstva osobe opisuju osobne podatke o nekoj osobi uspostavljajući npr. odnos između osobe i pojedine ustanove (npr. ustanove na kojima se osoba školovala), odnose između dvije osobe, odnose između osobe i karakteristika njezinog književnog stvaralaštva. U nastavku slijedi njihov prikaz iz Protege alata na Slici 32, te objašnjenje svakog pojedinog odnosa. Svojstva osobe se dijele na podsvojstva *OsobaOsobniPodaci, OsobaDjelatnost, OsobaOsoba, OsobaStvaralaštvo*. Svako od ovih svojstava sadrži podsvojstva kojima se opisuje ono na što navedeno nadsvojstvo upućuje (npr. o poslovima vezanim uz pojedine osobe). Tako, primjerice svojstvo *OsobaOsobniPodaci* sadrži podsvojstva koja opisuju osnovne podatke o osobi (*jeMjestoRođenja, jeStudirao* itd.).



Slika 32: Prikaz hijerarhije odnosa Svojstva Osoba i OsobaOsobniPodaci

label: jeMjestoRođenja

comment: Relacija povezuje književnika s mjestom rođenja.

D: Person,

R: Mjesto
jeRodniGrad inverse jeMjestoRođenja
SubProperty Of: OsobaOsobniPodaci

label: jeRodniGrad
comment: Relacija ukazuje u kojem je gradu određeni književnik rođen.
D: Mjesto
R: Person
jeMjestoRođenja inverse jeRodniGrad
SubProperty Of: OsobaOsobniPodaci

label: jeMjestoSmrti
comment: Relacija povezuje književnika s mjestom smrti.
D: Person
R: Mjesto
SubProperty Of: OsobaOsobniPodaci

label: jePoNarodnosti
comment: Relacija ukazuje na nacionalnost pojedinog književnika.
D: Person
R: Nacionalnost
SubProperty Of: OsobaOsobniPodaci

label: jePodručjeStudiranja
comment: Relacija definira koji su sve književnici studirali npr. povijest.
D: Područje
R: Person
jeStudirao inverse jePodručjeStudiranja
SubProperty Of: OsobaOsobniPodaci

label: jeStudirao
comment: Relacija kaže koje je područje studirao određeni književnik
D: Person
R: Područje
jeStudirao inverse jePodrucjeStudiranja
SubProperty Of: OsobaOsobniPodaci

label: jeStudiraoNaUstanovi
comment: Relacija ukazuje na kojoj je ustanovi (fakultetu) studirao književnik.
D: Person
R: Obrazovne ustanove
jeUstanovaStudiranja inverse jeStudiraoNaUstanovi
SubProperty Of: OsobaOsobniPodaci

label: jeUstanovaStudiranja

comment: Relacija povezuje književnika s ustanovom na kojoj je studirao te ovaj odnos ukazuje koji su sve književnici studirali na određenoj ustanovi (npr. Filozofski fakultet u Zagrebu).

D: Obrazovne ustanove

R: Person

jeStudiraoNaUstanovi inverse jeUstanovaStudiranja

SubProperty Of: OsobaOsobniPodaci

label: jeStipendist

comment: Relacija ukazuje na stipendiju koju je određena osoba primala.

D: Person

R: Stipendija

SubProperty Of: OsobaOsobniPodaci

Podsvojstvo *OsobaDjelatnost* sadrži podsvojstva prikazana na Slici 33 te objašnjena u nastavku. Opisuju poslove i djelatnosti za koje su se pojedini književnici obrazovali ili koje su obavljali tijekom svog djelovanja.



Slika 33: Prikaz hijerarhije ontologijskih odnosa OsobaDjelatnost

label: jeRadio

comment: Relacija definira ustanove, poduzeća, organizacije itd. za koje je osoba radila.

D: Person
R: Organization
Subproperty Of: OsobaDjelatnost

label: jeSviraoU
comment: Relacija definira ustanovu, udruženje i sl. u kojoj je osoba svirala.
D: Person
R: Facility
Subproperty Of: OsobaDjelatnost

label: jeDopisnik
comment: Relacija označava za koje novine, radio postaje itd. je određena osoba slala vijesti i komentare iz nekog mjesta izvan sjedišta glasila ili izvan inozemstva.
D: Person
R: Serijske publikacije, Radio postaje
Subproperty Of: OsobaDjelatnost

label: jeDramaturg
comment: Relacija definira u kojem je kazalištu određena osoba djelovala kao dramaturg
D: Person
R: Kazalište
Subproperty Of: OsobaDjelatnost

label: jeFeljtonist
comment: Relacija definira koje je serijske publikacije bio feljtonist određeni književnik.
D: Person
R: Serijske publikacije
Subproperty Of: OsobaDjelatnost

label: jeKolumnist
comment: Relacija definira koje je serijske publikacije bio kolumnist određeni književnik.
D: Person
R: Serijske publikacije
Subproperty Of: OsobaDjelatnost

label: jeKomentator
comment: Relacija definira koje je serijske publikacije ili organizacije bio komentator određeni književnik.
D: Person
R: Serijske publikacije, Organization
Subproperty Of: OsobaDjelatnost

label: jeKorektor
comment: Relacija definira koje je serijske publikacije bio korektor određeni književnik.

D: Person
R: Serijske publikacije
Subproperty Of: OsobaDjelatnost

label: jeKritičar
comment: Relacija definira koje je serijske publikacije bio kritičar određeni književnik.
D: Person
R: Serijske publikacije
Subproperty Of: OsobaDjelatnost

label: jeKazališniKritičar
comment: Relacija definira koje je serijske publikacije bio kazališni kritičar određeni književnik.
D: Person
R: Serijske publikacije
Superproperties: jeKritičar

label: jeKnjiževniKritičar
comment: Relacija definira koje je serijske publikacije bio književni kritičar određeni književnik.
D: Person
R: Serijske publikacije
Superproperties: jeKritičar

label: jeKustost
comment: Relacija određuje književnike koji su tijekom života bili kustosi u nekom muzeju.
D: Person
R: Muzej
Subproperty Of: OsobaDjelatnost

label: jeLektor
comment: Relacija definira kojeg je izdavača bio lektor određeni književnik.
D: Person
R: Izdavač
Subproperty Of: OsobaDjelatnost

label: jeNovinar
comment: Relacija definira koje je serijske publikacije ili organizacije bio novinar određeni književnik.
D: Person
R: Serijske publikacije, Organization
Subproperty Of: OsobaDjelatnost

label: jePisaoKulturološkeKomentare

comment: Relacija definira u kojoj je serijskoj publikaciji pisao kulturološke komentare određeni određeni književnik.

D: Person

R: Serijske publikacije

Subproperty Of: OsobaDjelatnost

label: jePotpredsjednik

comment: Relacija ukazuje da je određena osoba tj. književnik bio potpredsjednik određene organizacije tj. društva.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jePredavaoNaUstanovi

comment: Relacija definira ustanovu na kojoj je osoba predavala.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jeProfesorNaUstanovi

comment: Relacija definira u kojoj je obrazovnoj ustanovi osoba radila kao profesor.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jeIzvanredniProfesor

comment: Relacija definira u kojoj je obrazovnoj ustanovi osoba radila kao izvanredni profesor i iz kojeg je područja određena osoba bila izvanredni profesor.

D: Person

R: 'Obrazovne ustanove', Područje

Superproperties: jeProfesorNaUstanovi

label: jeRedovitiProfesor

comment: Relacija definira u kojoj je obrazovnoj ustanovi osoba radila kao redoviti profesor i iz kojeg je područja određena osoba bila redoviti profesor.

D: Person

R: 'Obrazovne ustanove', Područje

Superproperties: jeProfesorNaUstanovi

label: jePredsjednik

comment: Relacija ukazuje da je određena osoba tj. književnik bio predsjednik određene organizacije tj. društva.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jePrevoditelj

comment: Relacija definira za kojeg je izdavača radio kao prevoditelj određeni književnik.

D: Person

R: Izdavač

Subproperty Of: OsobaDjelatnost

label: jePriredivaoIzdanja

comment: Relacija definira koji je književnik priredivao izdanja određenog izdavača. D:

Person

D: Person

R: Izdavač

Subproperty Of: OsobaDjelatnost

label: jeSuradnikUstanove

comment: Relacija definira s kojom ustanovom je surađivao određeni književnik.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jeTajnik

comment: Relacija ukazuje na položaj tajnika određene osobe unutar neke organizacije.

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jeUrednikUstanove

comment: Relacija definira koji je književnik bio urednik određene ustanove, društva i sl. (npr. izdavača, radio-postaja...).

D: Person

R: Radio postaje, Društva, Izdavač, Organization

Subproperty Of: OsobaDjelatnost

label: jeUpravitelj

comment: Relacija ukazuje na to da je osoba obavljala funkciju upravljanja određenim organizacijom tijekom života.

altLabel: jeRavnatelj

altLabel: jeDirektor

D: Person

R: Organization

Subproperty Of: OsobaDjelatnost

label: jeČlan

comment: Relacija definira da je određena osoba bila član određene obrazovne ustanove i organizacije.

D: Person

R: Obrazovne ustanove, Organization

Subproperty Of: OsobaDjelatnost

label: jeDopisniČlan

comment: Relacija ukazuje da je određena osoba bila dopisni član obrazovne ustanove i organizacije.

D: Person

R: Obrazovne ustanove, Organization

Superproperties: jeČlan

label: jeRedovitiČlan

comment: Relacija ukazuje da je određena osoba redoviti član obrazovne ustanove i organizacije.

D: Person

R: Obrazovne ustanove, Organization

Superproperties: jeČlan

Podstvojstvo *OsobaOsoba* sadrži podstvojstva prikazana na Slici 34 te su njihova objašnjena dana u nastavku.



Slika 34: Prikaz hijerarhije ontologijskih odnosa OsobaOsoba

label: JeRekonstruiranoPremaKonceptiji

comment: Relacija ukazuje da je određena osoba rekonstruirala djelo prema koncepciji neke druge osobe tj. relacija povezuje te dvije osobe.

D: Književnik

R: Person

SubProperty Of: OsobaOsoba

label: jeImaoMentora

comment: Relacija povezuje osobu s osobom koja joj je bila mentor.

D: Književnici

R: Person

jeImaoMentora inverse jeMentor

SubProperty Of: OsobaOsoba

label: jeMentor

comment: Relacija ukazuje tko je bio mentor određenom književniku.

D: Person

R: Književnici

jeImaoMentora inverse jeMentor

SubProperty Of: OsobaOsoba

label: jeUzor

comment: Relacija ukazuje da je određena osoba je uzor nekom književniku. Relacija definira kojemu književniku je određena osoba bila uzor (Osoba --- jeUzor ---- Književnik).

D: Person

R: Književnici,

jeUzor inverse jeImaoUzora

SubProperty Of: OsobaOsoba

label: jeImaoUzora

comment: Relacija definira koja je osoba bila uzor određenom književniku (Književnik --- jeImaoUzora--- Osoba).

D: Književnici

R: Person

jeUzor inverse jeImaoUzora

SubProperty Of: OsobaOsoba

label: jeIzvršioUtjecaj

comment: Relacija navodi na koga je sve utjecao određeni književnik svojim stvaralaštvom.

D: Književnici

R: Person

jeIzvršioUtjecaj inverse jeUtjecaoNaKnjiževnika

SubProperty Of: OsobaOsoba

label: jeUtjecaoNaKnjiževnika
comment: Relacija navodi osobu koja je imala utjecaj na određenog književnika.
R: Književnici, Person
jeIzvršioUtjecaj inverse jeUtjecaoNaKnjiževnika
SubProperty Of: OsobaOsoba

label: jeNamijenioPosvetu
comment: Relaciju ukazuje da je jedna osoba namijenila posvetu drugoj osobi.
Vidi i: relaciju jePosvećeno koja ukazuje da je određeno književno djelo ili književna vrsta posvećena određenoj osobi
R: Person
SubProperty Of: OsobaOsoba

label: jeParafraziraoKnjiževnika
comment Književnik je parafrazirao neku drugu osobu (književnika, npr. "parafrazirao Ovidija, Vergilija")
R: Person

label: jePrepisivaoDjelaAutora
comment: Relacija definira autore čija je djela određeni književnik prepisivao.
R: Person
SubProperty Of: OsobaOsoba

label: jePrepisivaoZa
comment: Relacija definira za koje je osobe određeni književnik prepisivao djela.
R: Person
SubProperty Of: OsobaOsoba

label: jePrilagodioDjeloAutora
comment: Relacija ukazuje da je određeni književnik prilagodio književno djelo nekog drugog književnika (u tom slučaju relacija povezuje te dvije osobe).
D: Književnik
R: Književno djelo, Književnici, Person
SubProperty Of: OsobaOsoba

label: jePriredioDjelaAutora
comment: Relacija ukazuje da je određena osoba priredila djela određenog autora tj. ne imenuje ta djela. R: Književnik
SubProperty Of: OsobaOsoba

label: jePronašaoDjeloKnjiževnika
comment: Određena osoba je pronašla djelo književnika tj. imenuje se književnik, a ne navodi se naslov djela
R: Književnik

SubProperty Of: OsobaOsoba

label: jeUčenik

comment: Relacija ukazuje da je književnik bio učenik određene osobe

R: Person

label: jeUčitelj

comment: Relacija definira da je jedna osoba bila učitelj drugoj osobi.

R: Person

SubProperty Of: OsobaOsoba

label: seDopisivao

comment: Relacija ukazuje s kim se određena osoba dopisivala.

R: Person

SubProperty Of: OsobaOsoba

label: jePisaoPoslanice

comment: Relacija ukazuje da je osoba pisala poslanice nekoj osobi.

R: Person

SubProperty Of: OsobaOsoba

label: jePreradioDjelaAutora

comment: Relacija navodi autora čija je djela preradio određeni književnik.

R: Književnici

SubProperty Of: OsobaOsoba

label: jePrevodioAutora

comment: Relacija ukazuje kojeg je književnika prevodio određeni književnik.

R: Književnici

jePrevodioAutora inverse jePrevođeniAutorOd

SubProperty Of: OsobaOsoba

label: jePrevođeniAutorOd

comment: Relacija ukazuje koji je autor prevođen od određenog književnika.

R: Književnici

jePrevodioAutora inverse jePrevođeniAutorOd

SubProperty Of: OsobaOsoba

label: jeUglazbioDjeloAutora

comment: Relacija definira čije je književno djelo uglazbio određeni autor tj. navodi da je određena osoba uglazbila književno djelo određenog književnika.

R: Person

SubProperty Of: OsobaOsoba

Svojstvo *OsobaOsoba* sadrži podsvojstvo *Odnosi* i podsvojstvo *ProučavanjeOsoba*.

Podsvojstvo *Odnosi* sadrži sljedeća podsvojstva:

label: jeKum

comment: Relacija opisuje da je određena osoba kum nekoj osobi.

R: Person

Subproperty Of: Odnosi

label: jeKći

comment: Relacija povezuje određenu osobu s roditeljom.

D: Person

Subproperty Of: Odnosi

label: jeNećak

comment: Relacija ukazuje na srodstvo.

R: Person

Subproperty Of: Odnosi

label: jePrijatelj

comment: Relacija definira odnos prijateljstva između dvije osobe.

R: Person

Subproperty Of: Odnosi

label: jeSestra

comment: Relacija definira odnos srodstva.

R: Person

Subproperty Of: Odnosi

label: jeSin

comment: Relacija definira odnos srodstva.

R: Person

Subproperty Of: Odnosi

label: jeSupruga

comment: Relacija ukazuje na suprugu određenog književnika.

R: Person

Subproperty Of: Odnosi

label: jeSuprug

comment: Relacija ukazuje na suprugu određenog književnika.

R: Person

Subproperty Of: Odnosi

label: jeUnuk

comment: Relacija definira da je jedna osoba unuk/unuka drugoj osobi.

R: Person

Subproperty Of: Odnosi

Svojstvo *ProučavanjeOsoba* definira proučavanje određene osobe od strane književnika te navedeno svojstvo sadrži podsvojstva prikazana u nastavku.

label: jeBibliografskiObradio

comment: Relacija ukazuje koje je osobe bibliografski obradio određeni književnik.

R: Person

Subproperty Of: ProučavanjeOsoba

label: jeBiografOsobe

comment: Relacija definira koje su osobe biografski obradile određenog književnika.

R: Person

Subproperty Of: ProučavanjeOsoba

label: jePisaoKritiku

comment: Relacija ukazuje o kojim je osobama tj. književnicima određeni autor tj. književnik pisao kritiku.

R: Person

Subproperty Of: ProučavanjeOsoba

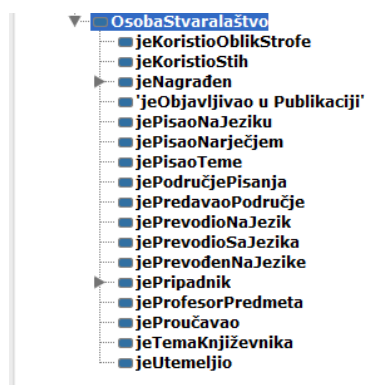
label: jePisaoODjelimaAutora

comment: Relacija definira o čijim je djelima (kojih autora) pisao određeni književnik.

R: Person

Subproperty Of: ProučavanjeOsoba

Podsvojstvo *OsobaStvaralaštvo* omogućuje opis stvaralaštva određenog književnika podsvojstvima prikazanim na Slici 35, čija objašnjena slijede u nastavku.



Slika 35: Prikaz hijerarhije ontologijskih odnosa OsobaStvaralaštvo

label: jeKoristioOblikStrofe

comment: Relacija opisuje koji je oblik strofe u svom stvaralaštvu upotrebljavao određeni književnik

D: Književnici

R: Strofa

Superproperties: OsobaStvaralaštvo

label: jeKoristioStih

comment: Relacija opisuje koji je stih upotrebljavao određeni književnik

D: Književnici

R: Stih

Superproperties: OsobaStvaralaštvo

label: jeNagrađen

comment: Relacija definira koju je nagradu primio književnik.

D: Književnici,

R: Nagrada

Superproperties: jeNagrađen

label: jeObjavljivaoUČasopisu

comment: Relacija ukazuje u kojem je časopisu objavljivao pojedini autor.

D: Književnici

R: Časopisi

Superproperties: OsobaStvaralaštvo

label: jePisaoNaJeziku

comment: Relacija definira jezike kojim je određeni književnik pisao.

D: Književnici

R: Jezik

Superproperties: OsobaStvaralaštvo

label: jePisaoNarječjem

comment: : Relacija definira narječje kojim je određeni književnik pisao.

D: Književnici,

R: Narječje

Superproperties: OsobaStvaralaštvo

label: jePisaoTeme

comment: Relacija definira teme koje je određeni književnik pisao.

D: Književnici,

R: Teme

jeTemaKnjiževnika inverse jePisaoTeme

Superproperties: OsobaStvaralaštvo

label: jeTemaKnjiževnika

comment: Relacija ukazuje koji su sve književnici pisali o određenoj temi.

D: Teme

R: Književnici

jeTemaKnjiževnika inverse jePisaoTeme

Superproperties: OsobaStvaralaštvo

label: jePisaoKnjiževneVrste

comment: Relacija definira odnos između književnika i književnih vrsta tj. ukazuje koje je književne vrste određeni književnik pisao u svom stvaralaštvu.

D: Književnik

R: Književni rodovi i vrste

SubProperty Of: OsobaStvaralaštvo

label: jePodručjePisanja

comment: Relacija definira iz kojeg je područja osoba pisala djela.

D: Književnici

R: Područje

Superproperties: OsobaStvaralaštvo

label: jePredavaoPodručje

Relacija definira područje koje je osoba predavala.

D: Književnici

R: Područje

Superproperties: OsobaStvaralaštvo

label: jePrevodioNaJezik

comment: Relacija definira jezike na koje je prevodio određeni književnik.

D: Književnici

R: Jezik

Superproperties: OsobaStvaralaštvo

label: jePrevodioSaJezika

comment: Relacija definira jezike sa kojih je prevodio određeni književnik.

D: Književnici

R: Jezik

Superproperties: OsobaStvaralaštvo

label: jePrevođenNaJezike

comment: Relacija definira na koje su jezike prevođena djela određenog književnika.

D: Književnici

R: Jezik

Superproperties: OsobaStvaralaštvo

label: jePripadnik

comment: Relacija definira kojem razdoblju su pripadali određeni književnici ili književna djela.

D: Književno djelo, Književnici, Person

R: 'Arhajsko doba' , 'Arhajsko razdoblje' , 'Barokna književnost' , 'Druga moderna' , 'Elizabetansko doba' , 'Grčka književnost' , 'Kasni modernizam' , 'Kasno carstvo' , 'Klasično doba' , 'Književnici u izbjeglištvu' , 'Književnost antike' , 'Književnost klasicizama' , 'Književnost modernizama' , 'Književnost prosvjetiteljstva' , 'Književnost realizma' , 'Književnost renesanse' , 'Književnost romantizma' , 'Književnost srednjovjekovlja' , 'Moderni objektivizam' , 'Periodizacija književnosti' ,

'Prva moderna' , 'Rimska književnost' , 'Rimsko doba' , 'Socijalistički realizam' , 'Srebrno razdoblje' , 'Stranačka organizacija' , 'Sturm und Drang' , 'Uvodno rimsko razdoblje' , 'Zlatno doba' , Akmeizam , Avangarda ,Bidermajer, Dadaizam, Egzistencijalizam, Ekspresionizam, Esteticizam, Eufizam, Futurizam, Gongorizam, Helenizam, Humanizam, Ilirizam, Imagizam, Knjizevnost_od_1929._do_1952, Konstruktivizam, Krugovaši, Manirizam, Moderna, Nadrealizam, Naturalizam, Neoromantizam, Parnasizam, Postmodernizam, Precioznost, Predrealizam, Predrenesansa, Predromantizam, Protureformacija, Razlogovci, Reformacija, Rokoko, Secesija, Sentimentalizam, Simbolizam, Stilski pluralizam, Verizam, Zenitizam
Superproperties: OsobaStvaralaštvo

label: jeKnjizevnikPripadnikKnjiževneGeneracije

comment: Relacija definira kojoj književnoj generaciji pripada određeni književnik.

D: Književnici

R: 'Književnici u izbjeglištvu', Krugovaši, Razlogovci,

Superproperties: jePripadnik

label: jeKnjizevnikPripadnikKnjiževneSkupine

comment: Relacija definira kojoj književnoj skupini pripada određeni književnik.

D: Književnici

R: Gričani

Superproperties: jePripadnik

label: jeKnjizevnikPripadnikKnjiževneŠkole

comment: Relacija definira kojoj književnoj školi pripada određeni književnik.

D: Književnici

R: Parnasizam

Superproperties: jePripadnik

label: jeKnjizevnikPripadnikKnjiževnogPokreta

comment: Relacija definira kojem književnom pokretu pripada određeni književnik.

D: Književnici

R: 'Sturm und Drang' , Avangarda, Dadaizam, Ekspresionizam, Futurizam, Ilirizam, Imagizam, Konstruktivizam, Nadrealizam, Zenitizam

Superproperties: jePripadnik

label: jePripadnikKnjiževnogPravca

comment: Relacija definira kojem književnom pravcu pripada određeni književnik ili književno djelo.

D: Književnici, Književno djelo

R: Egzistencijalizam, Ekspresionizam, Eufizam, Gongorizam, Manirizam, Naturalizam, Neoromantizam, Precioznost, Sentimentalizam, Simbolizam, Verizam

Superproperties: jePripadnik, Opis djela

label: jePripadnikKnjiževnogRazdoblja

comment: Relacija definira kojem književnom razdoblju pripada određeni književnik ili književno djelo tj. relacija navodi da određeno književno djelo ili autor pripadaju određenom književnom periodu.

D: Književno djelo', Književnik

R: 'Arhajsko doba', 'Arhajsko razdoblje', 'Barokna književnost', 'Druga moderna', 'Elizabetansko doba', 'Grčka književnost', 'Kasni modernizam', 'Kasno carstvo', 'Klasično doba', 'Književnost antike', 'Književnost klasicizama', 'Književnost modernizama', 'Književnost od 1929. do 1952.', 'Književnost prosvjetiteljstva', 'Književnost realizma', 'Književnost renesanse', 'Književnost romantizma', 'Književnost srednjovjekovlja', 'Prva moderna', 'Rimska književnost', 'Rimsko doba', 'Srebreno razdoblje', 'Uvodno rimsko razdoblje', 'Zlatno doba', Esteticizam, Helenizam, Humanizam, Moderna, Postmodernizam, Predrenesansa, Predromantizam, Protureformacija, Reformacija

Superproperties: jePripadnik, Opis djela

label: jeProfesorPredmeta

comment: Relacija definira iz kojeg je područja određena osoba bila profesor.

D: Književnik

R: Područje

Superproperties: OsobaStvaralaštvo

label: jeProučavao

comment: Relacija definira što je predmet proučavanja određenog književnika.

D: Književnici

R: Entity

Superproperties: OsobaStvaralaštvo

label: jeUtemeljio

comment: Relacija ukazuje koje je serijske publikacije, biblioteke, organizacije, čitaonice i sl. utemeljio pojedini književnik.

D: Književnici,

R: 'Serijske publikacije', Biblioteke, Organization, Čitaonice

Superproperties: OsobaStvaralaštvo

7.3.5.3 Svojstva autora

Svojstva autora proizlaze iz odnosa klasa *Književnik* i *Književno djelo*. Njihova hijerarhija je prikazana Slikom 36, a u nastavku su pojašnjenja za svaki od njih.



Slika 36: Hijerarhija SvojstvaAutora

label: jeAutor

comment: Relacija ukazuje koje je književno djelo, scenarij (filmski scenarij, televizijski scenarij) napisao određeni književnik.

D: Književnici

R: 'Filmski scenarij', 'Književno djelo', 'Televizijski scenarij', Scenarij

autorDjelaJe inverse jeAutor

Superproperties: Svojstva Autora

label: autorDjelaJe
comment: Relacija definira autora određenog djelo.
D: Književno djelo
R: Književnici
autorDjelaJe inverse jeAutor
Superproperties: Svojstva Autora

label: jeAutorScenarijaZaFilm
comment: Relacija ukazuje tko je autor scenarija za određeni film.
D: Person
R: Film
Superproperties: jeAutor

label: djeloJePriredio
comment: Relacija definira koji je književnik priređivao određene književne vrste, književna djela i izdanja određenog izdavača.
Inverse Of: jePriredio
D: Književno djelo, Izdavač, Književni rodovi i vrste
R: Književnik
Superproperties: Svojstva Autora

label: jePriredio
altLabel: jePripremio
comment: Relacija ukazuje da je određena osoba priredila određena književna djela ili vrste, ili pak djela određenog autora te djela određenog izdavača.
D: Književnici
R: 'Književni rodovi i vrste', 'Književno djelo', Izdavač, Književnici
jePriredio inverse djeloJePriredio
Superproperties: Svojstva Autora

label: jeDovršio
comment: Relacija ukazuje koje je književno djelo dovršila određena osoba.
D: Književnici
R: Književno djelo
Superproperties: Svojstva Autora

label: jeNapisaoDopunuDjela
comment: Relacija definira kojem je književnom djelu napisala dopunu određena osoba.
D: Književnici,
R: 'Književno djelo'
Superproperties: jeDovršio

label: jeIlustrirao
comment: Relacija definira koje je književno djelo ilustrirala određena osoba.

D: Person
R: 'Književno djelo'
Superproperties: Svojstva Autora

label: jeKomentirao
comment: Relacija definira koje je književno djelo komentirao određeni književnik.

D: Književnici
R: Književno djelo
Superproperties: Svojstva Autora
label: jeParafraziraoDjelo
comment: Književnik je parafrazirao književno djelo.

D: Književnici
R: 'Književno djelo'
Superproperties: Svojstva Autora

label: jePrejevao
comment: Relacija definira koje je književno djelo ili oratorij prepjevao određeni književnik.

D: Književnici
R: 'Književno djelo', Oratorij
Superproperties: Svojstva Autora

jePrejevDjelaAutora
comment: Relacija definira kojeg je autora (ne imenuje se djelo nego autor određenog književnog djela) određeno književno djelo prepjev.

D: Književno djelo
R: Književnik
Superproperties: Svojstva Autora

label: jePreradio
comment: Relacija ukazuje koje je književno djelo preradio određeni književnik. Preradba književnog djela je njegovo podvrgavanje obradi, preinačavanje, izmjenjivanje djela znatno ili u potpunosti.

D: Književnici,
R: 'Književno djelo'
jePrerađeno inverse jePreradio
Superproperties: Svojstva Autora

label: jePrerađenoOdStraneKnjiževnika
comment: Relacija odgovara na pitanje koji je književnik preradio određeno književno djelo.

Inverse Of: jePreradio
D: Književno djelo
R: Književnik
Superproperties: Svojstva Autora

label: jePrilagodioDjelo

label: jeAdaptiraoDjelo

comment:Relacija ukazuje da je određeni književnik prilagodio određeno književno djelo.

D: Književnik

R: Književno djelo

Subproperty: jePreradio

label: jeDramatizirao

comment: Relacija ukazuje da je određeni književnik dramatizirao određeno djelo : npr.

"dramatizirao epizodu iz Ovidijevih Prijetvorbi""

comment: „prilagodba književnog teksta, obično pripovjednoga ali i dramskoga, za potrebe kazališne izvedbe te filmske ili televizijske ekranizacije. Uglavnom proširuje dijalog i monolog i na one dionice teksta što su čitatelju posredovane glasom pripovjedača, pa bi zbog toga ostale nedostupne opažaju publike. U nekim se istraživanjima (npr. kod N. Batušića) razlikuju teorijska dramaturgija (teorija drame) i praktična dramaturgija (prilagodba teksta za kazališnu izvedbu)."

D: Književnik

R: Književno djelo

Superproperties: jePrilagodioDjelo

label: jePrilagodioZaFilm

altLabel: jeAdaptiraoZaFilm

altLabel: jeEkranizirao

comment: Relacija definira koji je književnik prilagodio određeno književno djelo za film, npr.„Brezu je vrlo uspješno ekranizirao A. Babaja u istoimenome filmu iz 1967."

D: Književno djelo

R: Književnik

Superproperties: jePrilagodioDjelo

label: jePrilagodioZaScenu

altLabel: jeAdaptiraoZaScenu

comment: Relacija definira koji je književnik prilagodio određeno književno djelo za scenu npr. "God. 1976. Kiklopa je za scenu adaptirao K. Spaić, a 1982. za film A. Vrdoljak. "

D: Književno djelo

R: Književnik

Superproperties: jePrilagodioDjelo

label: jePrevedeno

comment: Relacija ukazuje koji su sve autori prevodili određeno djelo ili književne vrste.

D: Književno djelo, Književni rodovi i vrste

R: Književnik

Inverse of: jePreveo

Superproperties: Svojstva Autora

label: jePreveo
comment: Relacija ukazuje koja je djela preveo određeni autor.
D: Književnik
R: Književno djelo, Književni rodovi i vrste
Inverse of: jePrevedeno
Superproperties: Svojstva Autora

label: jePrviPreveo
comment: Relacija ukazuje koje je književno djelo, rod ili vrstu prvi preveo određeni književnik.
D: Književnici,
R: 'Književni rodovi i vrste', 'Književno djelo'
Superproperties: jePreveo

label: jeRekonstruiraoKrajDjela
comment: Relacija definira koji je književnik rekonstruirao kraj određenog književnog djela, npr. Tek je 1950. kraj romana, prema Kozarčevoj koncepciji, rekonstruirao E. Štampar.
D: Književno djelo
R: Književnik
Superproperties: Svojstva Autora

label: jeRežirao
comment: Relacija ukazuje koja je književna djela, drame, filmove ili opere režirao određeni književnik.
D: Književnici
R: 'Književno djelo', Drama, Film, Opera
Superproperties: Svojstva Autora

label: jeSabrao
comment: Relacija ukazuje da je autor sabrao određene književne vrste ili književna djela.
D: Književnici
R: 'Književni rodovi i vrste', 'Književno djelo'
Superproperties: Svojstva Autora

label: jeScenirao
comment: Relacija definira koje je televizijske serije, drame, komedije, radio-drame scenirao određeni književnik.
D: Književnici,
R: 'Televizijska serija', 'Televizijske drame', Komedija, Radio-drama
Superproperties: Svojstva Autora

label: jeSkladao
altLabel: jeUglazbio

comment: Relacija definira koje je književne vrste, djela ili glazbu skladao određeni književnik.

D: Književnici,

R: Književni rodovi i vrste, Književno djelo, Glazba

Superproperties: Svojstva Autora

label: jeSuradnik

comment: Relacija definira kojih serijskih publikacija je suradnik određeni književnik, kao i koja dva književnika su surađivala u radu.

D: Književnici

R: Person, Serijske publikacije

Superproperties: Svojstva Autora

label: jeSuautor

altLabel: jeKoautor

comment: Relacija definira koja dva književnika su autori određenog djela.

R: Književnik

Symmetric: jeSuautor

Superproperties: jeSuradnik

label: jeUređivaoSa

comment: Relacija definira koja dva književnika su zajedno uređivala neku publikacija.

R: Književnik

Superproperties: jeSuradnik

label: jeUrednik

comment: Relacija definira kojih je književnih vrsta, djela, serijskih publikacija, izdavača, društava i organizacija bio urednik određeni književnik.

D: Književnici,

R: 'Književni rodovi i vrste', 'Književno djelo', 'Serijske publikacije', Društva, Izdavač, Organization

jeUrednik inverse uređivaoJe

Superproperties: Svojstva Autora

label: uređivaoJe

comment: Relacija definira književnike koji su uređivali određeno književno djelo, serijske publikacije, društva, izdavača ili organizacije.

D: 'Književno djelo', 'Serijske publikacije', Društva, Izdavač, Organization

R: Književnici,

jeUrednik inverse uređivaoJe

Superproperties: Svojstva Autora

label: jeEnciklopedijskiUrednik

comment: Relacija definira enciklopedije kojih je bio urednik određeni književnik.

D: Književnici,
R: Enciklopedije
Superproperties: jeUrednik

label: jeFilmskiUrednik
comment: Relacija definira enciklopedije kojih je bio urednik određeni književnik iz filmskog područja.

D: Književnici,
R: Enciklopedije
Superproperties: jeEnciklopedijskiUrednik

label: jeUrednikPoredbeneKnjiževnosti
comment: Relacija definira enciklopedije kojih je bio urednik određeni književnik iz područja poredbene književnosti.

D: Književnici,
R: Enciklopedije
Superproperties: jeEnciklopedijskiUrednik

label: jeUrednikTeorijeKnjiževnosti
comment: Relacija definira enciklopedije kojih je bio urednik određeni književnik iz područja teorije književnosti.

D: Književnici,
R: Enciklopedije
Superproperties: jeEnciklopedijskiUrednik

label: jeGlavniUrednik
comment: Relacija definira kojih je serijskih publikacija, biblioteka, listova, tjednika bio glavni urednik određeni književnik.

D: Književnici,
R: 'Serijske publikacije', Biblioteke, Listovi, Tjednik
Superproperty: jeUrednik

label: jePrviUrednik
comment: Relacija definira serijske publikacije kojih je bio prvi urednik određeni književnik.

D: Književnici,
R: 'Serijske publikacije'
Superproperties: jeUrednik

label: jeUmjetničkiUrednik
comment: Relacija definira serijske publikacije kojih je bio urednik određeni književnik iz područja umjetnosti.

D: Književnici,
R: 'Serijske publikacije'
Superproperties: jeUrednik

label: jeUrednikKazališneRubrike

comment: Relacija definira serijske publikacije kojih je bio urednik određeni književnik iz kazališne rubrike.

D: Književnici,

R: 'Serijske publikacije'

Superproperties: jeUrednik

label: jeUrednikKnjiževnosti

comment: Relacija definira radio postaje, serijske publikacije i biblioteke kojih je bio urednik određeni književnik iz područja književnosti.

D: Književnici,

R: 'Radio postaje', 'Serijske publikacije', Biblioteke

Superproperties: jeUrednik

label: jeUrednikKulturneRubrike

comment: Relacija definira serijske publikacije kojih je bio urednik određeni književnik iz kulturne rubrike.

D: Književnici

R: 'Serijske publikacije'

Superproperties: jeUrednik

label: jeUrednikLikovnihUmjetnosti

comment: Relacija definira radio postaje, serijske publikacije kojih je bio urednik određeni književnik iz područja likovnih umjetnosti.

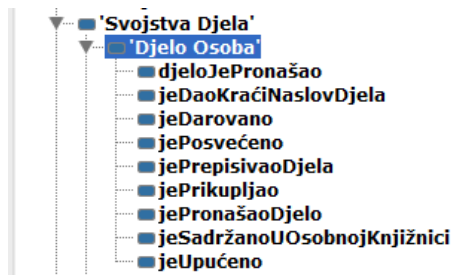
D: Književnici,

R: 'Radio postaje', 'Serijske publikacije'

Superproperties: jeUrednik

7.3.5.4 Svojstva djela

Svojstva djela proizlaze iz odnosa koje uspostavlja književno djelo s klasama *Person*, *Književnik*, podklasama klase *Periodizacija književnosti*, *Strofa*, *Književni rodovi i vrste*, *Izdavač*, *Organization*, *Figure*, *Stih*, *Društva*, *Kazalište*, *Jezik*, *Teme*, *Narječje*, sa svim podklasama klase *Dokument*, *Pismo*. Svojstva djela sadrži podsvojstva *DjeloOsoba* i *Opis djela*. *DjeloOsoba* povezuje određena djela s pojedinim osobama, sadrži podsvojstva prikazana na Slici 37 čija objašnjenja slijede u nastavku.



Slika 37: Podsojstva DjeloOsoba

label: djeloJePronasao

comment: Relacija definira koji je književnik pronašao određeno književno djelo.

Inverse Of: jePronasaoDjelo

D: Književno djelo,

R: Književnik

Sub Propetry Of: Djelo Osoba

label: jePronasaoDjelo

comment: Relacija definira koje je književno djelo pronašao određeni književnik.

D: Književnici

R: 'Književno djelo'

inverseOf: djeloJePronasao

Sub Propetry Of: Djelo Osoba

label: jeDaoKraciNaslovDjela

comment: Relacija ukazuje na književnika koji je djelu dodijelio kraći naslov.

D: Književno djelo,

R: Književnik

Sub Propetry Of: Djelo Osoba

label: jeDarovano

comment: Relacija definira kojem je književniku ili osobi darovano određeno književno djelo.

D: Književno djelo

R: Person

Sub Propetry Of: Djelo Osoba

label: jePosvećeno

comment: Relacija definira kojoj je osobi posvećeno određeno književno djelo ili vrsta.

Npr. "...malen zbornik hrvatskih poslovice i poslovičnih izraza Ilirske izreke (Adagia Illyrica), koji je posvetio dubrovačkomu plemiću Luji Saračiću."

D: Književni rodovi i vrste, Književno djelo

R: Person

Sub Propetry Of: Djelo Osoba

label: jePrepisivaoDjela

comment: Relacija definira književna djela i vrste koje je pojedina osoba prepisivala, ali iz nje proizlazi i svojstvo jePrepisivaoDjelaAutora koje navodi autore čija je djela određena osoba prepisivala, ali ne imenuje i sama djela.

D: Person

R: 'Književni rodovi i vrste', 'Književno djelo'

Sub Propetry Of: Djelo Osoba

label: jePrikupljao

comment: Relacija definira književna djela i vrste koje je određena osoba prikupljala.

D: Person

R: 'Književni rodovi i vrste', 'Književno djelo', Glazba

Sub Propetry Of: Djelo Osoba

label: jeSadržanoUOsobnojKnjižnici

comment: Relacija definira u čijoj knjižnici (kojeg književnika ili osobe) je određeno književno djelo ili vrsta sadržano (npr., Pjesmu je u svojoj knjižnici imao i M. Marulić).

D: Književno djelo, Književni rodovi i vrste

R: Person

Sub Propetry Of: Djelo Osoba

label: jeUpućeno

comment: Relacija definira kojem književniku ili osobi je upućeno određeno književno djelo (npr. „... pohvalna pjesma upućena B. Betteri.“).

D: Književno djelo

R: Person

Sub Propetry Of: Djelo Osoba

Opis djela uspostavlja odnose koje opisuju obilježja djela (npr. pismo djela, narječje u djelu i sl.), sadrži podsvojstva prikazana na Slici 38 čija objašnjenja slijede u nastavku:



Slika 38: Prikaz podsvojtva Opis djela

label: ElementiKnjiževnogPokreta

comment: Relacija definira koje karakteristike određenog književnog pokreta sadrži određeno književno djelo.

D: Književno djelo

R: Futurizam, Ilirizam, Imagizam, Nadrealizam, Avangarda, Konstruktivizam, Humanizam, Zenitizam, Dadaizam, Akmeizam, Sturm und Drang, Ekspresionizam

Sub Propetry Of: Opis djela

label: imaOblikStrofe

comment: Relacija definira koji oblik strofe sadrži određeno književno djelo.

D: Književno djelo

R: Strofa

Sub Propetry Of: Opis djela

label: imaPohranjenaDjela

comment: Relacija definira koja književna djela ili književna vrste ima pohranjena određena ustanova.

Inverse Of: jePohranjeno

D: Organization

R: Književni rodovi i vrste, Književno djelo

Sub Property Of: Opis djela

label: jePohranjeno

comment: Relacija definira u kojoj je ustanovi (Organization) pohranjeno određeno djelo

Inverse Of: imaPohranjenaDjela

D: Književni rodovi i vrste, Književno djelo

R: Facility, Samostani, Organization

Sub Property Of: Opis djela

label: imaPrisutneFigure

comment: Relacija definira koje književne figure su prisutne u određenom djelu.

D: Književno djelo

R: Figure

Sub Property Of: Opis djela

label: imaStih

comment: Relacija definira koje stihove sadrži određeno književno djelo.

D: Književni rodovi i vrste, Književno djelo

R: Stih

Sub Property Of: Opis djela

label: imaStilskoObilježje

comment: Relacija definira koje stilsko obilježje sadrži određeno književno djelo.

D: Književno djelo

R: Moderni objektivizam, Stilski plurizam, Rokoko, Socijalistički realizam, Secesija, Bidermajer

Sub Property Of: Opis djela

label: imaTravestiju

comment: Relacija definira koje književno djelo ima odgovarajuću travestiju.

Inverse Of: jeTravestijaDjela

R: Književno djelo

Sub Property Of: Opis djela

label: jeTravestijaDjela

comment: Relacija definira kojeg književnog djela je određeno djelo travestija (npr. „...travestija Smrt Babe Čengiđkinje na poznati Mažuranićev ep, 1880.“)

R: Književno djelo

Inverse Of: imaTravestiju

Sub Property Of: Opis djela

label: jeDodatak

comment: Samostalna cjelina ili književno djelo koji je dodatak nekoj publikaciji.

R: Književno djelo
Sub Propetry Of: Opis djela

label: jeGrađaZaPisanjeDjela
comment: Relacija definira koje je književno djelo poslužilo kao građa za pisanje nekog drugog književnog djela. npr.: Kao građa poslužilo je i Matijinu bratu Ivanu u pisanju spjeva Smrt Smail-age Čengića.

R: Književno djelo
Sub Propetry Of: Opis djela

label: jeJezikDjela
comment: Relacija definira jezik kojim je književno djelo ili književna vrsta napisana.
D: Književni rodovi i vrste, Književno djelo
R: Jezik
Sub Propetry Of: Opis djela

label: jeKnjiževnaVrsta
comment: Relacija definira kojoj književnoj vrsti pripada određeno književno djelo.
D: Književno djelo
R: Improvizirana komedija, Stručno pedagoški članci, radio-drama, epsko kazalište, novinarska reportaža, televizijske drame, misterij, vodvilj, komedija naravi, ulična komedija, pasion, satirične drame, epistola, putopisi, građanska drama, pregled, publicistika, autobiografija, farsa, mortalitet, komedija intrige, alegorija, burleska, filmski scenarij, auto, pismo, osvrt, parabola, travestija, komedija situacije, povijesne drame, kazališna kritika, članak, esej, eruditna komedija, tragikomedija, povijesna tragedija, komedija konverzacije, biografija, mim, pastoral, polemika, satira, glazbena kritika, hagiografska komedija, komedija karaktera, drama, pitalica, didaktičke vrste, lakrdija, drama apsurdna, povijesna drama, feljton, libreto, književni članci, napisi, filmska kritika, parodija, prikazi, memoari, igrokaz, egzemplum, televizijski scenarij, drama u stihu, maskareta, tekstološki radovi, crkvena prikazanja, diskurzivni književni oblici, satirični epigram, drama u užem smislu, basna, apologija, književni rodovi i vrste, reportaža, mirakul, tragedija, dnevnik, likovna kritika, koserije, epigram, kritika, komedija, kazališni članci, dramolet, hagiografija, dramska kronika, pjesnička poslanica, govor, melodrama, rasprava, scenarij
Sub Propetry Of: Opis djela

label: jeNadahnuto
comment: Relacija definira tematiku kojom je nadahnuto određeno književno djelo. Relacija također ukazuje na književno djelo koje je poslužilo kao nadahnuće za određeno djelo tj. povezuje dva djela.
altLabel: jeInspirirano
D: Književno djelo
R: Književno djelo, Teme
Sub Propetry Of: Opis djela

label: jeNapisanoPremaDjelu

comment: Relacija definira književno djelo prema kojemu je napisano određeno književno djelo.

R: Književno djelo

Sub Propetry Of: Opis djela

label: jeNarječjeDjela

comment: Relacija definira narječje književnog djela.

D: Književno djelo

R: Narječje

Sub Propetry Of: Opis djela

label: jeObjavljenoUDjelu

comment: Relacija definira u kojoj je publikaciji ili književnom djelu objavljeno određeno književno djelo ili književna vrsta.

D: Književni rodovi i vrste, Književno djelo

R: Publikacije, Serijske publikacije, Književno djelo

Sub Propetry Of: Opis djela

label: objavilaJe

comment: Relacija definira koje je književno djelo ili vrstu objavila određena publikacija ili određeno djelo.

Inverse Of: jeObjavljenoUDjelu

D: Serijske publikacije, Publikacije, Književno djelo

R: Književno djelo, Književni rodovi i vrste

Sub Propetry Of: Opis djela

label: jeOblikom

comment: Relacija definira oblik publikacije ili književnog djela.

D: Književno djelo

R: Katalog, Biblija, rubrika, Knjiga rasprava, misali, antički natpis, antologije, revije, časopisi za znanost o književnosti, knjiga pripovijetki, referentna djela, kodeks, zbirka autobiografskih zapisa, knjige članaka, knjiga drama, pojmovnik, traktat, kalendari, biblioteke, Knjiga ogleđa, Studija, Knjiga eseja, Listovi, Rijetke knjige, Kanconijer, Brošure, Prilozi, Zbornici, Zbirka novela, Zbirka članaka, Glasilo, Zbirka dječjih pjesama, Godišnjak, Pjesnički zbornici, Sastav, Molitvenici, Poučne knjižice, Zbirka aforizama, Zbirka pjesama, Inkunabule, Manifest, Biblija u stihu, Monografije, Trilogija, Teološki priručnici, Dokument, Zbirka pripovijedaka, Publikacije, Zapis, Zbirka proza, Knjiga govora.

Sub Propetry Of: Opis djela

label: jeOdgovorDjelu

comment: Relacija definira kojem je književnom djelu određeno djelo odgovor. Primjer:

...polemičku knjižicu stihova Opravdanje dobri žena (1802), odgovor na ženomrzačke stihove iz Kite cvitja razlikova I. Ivaniševića.

R: Književno djelo

Sub Propetry Of: Opis djela

label: jeParodija

comment: Relacija definira kojem je književnom djelu ili kojem autoru nekog književnog djela (nije imenovano djelo nego autor) parodija određeno književno djelo.

D: Književno djelo

R: Književno djelo, Književnik

Sub Propetry Of: Opis djela

label: jePisanoPismom

comment: Relacija definira kojim je pismom pisano određeno književno djelo.

Inverse Of: jePismoDjela

D: Književni rodovi i vrste, Književno djelo

R: Pismo teksta

Sub Propetry Of: Opis djela

label: jePismoDjela

comment: Relacija definira koja književna djela su pisana određenim pismom (npr. vjerskih knjiga tiskanih na glagoljici, ćirilici i latinici)

Inverse Of: jePisanoPismom

D: Pismo teksta

R: Književni rodovi i vrste, Književno djelo

Sub Propetry Of: Opis djela

label: jePrepjev

comment: Relacija definira kojem je književnom djelu određeno djelo prepjev.

"Jedino je njegovo poznato djelo hagiografski epilij Istorija svetoga Ivana Zlatoustoga, prepjev talijanske stihovane legende nepoznata autora La storia di S. Giovanni Boccadoro..."

R: Književno djelo

Sub Propetry Of: Opis djela

label: jePreradbaDjela

comment: Relacija definira kojeg je književnog djela preradba određeno djelo.

1)Najvažnije je njegovo djelo tragedija Jokasta, preradba istoimenoga djela

R: Književno djelo

Sub Propetry Of: Opis djela

label: jePretiskano

comment: Relacija definira u kojem djelu su pretiskane određene književne vrste ili određeno književno djelo. Npr. Eseji i rasprave pretiskani su mu u knjizi Hrvatsko-bošnjačke teme (1996).

D: Književni rodovi i vrste, Književno djelo

R: Književno djelo

Sub Propetry Of: Opis djela

label: jePretiskanoIzDjela

comment: Relacija definira iz kojeg je djela pretiskano određeno književno djelo ili književna vrsta. Npr. objavljeni u Hrvatskoj reviji pretiskani su mu u knjizi Hrvatsko-bošnjačke teme (1996).

D: Književno djelo

R: Književni rodovi i vrste, Književno djelo

Sub Propetry Of: Opis djela

label: jePrevedenoUDjelu

comment: Relacija definira u kojem su djelu prevedeni odlomci ili cijelo djelo, npr.: Odlomci putopisa prevedeni su u knjizi U dragom kraju (1942).

R: Književno djelo

Sub Propetry Of: Opis djela

label: jePrijevod

comment: Relacija definira kojeg je djela određeno književno djelo prijevod.

R: Književno djelo

Sub Propetry Of: Opis djela

label: jePripadnikKnjiževnogPravca

comment: Relacija definira kojem književnom pravcu pripada određeni književnik ili književno djelo.

D: Književnici, Književno djelo

R: Egzistencijalizam, Ekspresionizam, Eufizam, Gongorizam, Manirizam, Naturalizam, Neoromantizam, Precioznost, Sentimentalizam, Simbolizam, Verizam

Superproperties: jePripadnik, Opis djela

label: jePripadnikKnjiževnogRazdoblja

comment: Relacija definira kojem književnom razdoblju pripada pripada određeni književnik ili književno djelo tj. relacija navodi da određeno književno djelo ili autor pripadaju određenom književnom periodu.

D: Književno djelo', Književnik

R: 'Arhajsko doba', 'Arhajsko razdoblje', 'Barokna književnost', 'Druga moderna', 'Elizabetansko doba', 'Grčka književnost', 'Kasni modernizam', 'Kasno carstvo', 'Klasično doba', 'Književnost antike', 'Književnost klasicizama', 'Književnost modernizama', 'Književnost od 1929. do 1952.', 'Književnost prosvjetiteljstva', 'Književnost realizma', 'Književnost renesanse', 'Književnost romantizma', 'Književnost srednjovjekovlja', 'Prva moderna', 'Rimska književnost', 'Rimsko doba', 'Srebreno razdoblje', 'Uvodno rimsko razdoblje', 'Zlatno doba', Esteticizam, Helenizam, Humanizam, Moderna, Postmodernizam, Predrenesansa, Predromantizam, Protoreformacija, Reformacija

Superproperties: jePripadnik, Opis djela

label: jePrisutanIzričajDjela

comment: Relacija utvrđuje izričaj kojeg književnog djela je prisutan u određenom književnom djelu (tj. relacija utvrđuje da je u djelu prisutan utjecaj izričaja nekog drugog djela, npr.: pojedini izričaji Epitalamija odjekuju u Davidijadi).

R: Književno djelo

Sub Propetry Of: Opis djela

label: jeSadržajDjela

comment: Relacija definira sadržaj nekog djela, npr. evanđelja, poslanice su sadržaj Biblije.

Inverse Of: sadrži

R: Književno djelo

Sub Propetry Of: Opis djela

label: sadrži

comment: Relacija definira što sadrži određeno djelo tj. koje manje cjeline sadrži određeno djelo (npr. Biblija sadrži određene poslanice ili evanđelja, djelo se sastoji od pet pjevanja: „Građeno je od 1134 deseteračko-osmeračka stiha raspoređena u pet nejednakih pjevanja, u kojima su prikazana zlodjela tur. silnika Smail-age Čengića nad zarobljenim Crnogorcima (Agovanje), njihovo spremanje na osvetu (Noćnik), sakupljanje osvetničke čete (Četa), napad na zločinca i njegova smrt (Harač), te silnikov udes nakon smrti (Kob).

Inverse Of: jeSadržajDjela

R: Književno djelo

Sub Propetry Of: Opis djela

label: jeTemaDjela

comment: Relacija definira što je tema određenog književnog djela.

D: Književno djelo

R: Teme

Sub Propetry Of: Opis djela

label: jeTiskanUDjelu

comment: Relacija definira u kojem je djelu tiskano određeno književno djelo (npr. „...ciklus pjesama Tuga ljeta, tiskan u zajedničkoj knjizi Lirika (1935) s I. Dončevićem, A. Nizeteom i R. Žilićem.).

R: Književno djelo

Sub Propetry Of: Opis djela

label: jeUtjecaloNaDjelo

comment: Relacija definira koje je književno djelo izvršilo utjecaj na određeno književno djelo.

D: Književno djelo, Književni rodovi i vrste

R: Književno djelo

Sub Propetry Of: Opis djela

label: seIzvodilo

comment: Relacija definira koja je književna djela (naslov), drame ili književnike (tj. njihova djela koja nisu imenovana) izvodila određena kazališna družina ili kazalište.

D: Drama, Književno djelo, Književnik

R: Kazalište, Kazališna družina

Sub Propetry Of: Opis djela

label: suDijelovidjela

comment: Relacija definira dijelove od kojih je sastavljeno neko djelo, npr. epsko-lirsko djelo Zvončac (1670), sastavljeno od triju nepovezanih dijelova (Premišljanja zvrhu četirih poslednjih človeka, Plač smrtelnosti, Rasipnoga sina hištoria).

R: Književno djelo

Sub Propetry Of: Opis djela

7.3.5.5 Svojstva izdavača

Svojstva Izdavača sadrže podsvojstva *izdanoJeOdIzdavača* i *jeIzdao*. Svojstvo *izdanoJeOdIzdavača* povezuje individue klase *Književni rodovi i vrste* te klase *Književno djelo* s individuama klase *Književnici, Izdavač, Person*; a svojstvo *jeIzdao* je inverzno.

label: IzdanoJeOdIzdavača

comment: Relacija definira književnika, izdavača ili osobu koji su izdali određeno književno djelo ili književnu vrstu.

Inverse Of: jeIzdao

D: Književni rodovi i vrste, Književno djelo

R: Person, Izdavač, Književnik

label: jeIzdao

comment: Relacija definira koje književno djelo ili vrsta je izdano od određene osobe, književnika ili izdavača.

Inverse Of: IzdanoJeOdIzdavača

D: Person, Izdavač, Književnik

R: Književni rodovi i vrste, Književno djelo

7.3.5.6 Svojstva za ontologijsko zaključivanje

Ovu skupinu čine svojstva koja omogućuju zaključivanje i klasifikaciju o pojedinim ontologijskim klasama tj. entitetima na temelju njihovih pripadajućih svojstava. Svojstvo *imaOblik* omogućuje automatsko zaključivanje pripadaju li određene književne vrste

definiranim klasama *Poeziji* ili *Prozi* tj. definira književno djelo koji može biti nestihovano ili u obliku stiha (klasifikacija na temelju ovih svojstava je prikazana na Slici 39).

label: imaOblik

Sub Property Of: topObjectProperty

R: Oblik teksta

Svojstvo *imaZnačajke* definira značajke književnih vrsta kao književnoumjetničke ili znanstvene značajke te omogućuje klasifikaciju *Diskurzivnih književnih oblika*

label: imaZnačajke

Sub Property Of: topObjectProperty

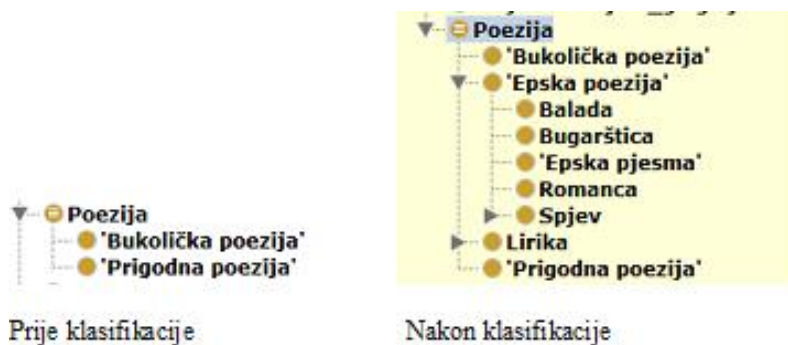
R: Značajke teksta

Svojstvo *jePoVeličini* također omogućuje automatsko zaključivanje jer definira male i velike književne vrste te na taj način zajedno sa svojstvom *imaOblik* omogućuje njihovu klasifikaciju kao npr. podklase definirane klase *Poezija* prikazane na Slici 39.

label: jePoVeličini

Sub Property Of: topObjectProperty

R: Veličina djela



Slika 39: Primjer zaključivanja i klasificiranja entiteta Poezije nakon primjene reasoner-a HermiT 1.3.8. u Protege-u.

Svojstvo *seIzvodi* omogućuje automatsko zaključivanje da je pojedino djelo namijenjeno izvođenju na pozornici i klasifikaciju određenih ontologijskih entiteta kao podklasa definirane klase *Dramski* koja označava pripadnost dramskom rodu (prikaz na Slici 40).

label: seIzvodi

Sub Property Of: topObjectProperty

D: Književni rodovi i vrste

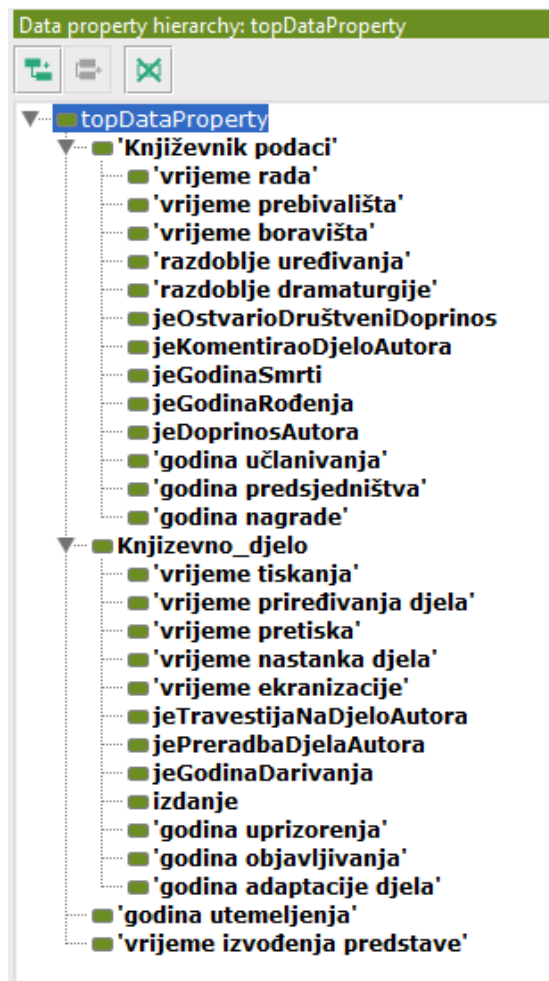
R: Pozornica



Slika 40: primjer klasificiranja knjiž. vrsta Drama nakon primjene reasoner-a HermiT 1.3.8

7.3.5.7 Podatkovni odnosi

Podatkovni odnosi (engl. *Data Properties*) uspostavljaju odnos između individue sa doslovnim vrijednostima (npr. godina, datum, literal itd.). Razvijena je hijerarhija pododnosa, a vrh hijerarhije predstavlja relacija *owl:topDataProperty*. Primjeri ontologijskih podatkovnih odnosa su prikazani na Slici 41, na kojoj je se uočavaju razvijeni podatkovni odnosi za opis književnika (*Književnik podaci*) i književnog djela (*Književno djelo*), *godina utemeljenja* (odnosi se na organizaciju, društvo i sl.) i *vrijeme izvođenja predstave*. Slijedi pojašnjenje svakog podatkovnog odnosa pojedinačno.



Slika 41: Podatkovni odnosi ontologije hrvatskih književnika

Uspostavljeni podatkovni odnosi koji se odnose na književnika su smješteni u podhijerarhiju odnosa *Književnik podaci*:

label: godina nagrade

comment: Svojstvo definira godinu kada je određeni književnik primio nagradu.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: godina predsjedništva

comment: Godina ili razdoblje kada je osoba obavljala službu predsjednika.

D: Person

R: int

Subproperty of: Književnik podaci

label: godina ućlanjivanja

altLabel: godina članstva

comment: Svojstvo definira godinu članstva u pojedinoj ustanovi, društvu, redakciji i sl. određenog književnika.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: jeDoprinosAutora

comment: Svojstvo definira poseban doprinos autora na određenom području (npr. književnost).

D: Književnik

R: Literal

Subproperty of: Književnik podaci

label: jeGodinaRođenja

comment: Relacija povezuje osobu s godinom rođenja.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: jeGodinaSmrti

comment: Relacija povezuje osobu s godinom smrti.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: jeKomentiraoDjeloAutora

comment: Svojstvo definira čije je djelo komentirao određeni književnik, npr. „prokomentirao je Aristotelov traktat De anima“

D: Književnik, Person

R: Literal

Subproperty of: Književnik podaci

label: jeOstvarioDruštveniDoprinos

comment: Svojstvo definira zasluge osobe koje se odnose na društvene promjene ili reforme koje je određena osoba provela u društvu, npr.:

„Kao ban 1873–80. potaknuo je živu reformsku djelatnost kojom su u Hrvatsku uvedeni mnogobrojni moderni instituti, ugl. preuzeti iz austr. zakonodavstva. Tada su uvedeni odgovornost bana (vlade) Saboru, dioba sudstva i uprave te neovisnost sudaca, moderna upravna organizacija, racionalnija sudbena organizacija i racionalniji kazneni postupak, moderniji penološki sustav, sloboda tiska i porotno suđenje, uređeno je pravo na javno okupljanje, donesen je zakon o zavičajnoj pripadnosti, liberaliziran izborni postupak, uvedeni su zametci nekih javnih službi (statistička služba, javno zdravstvo, unaprjeđenje poljoprivrede), laicizirano je i poboljšano pučko školstvo i osnovano Zagrebačko sveučilište“

D: Književnik

R: Literal

Subproperty of: Književnik podaci

label: razdoblje dramaturgije

comment: Relacija utvrđuje godinu ili razdoblje u kojemu je određena osoba bila dramaturg.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: razdoblje uređivanja

comment: Relacija utvrđuje godinu ili razdoblje kada je određeni književnik obavljao uredničke poslove.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: vrijeme boravišta

comment: Svojstvo definira razdoblje boravišta određenog književnika u određenom mjestu.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: vrijeme prebivališta

comment: Svojstvo definira razdoblje prebivališta određenog književnika u određenom mjestu.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: vrijeme rada

comment: Svojstvo definira godinu/razdoblje kada je radio određeni književnik.

D: Književnik

R: int

Subproperty of: Književnik podaci

label: vrijeme članstva

comment: Svojstvo definira razdoblje članstva u određenoj organizaciji, društvu i sl.

D: Književnik

R: int

Subproperty of: Književnik podaci

Uspostavljeni podatkovni odnosi koji se odnose na književno djelo smješteni su u podhijerarhiju podatkovnih odnosa *Književno djelo*:

label: godina adaptacije djela
altLabel: godina prilagodbe djela
comment: Svojstvo definira godinu adaptacije određenog književnog djela.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: godina objavljivanja
altLabel: godina izdanja
comment: Svojstvo definira koje je godine objavljeno određeno književno djelo.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: godina uprizorenja
comment: Svojstvo definira godinu kada je uprizoreno određeno književno djelo.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: izdanje
comment: Svojstvo definira izdanje određenog književnog djela, npr. 2. prošireno izdanje.
D: Književno djelo
R: Literal
Subproperty of: Književno_djelo

label: jeGodinaDarivanja
comment: Relacija opisuje koje godine je određeno djelo darovano nekoj osobi.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: jePreradbaDjelaAutora
comment: Svojstvo definira kojeg autora (djelo od kojega autora: Molièreove komedije) određeno djelo predstavlja preradbu. Vrijednost ima oblik navođenja književnika i djela (Molièreove komedije Tartuffe) te je stoga ovo svojstvo iskazano kao podatkovno svojstvo.
D: Književno djelo
R: Literal
Subproperty of: Književno_djelo

label: jeTravestijaNaDjeloAutora
comment: Svojstvo definira na čijeg autora je travestija određeno književno djelo. Vrijednost ima oblik navođenja književnika i djela (Molièreove komedije Tartuffe) te je stoga ovo svojstvo iskazano kao podatkovno svojstvo.

D: Književno djelo
R: Literal
Subproperty of: Književno_djelo

label: vrijeme ekranizacije
comment: Svojstvo definira godinu kada je određeno književno djelo ekranizirano.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: vrijeme nastanka djela
comment: Svojstvo definira vrijeme kada je nastalo djelo.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: vrijeme pretiska
comment: Svojstvo definira godinu pretiska određenog djela.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

label: vrijeme priređivanja djela
comment: Svojstvo definira godinu ili razdoblje kada je priređeno određeno djelo.
D: Književno djelo
R: int
Subproperty of: Književno_djelo
label: vrijeme tiskanja
comment: Svojstvo definira godinu kada je tiskano određeno književno djelo.
D: Književno djelo
R: int
Subproperty of: Književno_djelo

Uspostavljena su dva podatkovna odnosa koji su predstavljeni kao pododnosi *owl:topDataProperty* jer se ne odnose ni na podatke o književnicima, kao ni na podatke o književnom djelu. Slijedi njihovo pojašnjenje:

label: godina utemeljenja
altLabel: godina osnivanja
comment: Svojstvo definira godinu kada je utemeljenja određena serijska publikacija, biblioteka, organizacija i sl.
D: Biblioteke, Organization, Serijske publikacije
R: int

Subproperty of: topDataProperty

label: vrijeme izvođenja predstave

comment: Svojstvo definira godinu kad se izvodila određena predstava.

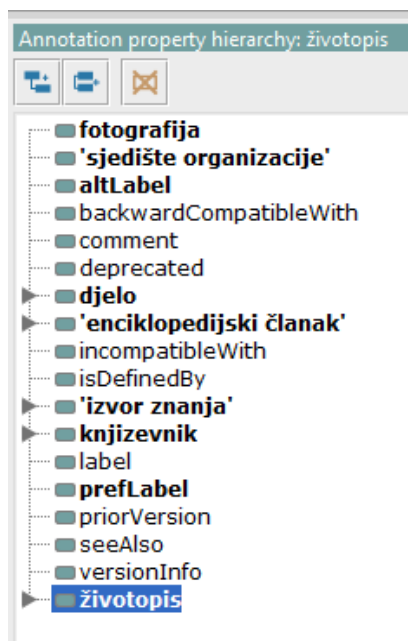
D: Drama

R: int

Subproperty of: topDataProperty

7.3.5.8 Opisna svojstva

Opisna svojstva opisuju individue. Slika 42 prikazuje razvijena opisna svojstva ontologije hrvatskih književnika te hijerarhijsku organizacije pojedinih svojstava (životopis, književnik, djelo, enciklopedijski članak, izvor znanja). Slijedi objašnjene za svako opisno svojstvo.



Slika 42: Opisna svojstva

Nakon kreiranja klasa, individua i relacija potrebno je svima dodati lako čitljiva imena za čovjeka, odnosno doslovne vrijednosti preko *rdfs:label* i *rdfs:comment* relacija. Slika 43 prikazuje opisna svojstva koja se odnose na opis životopisa svakog književnika, stoga je otvorena podhijerarhija svojstva *životopis* koja sadrži podsvojstva *osobni podaci*, *posao*, *školovanje*.



Slika 43: Podhijerarhija svojstava životopis

label: životopis

comment: Relacija opisuje cijeli životopis neke osobe, prikazan je tekstualni dio članka koji donosi informacije o svim osobnim podacima, poslu i školovanju neke osobe.

D: Person

R: Literal

Svojstvo *osobni podaci* je podsvojstvo svojstva životopis te okuplja sve informacije koje govore o osobnim podacima neke osobe:

label: boravište

comment: Svojstvo definira mjesto boravka tj. privremene adrese koje je književnik imao tijekom svog života.

D: Person

R: Mjesto

Superproperties: osobni podaci

label: datum rođenja
comment: Svojstvo definira datum rođenja određenog književnika.
altLabel: birthDate
D: Književnik
R: date
Superproperties: osobni podaci

label: datum smrti
comment: Svojstvo definira datum smrti određenog književnika.
altLabel: deathDate
D: Književnik
R: date
Superproperties: osobni podaci

label: datum zaredenja
comment: Svojstvo definira datum zaredenja određenog književnika.
D: Književnik
R: date
Superproperties: osobni podaci

label: ime
comment: Svojstvo omogućuje imenovanje književnika.
D: Person
R: Literal
Superproperties: osobni podaci

label: ime člana Arkadije
altLabel: pastirsko ime Accademia degli Arcadi
comment: Svojstvo definira ime koje je određeni književnik imao kao član Arkadije.
D: Književnici
R: Literal
Superproperties: osobni podaci

label: potpis
comment: Svojstvo definira koji je potpis koristio određeni književnik.
D: Književnici
R: Literal
Superproperties: osobni podaci

label: mjesto zaredenja
comment: Svojstvo definira koje je mjesto zaredenja određenog književnika.
D: Književnici
R: Mjesto
Superproperties: osobni podaci

label: nadimak
comment: Svojstvo definira koji je nadimak određenog književnika.
D: Književnici
R: Literal
Superproperties: osobni podaci

label: politički stav
comment: Svojstvo definira politički stav određene osobe.
D: Književnici
R: Literal
Superproperties: osobni podaci

label: prebivalište
comment: Svojstvo definira koje je mjesto prebivanja (mjesto stalnog boravka) određenog književnika
D: Književnici
R: Mjesto
Superproperties: osobni podaci

label: pripada tipu obitelji
comment: Svojstvo definira kojem tipu obitelji pripada određeni književnik.
D: Književnici
R: Literal
Superproperties: osobni podaci

label: pseudonim
comment: Svojstvo definira koji je pseudonim određenog književnika.
D: Književnici
R: Literal
Superproperties: osobni podaci

Svojstvo *posao* je podsvojstvo svojstva *životopis* te okuplja sve podatke koje donose informacije o poslovnoj djelatnosti nekog književnika. Njegova podsvojstva su objašnjenja u nastavku:

label: posao
comment: Svojstvo definira koje je poslove obavljala određena osoba.
Navode se poslovi koje je osoba obavljala
D: Književnici
R: Literal
Superproperties: životopis

label: godina umirovljenja
comment: Svojstvo definira godinu umirovljenja određenog književnika.
D: Književnici
R: int
Superproperties: posao

label: mjesto rada
comment: Svojstvo definira koje je mjesto rada određenog književnika.
D: Književnici
R: Mjesto
Superproperties: posao

label: naslov
comment: Svojstvo definira naslov u nekom poslu, npr. rektor Sveučilišta, dekan i sl.
D: Književnici
R: Literal
Superproperties: posao

label: vrijeme predavanja
comment: Svojstvo definira vrijeme (godina, razdoblje) kada određeni književnik predavao.
D: Književnici
R: int
Superproperties: posao

Svojstvo *školovanje* je podsvojstvo svojstva *životopis* te okuplja sve podatke koje donose informacije o obrazovanju nekog književnika. Sadrži određeni broj podsvojstva koji su objašnjeni u nastavku:

label: školovanje
comment: Relacija opisuje školovanje pojedine osobe kroz duži tekst (nekoliko rečenica, odlomak).
D: Književnici
R: Literal
Superproperties: školovanje

label: godina diplomiranja
comment: Svojstvo definira godinu kada je diplomirao određeni književnik.
D: Književnici
R: int
Superproperties: školovanje

label: godina doktorata
comment: Svojstvo definira godinu kada je doktorirao određeni književnik.
D: Književnici

R: int

Superproperties: školovanje

label: godina mature

comment: Svojstvo definira godinu kada je maturirao određeni književnik.

D: Književnici

R: int

Superproperties: školovanje

label: godine školovanja

comment: Svojstvo definira godine školovanja određenog književnika.

D: Književnici

R: int

Superproperties: školovanje

label: mjesto doktorata

comment: Svojstvo definira koje je mjesto doktorata određenog književnika.

D: Književnici

R: Mjesto

Superproperties: školovanje

label: mjesto srednje škole

comment: Svojstvo definira u kojem je mjestu određeni književnik završio srednju školu.

D: Književnici

R: Mjesto

Superproperties: školovanje

label: mjesto studiranja

comment: Svojstvo definira koje je mjesto studiranja određenog književnika.

D: Književnici

R: Mjesto

Superproperties: školovanje

label: mjesto školovanja

comment: Svojstvo definira koje je mjesto školovanja određenog književnika.

D: Književnici

R: Mjesto

Superproperties: školovanje

label: naslov doktorata

comment: Svojstvo definira naslov doktorata koje je napisao i obranio određeni književnik.

R: Literal

Superproperties: školovanje

label: naslov neobjavljenog doktorata
comment: Svojstvo definira naslov doktorata kojega nije objavio određeni književnik.
R: Literal
Superproperties: školovanje

label: područje doktorata
comment: Svojstvo definira područje iz kojega je napisao doktorat određeni književnik.
D: Književnici
R: Područje
Superproperties: školovanje

label: razdoblje studiranja
comment: Svojstvo definira razdoblje studiranja određenog književnika.
D: Književnici,
R: dateTime
Superproperties: školovanje

label: tema doktorata
comment: Svojstvo definira koja je tema doktorata određenog književnika.
D: Književnici
R: Literal
Superproperties: školovanje

label: tema magisterija
comment: Svojstvo definira koja je tema magisterija određenog književnika.
D: Književnici
R: Literal
Superproperties: školovanje

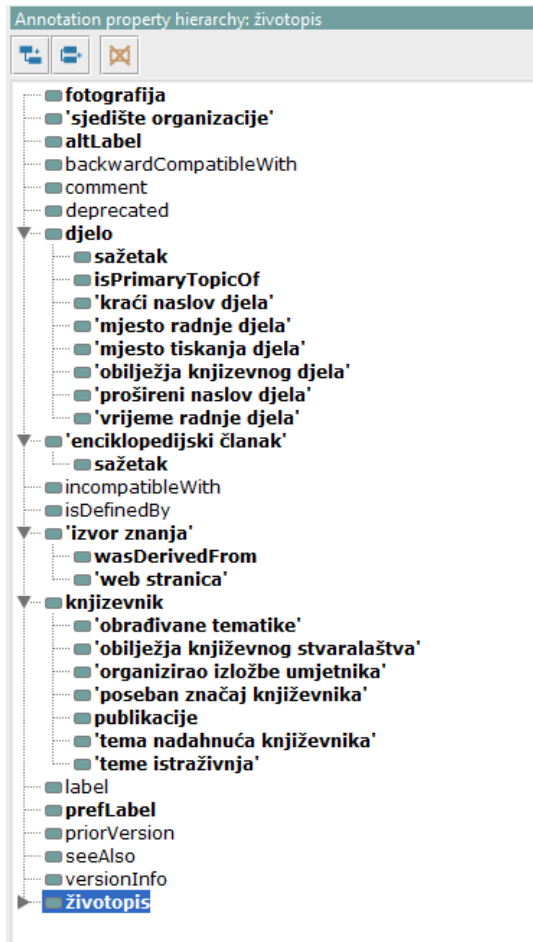
label: titula
comment: Svojstvo definira koji titulu ima određeni književnik.
D: Književnici
R: Literal
Superproperties: školovanje
Superproperties: školovanje

label: ustanova magisterija
comment: Svojstvo definira na kojoj je obrazovnoj ustanovi magistrirao određeni književnik.
D: Književnici
R: Obrazovne ustanove
Superproperties: školovanje

label: ustanova_doktorata
comment: Svojstvo definira na kojoj je obrazovnoj ustanovi doktorirao određeni književnik.

D: Književnici
R: Obrazovne ustanove
Superproperties: školovanje

Slijede opisi ostalih opisnih svojstava koji su prikazani na Slici 44.



Slika 44: Prikaz opisnih svojstava

Načinjeno je opisno svojstvo *knjizevnik* s ciljem da donese podsojstva kojima će se opisati neki književnik, a pojašnjenja njegovih podsojstava slijede u nastavku.

label: obilježja književnog stvaralaštva

comment: Svojstvo donosi opis obilježja književnog stvaralaštva primijenjenog u djelima određenog književnika.

D: Književnici

R: Literal

Superproperties: knjizevnik

label: obrađivane tematike

comment: Relacija definira teme koje je književnik obrađivao u svojim djelima. Za razliku od objektnog svojstva je TemaKnjiževnika, ovo svojstvo definira skup tema koje je književnik obrađivao u svojim djelima.

D: Književnici

R: Literal

Superproperties: književnik

label: organizirao izložbe umjetnika

comment: Svojstvo definira umjetnike čije je izložbe organizirao određeni književnik.

D: Književnici

R: Literal

Superproperties: književnik

label: poseban značaj književnika

comment: Relacija opisuje po čemu je osoba istaknuta tj. značajna, npr. po čemu je ostvarila prvenstvo u svom radu i sl.

D: Književnici

R: Literal

Superproperties: književnik

label: publikacije

comment: Svojstvo definira publikacije u kojima je određeni književnik objavljivao. Za razliku od svojstva je ObjavljivaoUČasopisu koja definira samo časopise u kojima je književnik objavljivao, ovdje se mogu staviti sve vrste publikacija.

D: Književnici

R: InformationResource

Superproperties: književnik

label: tema nadahnjuća književnika

comment: Svojstvo definira koja je tema posebno poticala (nadahnjivala) određenog književnika na pisanje.

D: Književnici

R: Literal

Superproperties: književnik

label: teme istraživanja

comment: Relacija opisuje teme koje je autor istraživao tj. kojima se bavio u svom radu, pisanju i sl.

D: Književnici

R: Literal

Superproperties: književnik

Svojstvo *djelo* je stvoreno kako bi se dodijelila opisna svojstva nekom književnom djelu. Sadrži određeni broj podsvojstva koji su objašnjeni u nastavku:

label: naslov djela

comment: Svojstvo povezuje individu u klase Književno djelo s naslovom te individue

D: Književno djelo

R: Literal

Superproperties: djelo

label: kraći naslov djela

comment: Svojstvo definira kraći naslov određenog djela (publikacije), ukoliko postoji.

D: Književno djelo

R: Literal

Superproperties: djelo

label: alternativni naslovdjela

comment: Relacija definira koje alternativni naslov nekog književnog djela tj. pridruživanje tog alternativnog naslova književnom djelu, npr. Gl. je Mažuranićevo knjiž. djelo Smrt Smail-age Čengića, koje je objavio 1846. u almanahu Iskra pod naslovom Smrt Čengić-age.

D: Književno djelo

R: Literal

Superproperties: djelo

label: mjesto radnje djela

comment: Svojstvo definira koje je mjesto radnje određenog djela.

D: Književno djelo

R: Mjesto

Superproperties: djelo

label: mjesto tiskanja djela

comment: Svojstvo definira koje je mjesto tiskanja određenog književnog djela.

D: Književno djelo

R: Mjesto

Superproperties: djelo

label: obilježja književnog djela

comment: Svojstvo definira obilježja o nekom književnom djelu ukoliko se žele istaknuti.

D: Književno djelo

R: Literal

Superproperties: djelo

label: prošireni naslov djela

comment: Svojstvo definira prošireni naslov nekog djela (publikacije) ukoliko postoji.

D: Književno djelo

R: Literal

Superproperties: djelo

label: sažetak

comment: Svojstvo omogućuje prikaz sažetka nekog djela ili enciklopedijskog članka.

D: Književno djelo, InformationResource

R: Literal

Superproperties: djelo

Superproperties: enciklopedijski članak (svojstvo *enciklopedijski članak* je nadređeno svojstvu podsvojstvu *sažetak* koji ima za cilj donijeti sažetak određenog enciklopedijskog članka)

label: vrijeme radnje djela

comment: Svojstvo definira koje je vrijeme radnje određenog književnog djela.

D: Književno djelo

R: int

Superproperties: djelo

label: isPrimaryTopicOf

comment: Svojstvo definira individu u kao primarnu temu nekog izvora tj. web stranice.

D: Književnici

R: Literal

Superproperties: djelo

Preostala opisna svojstva su:

label: fotografija

comment: Svojstvo uspostavlja link opisivane individue na fotografiju koja je dostupna na webu.

D: Entity

R: anyURI

label: sjedište organizacije

comment: Svojstvo definira sjedište (mjesto) određene organizacije (ustanove, društva itd.).

D: Organization

R: Mjesto

Svojstvo *izvor znanja* opisuje određene izvore znanja povezane s opisivanim individuama. Opisuje se pomoću sljedećih podsvojstava:

label: wasDerivedFrom

comment: Svojstvo definira iz koje su web stranice ekstrahirani određeni strukturirani podaci.

Superproperties: izvor znanja

D: web stranica

R: anyURI

label: web stranica

comment: Svojstvo definira koji je link poveznica s vanjskom web stranicom.

Superproperties: izvor znanja

D: web stranica

R: anyURI

7.3.6 Definiranje ograničenja nad relacijama

Svojstva mogu imati različita ograničenja opisujući vrstu vrijednost, dozvoljene vrijednosti, broj vrijednosti (engl. cardinality – kardinalnost) i druge značajke vrijednosti koje svojstvo može imati. Primjerice, vrijednost opisne relacije *label* (koja definira ime književnika) je *string*. Stoga opisna relacija *label* je relacija s vrstom vrijednosti *string*. Primjerice, relacija *jeAutor* može imati višestruke vrijednosti, a vrijednosti su instance klase *Književno djelo*. Stoga, *jeAutor* je relacija s vrstom vrijednosti instanca *Književno djelo* kao dozvoljenom klasom. Tako primjerice opća ograničenja mogu biti:

- Kardinalnost svojstava (eng. *cardinality restraints*) definira koliko vrijednosti svojstvo može imati. Primjerice, ukoliko se želi iskazati da bilo koji objekt klase *Književnik* mora biti autor najmanje jednog objekta klase *Književno djelo*, to se može učiniti na sljedeći način. Zna se da bilo koje djelo koje je napisao *Književnik* mora pripadati klasi *Književno djelo*. Međutim, ako se želi iskazati *Književnik jeAutor* najmanje jednog *Književnog djela*, to se može izraziti pomoću konstrukata *owl:minQualifiedCardinality* tzv. „najmanje“ ograničenje, jer se pomoću njega može reći da klasa *Književnik* može imati najmanje jedno svojstvo *jeAutor*. Ako se umjesto *owl:minQualifiedCardinality* koristi *owl:maxQualifiedCardinality* („najviše“ ograničenje) onda *Književnik* može napisati najviše jedno *Književno djelo*, a ako se koristi svojstvo *owl:qualifiedCardinality* („točno“ određeno ograničenje) onda *Književnik* može napisati točno jedno *Književno djelo*.
- Boolova svojstva daju mogućnost i korištenja skupa operatora pri kreiranju klasa, kojima se izražava boolova logika, tj. predstavljaju se AND, OR i NOT operatori,

odnosno presjek (*intersectionOf*), unija (*unionOf*) i komplement (*complementOf*). Tako npr. svojstvo unija (*owl:unionOf*) povezuje neku klasu sa listom klasa koje ju opisuju te *owl:unionOf* izjava opisuje klasu koja se sastoji od dvije ili više klasa koje su članovi te klase koju opisujemo. Primjerice, u ontologiji se želi istaknuti kako je klasa *Epika* unija klasa *Epska poezija* i (and) *Proza*.

- *Slot-value type* (Vrijednost vrste) ograničenja opisuje koje vrste vrijednosti se mogu ispuniti u svojstvu. Popis uobičajenih vrsta vrijednosti je *string* koji definira niz podataka (npr. opisna relacija *label* ima vrijednost jednostavan niz ili string), *float* i *int* predstavljaju specifičnije vrste vrijednosti za opis numeričkih vrijednosti (npr. opisna relacija godina objavljivanja ima vrstu vrijednosti integer), *instance* omogućava definiranje odnosa između individua te relacije s vrstom vrijednosti instance trebaju također definirati popis dozvoljenih klasa za koje mogu doći instance. Na primjer, relacija *jeAutor* za klasu *Književnik* može imati instance klasa *Književno djelo*, *Filmski scenarij*, *Televizijski scenarij*, *Scenarij* kao svoje vrste vrijednosti.

7.3.7 Domena i doseg svojstva

Kada se opisuje rječnik koji se koristi za podatke, korisno je istaknuti kako neka relacija može biti predstavljena u odnosu na definiranu klasu, tj. želi se pomoću neke relacije naglasiti kako subjekt unutar trojke u kojoj se koristi ta relacija dolazi iz jedne klase, dok objekt dolazi iz druge. Za to se koristi *rdfs:domain* i *rdfs:range*. U matematičkom smislu, domena (eng. *domain*) neke funkcije je skup vrijednosti za koje je ta funkcija definirana, dok je doseg (eng. *range*) te funkcije skup vrijednosti koje ta funkcija može poprimiti. Npr. funkcija kvadratni korijen kao svoju domenu ima sve pozitivne brojeve (budući u skupu realnih brojeva negativni brojevi nemaju kvadratni korijen), dok su doseg te funkcije svi realni brojevi (budući da postoji i pozitivan i negativan kvadratni korijen). Unutar neke trojke koja koristi P kao predikat, domena je subjekt te trojke dok je doseg objekt. Kada se istakne svojstvo P ima domenu D, zapravo se ističe da kada god se koristi to svojstvo P, subjekt te trojke u kojoj se koristi pripada klasi D. Isto tako, kada se kaže kako svojstvo P ima doseg R,

zapravo se govori da kada god se koristi to svojstvo P, objekt te trojke u kojoj ga se koristi pripada klasi R.³⁷⁰

Svojstva *rdfs:range* i *rdfs:domain* se koriste kako bi se dodala neka ograničenja u rječnik. Ograničenja definiraju koje klase se mogu koristiti za određena svojstva. *rdfs:domain* se koristi da se odredi koja svojstva mogu biti korištena unutar neke klase, tj. da se poveže svojstvo sa resursom koji se opisuje (subjekt trojke). *rdfs:range* svojstvo se koristi da se odredi koje klase neko svojstvo može imati za svoje vrijednosti (objekt trojke). Npr. nekada se *jeAutor* može koristiti pisac *Knjiga*, *Članak*, *Kolumna* itd. Međutim, ukoliko se želi ograničiti svojstvo *jeAutor* samo unutar klase *Knjiga* onda se koristi *rdfs:range*, kako bi se istaknulo da npr. vrijednost svojstva *jeAutor* može biti samo nešto što se nalazi u klasi *Knjiga*. Dakle, dozvoljene klase za relaciju vrijednosti vrste instance zovu se doseg relacije (engl. *range*), npr. klasa *Knjiga* je doseg za *jeAutor* relaciju. Klase na koje je relacija dodana ili klase koje relacija opisuje nazivaju se domena svojstva (engl. *domain*), npr. klasa *Književnik* je domena za *jeAutor* relaciju.

Prilikom određivanja domene i dosega ontologije književnosti poštivana su sljedeća pravila:

- ako popis klasa koji definira doseg ili domenu relacije uključuje klasu i njezine podklase, potrebno je ukloniti podklase,
- ako popis klasa koji definira doseg ili domenu relacije sadrži sve podklase klase A, ali ne samu klasu A, doseg treba sadržati samo klasu A a ne podklase,
- ako popis klasa koji definira doseg ili domenu svojstava sadrži sve ili nekoliko podklasa klase A, potrebno je razmotriti da li bi klasa A činila primjereniju definiciju dosega.

U prikazu ontologije su navedene domene i dosezi za pojedine relacije.

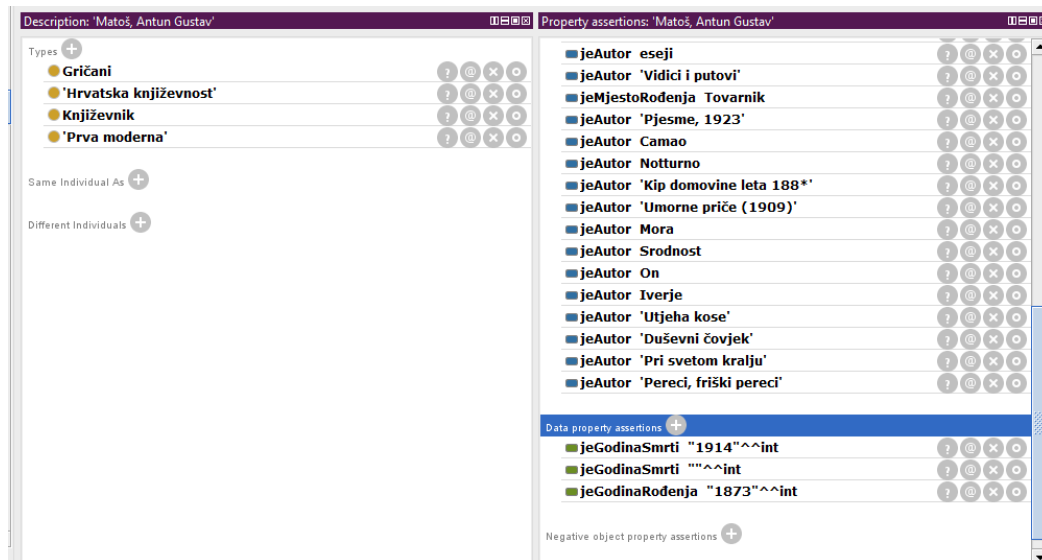
7.3.8 Stvaranje individua

Posljednji postupak prilikom razvoja ontologije književnosti je stvaranje individualnih instanci klasa u hijerarhiji. Definiranje individualne instance klase zahtijeva (1) izbor klase, (2) stvaranje individualne instance te klase i (3) popunjavanje vrijednosti svojstava.

³⁷⁰ Allemang, Dean; Hendler, James A. Semantic web for the working ontologist : modeling in RDF, RDFS and OWL. Amsterdam ; Boston : Morgan Kaufmann Publishers/Elsevier, 2008. Str. 115.

Primjerice, stvorena je individualna instanca *Matoš, Antun Gustav*, prikazana na Slikama 45 i 46. Iz prikazanog (Slika 45) je lako uočiti da je individua *Matoš, Antun Gustav* dodijeljena u klase *Književnik*, *Hrvatska književnost*, *Gričani*, *Prva moderna*. Nadalje, imenovana je preko *rdfs:label* kao *Matoš, Antun Gustav*. Uspostavljeni su objektni odnosi *jeAutor* (s njom se književnik povezuje s književnim djelom koje je napisao), *jePripadnikKnjiževneSkupine*, *jePisaoKritiku*, *jeObjavljivaoUPublikaciji*. Uspostavljeni su objektni odnosi *jeGodinaRođenja* i *jeGodinaSmrti*.

Opisnim relacijama *datum rođenja*, *datum smrti*, *mjesto rođenja*, *mjesto smrti*, *obilježja književnog stvaralaštva*, *školovanje*, *životopis*, *boravište* se opisala navedena individua (Slika 46).



Slika 45: Prikaz objektivnih i podatkovnih relacija individue Antun Gustav Matoš

Individual: 'Matoš, Antun Gustav'

http://www.semanticweb.org/ontologies/Knji.owl#Matos_Antun_Gustav

Annotations (17)

- boravište "Beograd" (hr)
- boravište "Beč" (hr)
- boravište "München" (hr)
- boravište "Pariz" (hr)
- boravište "Zagreb" (hr)
- boravište "Ženeva" (hr)
- 'datum rođenja' "1873-06-13" (date)
- 'datum smrti' "1914-03-17" (date)
- 'mjesto rođenja' "Tovarnik" (hr)
- 'mjesto smrti' "Zagreb" (hr)
- 'obilježja književnog stvaralaštva' "I kao pjesnik Matoš je u hrvatskoj književnosti bio promotorom modernističkih strujanja koja su dolazila iz europskih književnih središta. Poput E. A. Poea, Ch. Baudelaira i pjesnika simbolističkoga pokreta inzistirao je na dotjeranoj formi pa je sonet u njegovu pjesničkom opusu postao kanonskim oblikom. Poetski artizam ogleda se i u strogom poštivanju pravila akcenatsko-silabičke versifikacije. U okviru Matoševe škole velika se pozornost pridavala sinesteziji i zvukovnim (»glazbenim«) kvalitetama izraza, a osobito se njegovala čista i bogata rima." (hr)
- 'obilježja književnog stvaralaštva' "Matoš je i jedan od najznačajnijih hrvatskih kritičara s početka XX. st. Pisao je o hrvatskim (S. S. Kranjčević, V. Vidrić, D. Domjanić, J. Polić Kamov) i o srpskim autorima (S. Srećak, J. Veselinović, S. Pandurović). Njegovao je impresionističku kritiku, vodeći se artističkim, ali i nacionalnim kriterijima. I u književnoj kritici i u žurnalističkim tekstovima nerijetko je ulazio i u političke teme. Matoševi politički nazori, s obzirom na protuaustrijska stajališta i skepsu prema jugoslavenskom unitarizmu, bili su bliski Starčeviću pravaštvu, ali mu je u kulturnom smislu bio stran svaki izolacionizam. Dio njegove esejistike posvećen je istaknutim inozemnim piscima, o kojima je pisao s visokim stupnjem kompetencije pa je hrvatsku sredinu sustavno upoznao s glavnim kretanjima suvremene europske i svjetske literature. Matoš je znatno modernizirao hrvatsku književnost, nametnuo se kao uzor mnogim mladim piscima i uspostavio umjetničke kriterije kao glavni element literarnoga djela te se danas općenito smatra jednim od najboljih hrvatskih književnika" (hr)
- posao "" (hr)
- posao "novinar" (hr)
- školovanje "U Zagrebu, gdje mu se obitelj naselila ubrzo nakon njegova rođenja, pohađao osnovnu školu i gimnaziju. Nakon nedovršene gimnazije propao mu je i pokušaj studija na Vojnome veterinarskom fakultetu u Beču." (hr)
- životopis "1894. dezertirao te morao napustiti Hrvatsku. Otišao je u Beograd, gdje je svirao violončelo u tamošnjim orkestrima i radio kao novinar. Početkom 1898. nakratko je boravio u Münchenu, Beču i Ženevi, a zatim otišao u Pariz, gdje je ostao pet godina. God. 1904. vratio se u Beograd, gdje je ostao do konačnoga povratka u Zagreb 1908., kada je bio pomilovan." (hr)
- label "Matoš, Antun Gustav" (hr)

Slika 46: Prikaz opisnih relacija individue Antun Gustav Matoš

7.4 Rezultati

7.4.1 Prikaz ontologijskih modula s pripadajućim svojstvima

Nakon razvoja ontologije književnosti tj. njezinih temeljnih klasa i ontologijskih odnosa, slijedi tabelarni prikaz temeljnih ontologijskih modula u Tablici 13 tj. klasa sa pripadajućim im svojstvima za svaku od njih u svrhu lakšeg upravljanja ontologijom književnosti prilikom njezine upotrebe.

Tabela 13: Prikaz ontologijskih modula

Modul Agent	
Objektni odnosi	
label: jeMjestoRođenja	label: jeKoristioOblikStrofe
label: jeRodniGrad	label: jeKoristioStih
label: jeMjestoSmrti	label: jeNagrađen
label: jePoNarodnosti	label: jeObjavljiivaoUČasopisu
label: jePodručjeStudiranja	label: jePisaoNaJeziku
label: jeStudirao	label: jePisaoNarječjem
label: jeStudiraoNaUstanovi	label: jePisaoTeme
label: jeUstanovaStudiranja	label: jeTemaKnjiževnika
label: jeStipendist	label: jePisaoKnjiževneVrste
	label: jePodručjePisanja
label: jeRadio	label: jePredavaoPodručje

label: jeSviraoU	label: jePrevodioNaJezik
label: jeDopisnik	label: jePrevodioSaJezika
label: jeDramaturg	label: jePrevođenNaJezike
label: jeFeljtonist	label: jePripadnik
label: jeKolumnist	label: jeKnjizevnikPripadnikKnjiževneGeneracije
label: jeKomentator	label: jeKnjizevnikPripadnikKnjiževneSkupine
label: jeKorektor	label: jeKnjizevnikPripadnikKnjiževneŠkole
label: jeKritičar	label: jeKnjizevnikPripadnikKnjiževnogPokreta
label: jeKazališniKritičar	label: jePripadnikKnjiževnogPravca
label: jeKnjiževniKritičar	label: jePripadnikKnjiževnogRazdoblja
label: jeKustost	label: jeProfesorPredmeta
label: jeLektor	label: jeProučavao
label: jeNovinar	label: jeUtemeljio
label: jePisaoKulturološkeKomentare	
label: jePotpredsjednik	label: jeAutor
label: jePredavaoNaUstanovi	label: autorDjelaJe
label: jeProfesorNaUstanovi	label: jeAutorScenarijaZaFilm
label: jeIzvanredniProfesor	label: djeloJePriredio
label: jeRedovitiProfesor	label: jePriredio
label: jePredsjednik	label: jeDovršio
label: jePrevoditelj	label: jeNapisaoDopunuDjela
label: jePriredivaoIzdanja	label: jeIlustrirao
label: jeSuradnikUstanove	label: jeKomentirao
label: jeTajnik	label: jeParafraziraoDjelo
label: jeUrednikUstanove	label: jePrepjevao
label: jeUpravitelj	label: jePreradio
label: jeČlan	label: djeloJePronašao
label: jeDopisniČlan	label: jePrerađenoOdStraneKnjiževnika
label: jeRedovitiČlan	label: jePrilagodioDjelo
	label: jeAdaptiraoDjelo
label: JeRekonstruiranoPremaKonceptiji	label: jeDramatizirao
label: jeImaoMentora	label: jePrilagodioZaFilm
label: jeMentor	label: jePrilagodioZaScenu
label: jeUzor	label: jePrevedeno
label: jeImaoUzora	label: jePreveo
label: jeIzvršioUtjecaj	label: jePrviPreveo
label: jeUtjecaoNaKnjiževnika	label: jeRekonstruiraoKrajDjela
label: jeNamijenioPosvetu	label: jeRežirao
label: jeParafraziraoKnjiževnika	label: jeSabrao
label: jePrepisivaoDjelaAutora	label: jeScenirao
label: jePrepisivaoZa	label: jeSkladao
label: jePrilagodioDjeloAutora	label: jeSuradnik
label: jePriredioDjelaAutora	label: jeSuautor
label: jePronašaoDjeloKnjiževnika	label: jeUredivaoSa
label: jeUčenik	label: jeUrednik
label: jeUčitelj	label: uredivaoJe
label: seDopisivao	label: jeEnciklopedijskiUrednik
label: jePisaoPoslanice	label: jeFilmskiUrednik
label: "jePreradioDjelaAutora"	label: jeUrednikPoredbeneKnjiževnosti
label: jePrevodioAutora	label: jeUrednikTeorijeKnjiževnosti
label: jePrevođeniAutorOd	label: jeGlavniUrednik
label: jeUglazbioDjeloAutora	label: jePrviUrednik
label: izdanoJeOdIzdavača	label: jeUmjetničkiUrednik
label: jeDaoKraćiNaslovDjela	label: jeUrednikKazališneRubrike
label: jeIzdao	label: jeUrednikKnjiževnosti
label: jeParodija	label: jeUrednikKulturneRubrike
label: jePisaoKnjiževneVrste	label: jeUrednikLlik_Umjetnosti
label: jePrepjevDjelaAutora	
label: jePronašaoDjelo	

label: seIzvodilo	
<i>Podatkovni odnosi</i>	
	label: godina nagrade label: godina predsjedništva label: godina ućlanivanja label: jeDoprinosaAutora label: jeGodinaRoćjenja label: jeGodinaSmrti label: jeKomentiraoDjeloAutora label: jeOstvarioDruštveniDoprinosa label: razdoblje dramaturgije label: razdoblje urećivanja label: vrijeme boravišta label: vrijeme rada label: vrijemeĆlanstva label: vrijeme prebivališta
<i>Opisni odnosi</i>	
label: životopis label: boravište label: datum roćjenja label: datum smrti label: datum zarećjenja label: ime label: ime ćlana Arkadije label: potpis label: mjesto zarećjenja label: nadimak label: politićki stav label: prebivalište label: pripada tipu obitelji label: pseudonim label: posao label: godina umirovljenja label: mjesto rada label: naslov label: vrijeme predavanja	label: školovanje label: godina diplomiranja label: godina doktorata label: godina mature label: godine školovanja label: mjesto doktorata label: mjesto srednje škole label: mjesto studiranja label: mjesto školovanja label: naslov doktorata label: naslov neobjavljenog doktorata label: područje doktorata label: razdoblje studiranja label: tema doktorata label: tema magisterija label: titula label: ustanova magisterija label: ustanova_doktorata label: obilježja književnog stvaralaštva label: obraćivane tematike label: organizirao izlozbe umjetnika label: poseban znaćaj književnika label: publikacije label: tema nadahnjuća književnika label: teme istraživanja
Modul Književno djelo	
<i>Objektni odnosi</i>	

<p>label: djeloJePronašao label: jePronašaoDjelo label: jeDaoKraćiNaslovDjela label: jeDarovano label: jePosvećeno label: jePrepisivaoDjela label: jePrikupljao label: jeSadržanoUOsobnojKnjižnici label: jeUpućeno label: jeAutor label: autorDjelaJe label: izdanoJeOdIzdavača label: jeDovršio label: jeDramatizirao label: jeIlustrirao label: jeIzdao label: jeKomentirao label: jeNapisaoDopunuDjela label: jeParafraziraoDjelo label: jePrepjevao label: jePrepjevDjelaAutora label: jePreradio label: jePrerađenoOdStraneKnjiževnika label: jePreveo label: jePrilagodioDjelo label: jePrilagodioZaFilm label: jePrilagodioZaScenu label: jePripadnik label: jePriredio label: jePrviPreveo label: jeRekonstruiraoKrajDjela label: jeRežirao label: jeSabrao label: jeSkladao label: jeUrednik label: seIzvodilo label: uređivaoJe</p>	<p>label: ElementiKnjiževnogPokreta label: imaOblikStrofe label: imaPohranjenaDjela label: jePohranjeno label: imaPrisutneFigure label: imaStih label: imaStilskoObilježje label: imaTravestiju label: jeTravestijaDjela label: jeDodatak label: jeGrađaZaPisanjeDjela label: jeJezikDjela label: jeKnjiževnaVrsta label: jeNadahnuo label: jeNapisanoPremaDjelu label: jeNarječjeDjela label: jeObjavljeno u Djelu label: objavilaJe label: jeOblikom label: jeOdgovorDjelu label: jeParodija label: jePisanoPismom label: jePismoDjela label: jePrepjev label: jePreradbaDjela label: jePretiskano label: jePretiskanoIzDjela label: jePrevedeno label: jePrevedenoUDjelu label: jePrijevod label: jePripadnikKnjiževnogPravca label: jePripadnikKnjiževnogRazdoblja label: jePrisutanIzričajDjela label: jeSadržajDjela label: sadrži label: jeTemaDjela label: jeTiskanUDjelu label: jeUtjecaloNaDjelo label: suDijelovidjela</p>
<i>Podatkovni odnosi</i>	
<p>label: godina adaptacije djela label: godina objavljivanja label: godina uprizorenja label: izdanje label: jeGodinaDarivanja label: jePreradbaDjelaAutora label: jeTravestijaNaDjeloAutora label: vrijeme ekranizacije label: vrijeme nastanka djela label: vrijeme pretiska label: vrijeme priređivanja djela</p> <p>label: vrijeme tiskanja</p>	
<i>Opisni odnosi</i>	

<p>label: naslov djela label: kraći naslov djela label: alternativni naslovdjela label: mjesto radnje djela label: mjesto tiskanja djela label: obilježja književnog djela label: prošireni naslov djela label: sažetak label: vrijeme radnje djela label: fotografija label: naslov djela label: kraći naslov djela label: alternativni naslovdjela label: mjesto radnje djela label: mjesto tiskanja djela label: obilježja književnog djela label: prošireni naslov djela label: sažetak label: vrijeme radnje djela label: fotografija</p>	
Modul književni rodovi i vrste	
<i>Objektni odnosi</i>	
<p>djeloJePriredio imaPohranjenaDjela imaStih izdanoJeOdIzdavača jeIzdao jeJezikDjela jeKnjiževnaVrsta jeObjavljenoUDjelu jePreveo jePrviPreveo jePrikupljao jePriredio jeSabrao seIzvodi</p>	<p>jePisanoPismom jePisaoKnjiževneVrste jePismoDjela jePohranjeno jePosvećeno jePrepisivaoDjela jePretiskano jePrevedeno jeSadržanoUosobnojKnjižnici jeSkladao jeUrednik jeUtjecaloNaDjelo objavilaJe uredivaoJe</p>
Modul Periodizacija književnosti	
<i>Objektni odnosi</i>	
<p>jePripadnik jeKnjiževnikPripadnikKnjiževneGeneracija jeKnjiževnikPripadnikKnjiževneSkupine jeKnjiževnikPripadnikKnjiževneŠkole jeKnjiževnikPripadnikKnjiževnogPokreta jePripadnikKnjiževnogPravca jePripadnikKnjiževnogRazdoblja</p>	
Modul Dokument	

<i>Objektni odnosi</i>
jeOblikom
<i>Podatkovni odnosi</i>
izdanje

7.4.2 Enciklopedijski članak i ontologija

Nakon dobivenih ontologijskih klasa (pojmovi) i svojstava sama struktura i slijed informacija biografskog enciklopedijskog članka se može prikazati kao skup ontologijskih odnosa poštujući temeljnije dijelove tj. strukturu enciklopedijskog članka (glavu i tijelo članka):

- Glava članka

Natuknica članka: *label, altlabel, ime, nadimak, ime člana Arkadije, talijanizirano ime, pseudonim*

- Tijelo članka

Definicija članka: *jePoNarodnosti, jeMjestoRođenja (jeRodniGrad), datumRođenja, jeMjestoSmrti, datumSmrti, jeGodinaRođenja i jeGodinaSmrti*

Tekst članka: U tekstu članka mogu biti navedene činjenice o obiteljskim ili drugim odnosima između književnika i neke osobe, koji su opisani npr. svojstvima poput *jeKći, jeNećak, jeSestra, jeSin, jeSupruga, jeUnuk, jePrijatelj, jeKum, jeUčenik, seDopisivao*. Slijede činjenice iz životopisa koje mogu biti opisane ontologijskim svojstvom *životopis* koji može obuhvatiti veći dio enciklopedijskog teksta koji se odnosi na obrazovanje, mjesta življenja i djelovanja književnika. Navedeni veći dio teksta koji je obuhvaćen svojstvom *životopis* može se razlomiti na manje cjeline tako što ćemo ga opisati njegovim ontologijskim podsvojstvima *školovanje, boravište, vrijeme boravišta, prebivalište, vrijeme prebivališta*. Ontologijsko svojstvo *školovanje* se pak može još detaljnije opisati primjenom njegovih podsvojstava *mjesto srednje škole, godina mature, jePodručjeStudiranja, jeStudirao, jeStudiraoNaUstanovi, jeUstanovaStudiranja, fakultet, jeImaoMentora, jeMentor, ustanova magisterija, ustanova doktorata, godina diplomiranja, godina doktorata, mjesto doktorata, naslov*

doktorata, naslov neobjavljenog doktorata, mjesto studiranja, mjesto školovanja, godine školovanja, područje doktorata, razdoblje studiranja, tema doktorata, tema magisterija, titula. Upravo ovakva mogućnost izbora ontologijskih svojstava pruža mogućnost izbora i primjene različitih razina opisa znanja pohranjenog unutar članka čime se osigurava fleksibilnost ontologije, kao i prilagodba određenim vrstama korisnika te njihovim informacijskim potrebama. U tekstu članka se također mogu pronaći informacije koje omogućuju ekstrakciju svojstava: *pripada tipu obitelji, politički stav.* Nakon toga slijede svojstva koja opisuju poslove koje je književnik tijekom života obavljao. Opširniji opis posla koje je književnik obavljao se može izvršiti upotrebom ontologijskog svojstva *posao* te kao vrijednost navesti dio enciklopedijskog teksta koji u cijelosti donosi opis svih poslova i djelatnosti književnika. Detaljniji opis djelatnosti književnika može se provesti ontologijskim svojstvima: *jeRadio, jeDopisnik, jeDramaturg, jeEnciklopedijskiUrednik, jeFeljtonist, jeFilmskiUrednik, jeGlavniUrednik, jeIzvanredniProfesor, jeKazališniKritičar, jeKnjiževniKritičar, jeKolumnist, jeKomentator, jeKorektor, jeKritičar, jeKustost, jeLektor, jeNovinar, jePotpredsjednik, jePisaoKulturološkeKomentare, jePrevoditelj, jePredavaoNaUstanovi, jePredsjednik, godina predsjedništva, jePrevoditelj, jePriredivaoIzdanja, jeProfesorNaUstanovi, jeRedovitiProfesor, jeSuradnik, jeTajnik, jeUmjetničkiUrednik, jeUpravitelj, jeUrednik, jePrviUrednik, jeUrednikUstanove, jeUrednikKazališneRubrike, jeUrednikKnjiževnosti, jeUrednikKulturneRubrike, jeUrednikLlik_Umjetnosti, jeUrednikPoredbeneKnjiževnosti, jeUrednikTeorijeKnjiževnosti, jeUčitelj, jeIzvanredniProfesor, jeRedovitiProfesor, jeSuradnikUstanove, jeSviraoU, naslov funkcije, organizirao izložbe umjetnika, datum zaređenja, mjesto zaređenja, godina umirovljenja, mjesto rada, razdoblje dramaturgije, razdoblje uređivanja, vrijeme rada, vrijeme predavanja, godina utemeljenja.* U tekstu se navodi i članstvo (djelovanje) u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama, što je opisano ontologijskim svojstvima *jeČlan, jeČlanRedakcije, jeDopisniČlan, jeRedovitiČlan, godina ućlanivanja, vrijemeČlanstva.* Također se navode prestižne nagrade i priznanja, što je opisano ontologijskim svojstvima: *jeNagrađen, godina nagrade, jeStipendist.* Djela se navode u reprezentativnom izboru, a ekstrahirana ontologijska svojstva opisuju informacije vezane uz djela: *naslov djela, alternativni naslov djela, publikacije, djeloJePriredio, djeloJePronasao, ElementiKnjiževnogPokreta, JeRekonstruiranoPremaKonceptiji, imaOblikStrofe, imaPohranjenaDjela, imaPrisutneFigure, imaStih,*

imaStilskoObilježje, imaTravestiju, jeDaoKraćiNaslovDjela, jeDarovano, jeDodatak, jeGrađaZaPisanjeDjela, jeJezikDjela, jeKnjiževnaVrsta, jeNadahnuo, jeNapisanoPremaDjelu, jeNarječjeDjela, jeObjavljeno u Djelu, jeOblikom, jeOdgovorDjelu, jeParodija, jePisanoPismom, jePismoDjela, jePohranjeno, jePosvećeno, jePrepisivaoDjela, jePrikupljao, jePronašaoDjelo, jePrepjev, jePreradbaDjela, jePrerađenoOdStraneKnjiževnika, jePretiskano, jePretiskanoIzDjela, jePrevedeno, jePrevedenoUDjelu, jePrijevod, jePripadnikKnjiževnogPravca, jePripadnikKnjiževnogRazdoblja, jePrisutanIzričajDjela, jeSadržajDjela, jeSadržanoUOsobnojKnjižnici, jeTemaDjela, jeTiskanUDjelu, jeTravestijaDjela, jeUpućeno, jeUtjecaloNaDjelo, sadrži, objavilaJe, godina objavljivanja, godina adaptacije djela, godina uprizorenja, vrijeme izvođenja predstave, seIzvodilo, suDijeloviDjela, izdanje, jeGodinaDarivanja, jePreradbaDjelaAutora, jeTravestijaNaDjeloAutora, izdanoJeOdIzdavača, jeIzdao, vrijeme ekranizacije, vrijeme pretiska, vrijeme tiskanja, vrijeme nastanka djela, vrijeme priređivanja djela, kraći naslov djela, mjesto radnje djela, mjesto tiskanja djela, obilježja književnog djela, prošireni naslov djela, vrijeme radnje djela. U ovom dijelu teksta također slijede i svojstva autorstva koja nisu navedena u prethodnim dijelovima članka jeAutor , autorDjelaJe, jeSuautor, jeUređivaoSa, jeAutorScenarijazaFilm, koristio potpis , jeDovršio, jeNpisaoDopunuDjela, jeIlustrirao, jeKomentirao, jeParafraziraoDjelo, jePrepjevao , jePrepjevDjelaAutora, jePreradio , jePreveo, jePrviPreveo, jePrilagodioDjelo, jeDramatizirao, jePrilagodioZaFilm, jePrilagodioZaScenu, jePriredio , jeRekonstruiraoKrajDjela, jeRežirao , jeSabrao , jeScenirao , jeSkladao , jeSuradnik , jeUrednik, uređivaoJe. U tekstu se također mogu pronaći informacije o djelima ili književnikovom stvaralaštvu upućujući na odnos između dvije osobe (tj. dva književnika), koje su opisane ontologijskim svojstvima jeBiografOsobe, jeImaoUzora, jeIzvršioUtjecaj, jeUglazbioDjeloAutora, jePisaoODjelimaAutora , jePisaoPoslanice, jePrevodioAutora , jePrevođeniAutorOd ,jeBibliografskiObradio, jeKomentiraoDjelaAutora, jeNamijenioPosvetu, jeParafraziraoKnjiževnika, jePrepisivaoDjelaAutora, jePrepisivaoZa, jePreradioDjelaAutora, jePrevodioDjelaAutora, jePrilagodioDjeloAutora, jePriredioDjelaAutora, JePisaoKritiku, jePronašaoDjeloKnjiževnika, jeUtjecaoNaKnjiževnika, jeUzor. Tekst članaka navodi i obilježja književnog stvaralaštva književnika, što je opisano ontologijskim svojstvima: jeKoristioOblikStrofe, jeKoristioStih,

jePisaoKnjiževneVrste, jeObjavljivaoUČasopisu, jePisaoNaJeziku, jePisaoNarječjem, jePisaoTeme, jePodručjePisanja, jePredavaoPodručje, jePrevođenNaJezike, jePrevodioNaJezik, jePrevodioSaJezika, jeProfesorPredmeta, jeProučavao, jeTemaKnjiževnika, jeUtemeljio, obrađivane tematike, obilježja književnog stvaralaštva, poseban značaj književnika, tema nadahnuća književnika, teme istraživnja, jeDoprinosAutora, jeOstvarioDruštveniDoprinos.

Također su ekstrahirana i svojstva: *sažetak, wasDerivedFrom* (kako bi se navela stranica iz koje su ekstrahirani metapodaci za opis individue), *web stranica, fotografija, isPrimaryTopicOf* (individua je primarna tema izvora tj. web stranice).

Očito je kako se svi dobiveni odnosi neće primijeniti u svim člancima te će se javljati s obzirom na potrebe opisa autora. Međutim, ipak postoji onaj minimalni skup informacija koji se javlja u svakom enciklopedijskom članku, a koji obuhvaća sljedeće ontologijske odnose:

- Glava članka

Natuknica članka: *label, altlabel, ime*

- Tijelo članka

Definicija članka: *jePoNarodnosti, jeMjestoRođenja, datumRođenja, jeMjestoSmrti, datumSmrti, jeGodinaRođenja i jeGodinaSmrti*

Tekst članka: *životopis, boravište, školovanje, posao, jeRadio, naslov djela*

Stoga se svakako predlaže sagledavanje navedenih obaveznih elementa članka prikazanih u vidu ontologijskih svojstava kao mogućih gradivnih svojstava za izgradnju infookvira književnih biografskih enciklopedijskih članaka Hrvatske enciklopedije. Preporučuje se i uvođenje svojstva u infookvir kao bitnih za isticanje i opis posebnosti određenog književnika, iako se radi o neobveznim elementima članka: *pseudonim, jePisaoKnjiževneVrste, jeNagrađen, jeTema književnika.*

7.5 Primjena ontologije

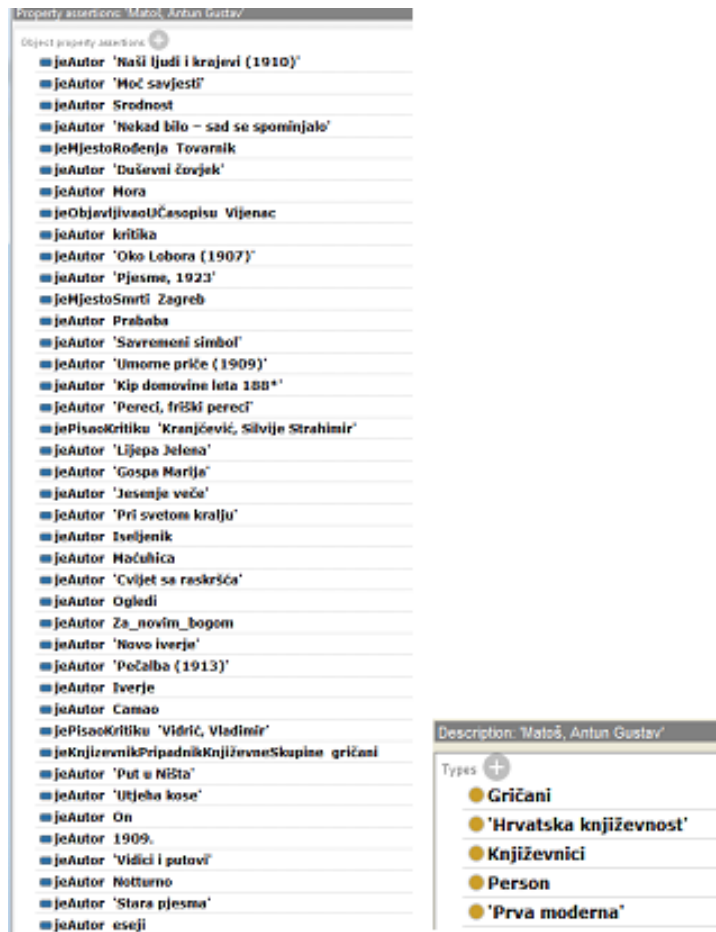
7.5.1 Primjena ontologije hrvatske književnosti za različite razine pretraživanja

Dobivena ontologija iz tekstova enciklopedijskih biografskih članaka hrvatske književnosti omogućuje opis pojedinih individua tj. članaka kroz strukturirane podatke do različitih razina detaljnosti. Ukoliko se razmotri Slika 46 u odlomku Stvaranje individua uočljive su različite razine opisa tj. individui *Matoš*, *Antun Gustav* se može dodijeliti samo opisno svojstvo *životopis* koji će prilikom pretraživanja ponuditi veći dio teksta enciklopedijskog biografskog članka dodijeljen navedenom svojstvu, a koji se odnosi na opisne informacije o životopisu individue te će korisnik dobiti širi uvid u život književnika. Ukoliko korisnik poželi ciljanu i kratke odgovore, može umjesto svojstva *životopis* odabrati svojstvo *posao* (saznat će kako je književnik bio npr. novinar tj. koje je poslove tijekom svog života književnik konkretno obavljao) ili npr. svojstvo *boravište* kako bi saznao u kojim je sve gradovima književnik boravio tijekom života. Navedeno je rezultat stvorene ontologije književnosti koja se primjenjuje na opis stvorene individue. Dakle, ne pružaju istu razinu opisa sva ontologijska svojstva, nego određena svojstva pružaju opširniju i općenitiju razinu opisa, kao npr. svojstva *životopis*, *školovanje*, *obilježja književnog stvaralaštva*; dok preciznu i kratku razinu opisa pružaju ontologijska svojstva kao što su npr. *posao*, *boravište*, *naslov doktorata i sl.*

Nadalje, Slika 47 ispod prikazuje klase kojima individua A. G. Matoš pripada (*Gričani*, *Hrvatska književnost*, *Književnici*, *Person*, *Prva moderna*). Uočavaju se sljedeći odnosi:

- odnos *jeAutor*, koje je dodijeljeno naslovima Matoševih djela i time omogućuje prilikom pretraživanja na postavljeni upit „Koja je djela napisao A. G. Matoš?“ dobivanje preciznog odgovora svih djela navedenih u enciklopedijskom članku. Navedeno svojstvo također povezuje književnika s književnim vrstama, stoga je moguće prilikom pretraživanja na postavljeni upit „Koje je književne vrste pisao A. G. Matoš?“ dobiti kao odgovor kritike, ogledi, eseji.
- odnos *jeMjestoRođenja* povezuje književnika s mjestom u kojem je rođen te time omogućuje prilikom pretraživanja na postavljeni upit „U kojem je mjestu rođen književnik A. G. Matoš?“ dobivanje preciznog odgovora Tovarnik. Isto vrijedi i za mjesto smrti tj. odnos *jeMjestoSmrti*.

- Odnos *jeObjavljiavaoUČasopisu* povezuje književnika s časopisima te time omogućuje prilikom pretraživanja na postavljeni upit „U kojem je časopisu objavljivaio književnik A. G. Matoš?“ dobivanje preciznog odgovora npr. časopis Vjenac.
- Odnos *jePisaoKritiku* povezuje književnika s osobama o kojima je pisao kritiku te time omogućuje prilikom pretraživanja na postavljeni upit „O kojim je književnicima A. G. Matoš pisao kritiku?“ dobivanje preciznog odgovora Kranjčević, Silvije Strahimir i Vidrić, Vladimir

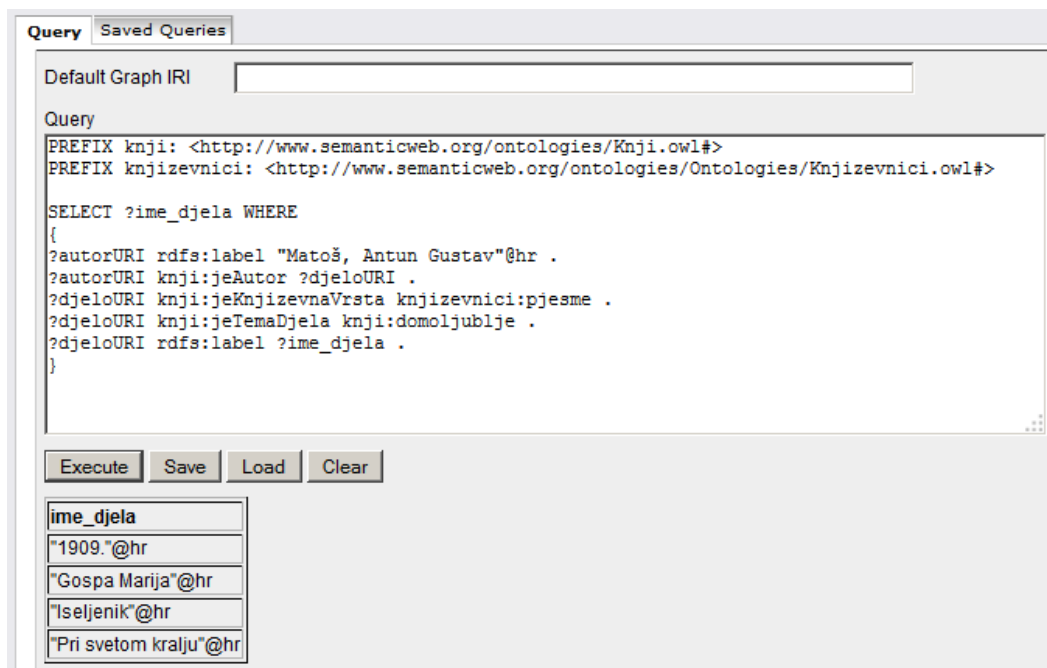


Slika 47: Svojtstvo jeAutor povezuje individuu A. G. Matoš sa svim njegovim djelima i prikazuje njegovu pripadnost pojedinim klasama.

Osim dobivenih prethodno postavljenih jednostavnih upita, razvijena ontologija književnosti također omogućuje i postavljanje složenih upita te dobivanje preciznih i smislenih odgovora, što do sada nije bilo moguće u trenutno dostupnoj mrežnoj online elektroničkoj enciklopediji. Slike 48-54 prikazuju u gornjem djelu prozora postavljeni upit

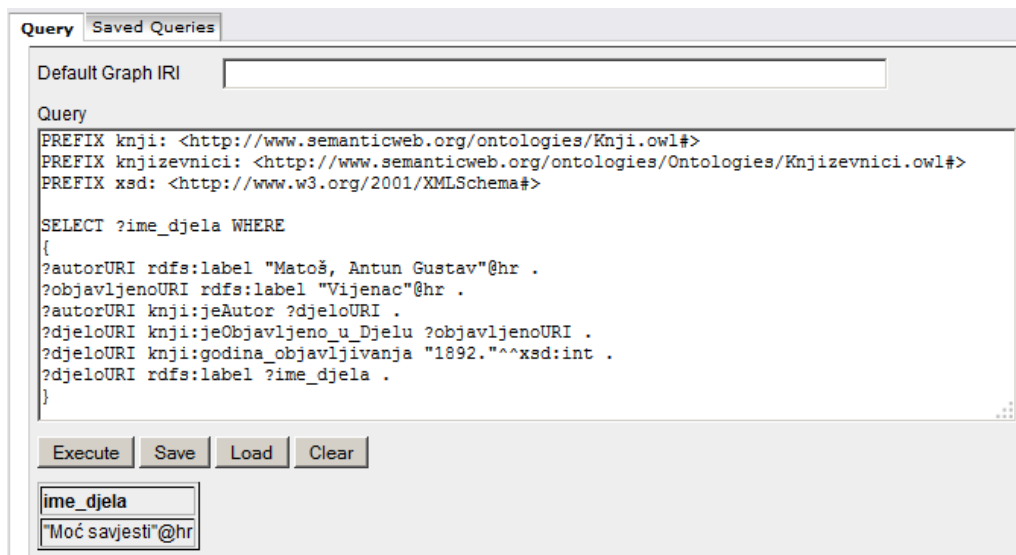
SPARQL jezikom, a u donjem dijelu prozora dobivamo precizne odgovore. Prikazani primjeri ukazuju kako je ovaj odjeljak ilustrativno prikazao primjenu razvijene ontologije književnosti prilikom opisa individue *Matoš, Antun Gustav* te je ukazalo na mogućnost pretraživanja i dobivanja točnih odgovora tj. informacija na postavljanje jednostavne i složene upite.

Primjer 1: Na Slici 48 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Koje je pjesme domoljubne tematike napisao Antun Gustav Matoš?“, a u donjem dijelu prozora je dobiven kao odgovor djela: 1909, Gospa Marija, Iseljenik i Pri svetom kralju.



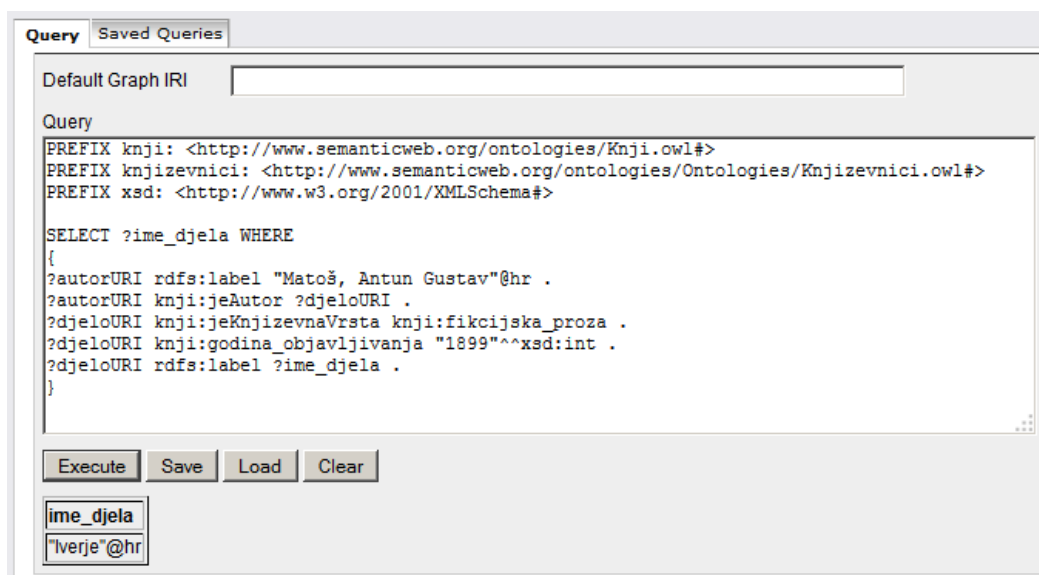
Slika 48: Odgovor na složeni upit „Koje je pjesme domoljubne tematike napisao Antun Gustav Matoš?“

Primjer 2: Na Slici 49 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Koje je djelo Antuna Gustava Matoša objavljeno u Vijencu 1892?“, a u donjem dijelu prozora je dobiven kao odgovor djelo *Moć savjest*.



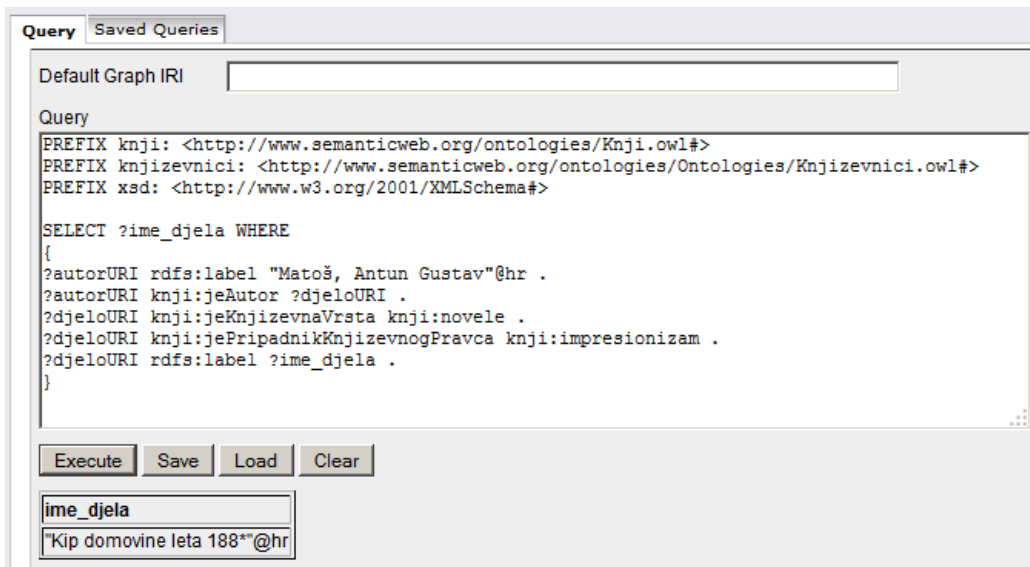
Slika 49: Odgovor na složeni upit „Koje je djelo Antuna Gustava Matoša objavljeno u Vijencu 1892?“

Primjer 3 Na Slici 50 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Koji je naslov fikcijske proze Antun Gustav Matoš objavio 1899?“, a u donjem dijelu prozora je dobiven kao odgovor djelo Iverje.



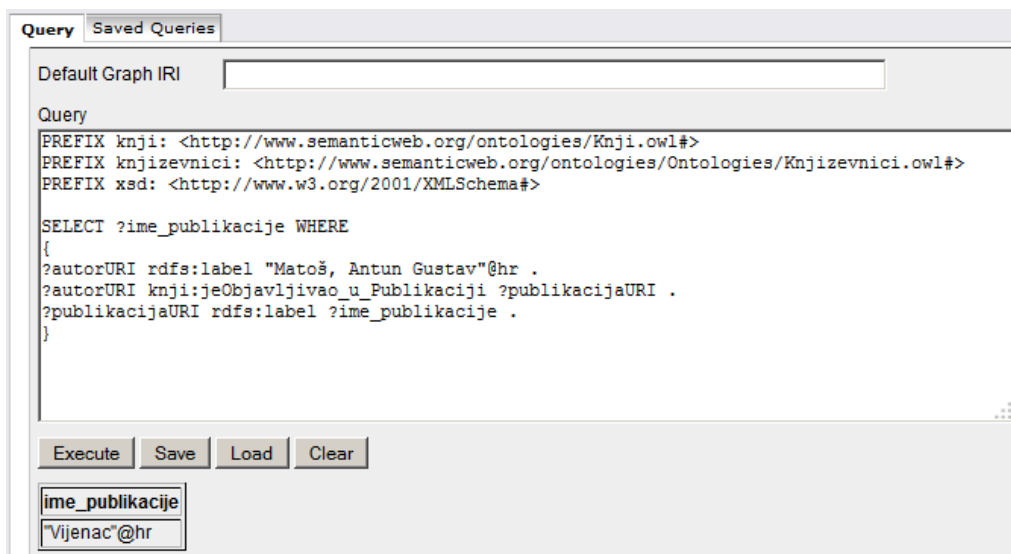
Slika 50: Odgovor na složeni upit „Koji je naslov fikcijske proze Antun Gustav Matoš objavio 1899?“

Primjer 4: Na Slici 51 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Kuju je impresionističku novelu objavio Antun Gustav Matoš?“, a u donjem dijelu prozora je dobiven kao odgovor djelo Kip domovine leta 188*“



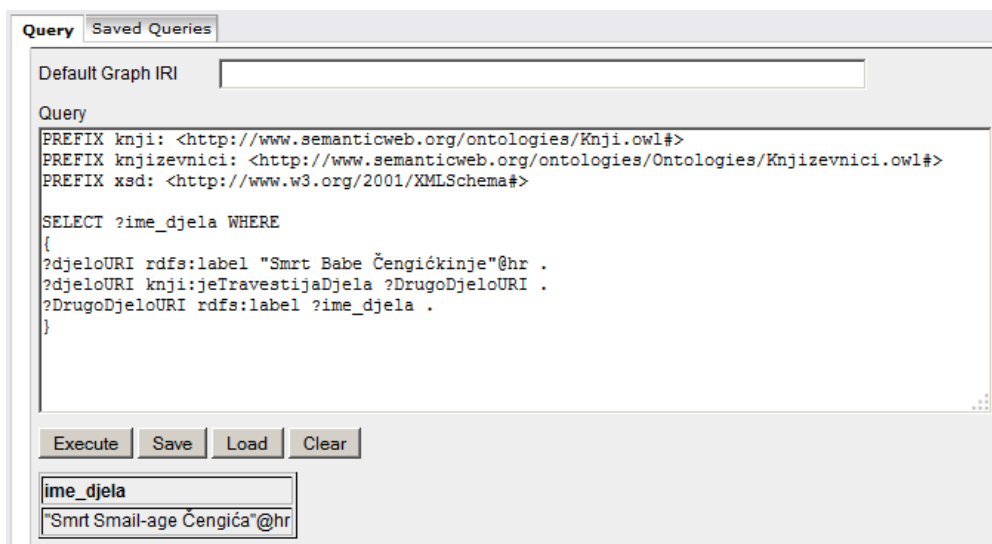
Slika 51: Odgovor na složeni upit „Kuju je impresionističku novelu objavio Antun Gustav Matoš?“

Primjer 5: Na Slici 52 je u gornjem djelu prozora je zadan SPARQL jezikom upit „U kojim je sve publikacijama objavljivao A. G. Matoš?“, a u donjem dijelu prozora je dobiven kao odgovor Vijenac.



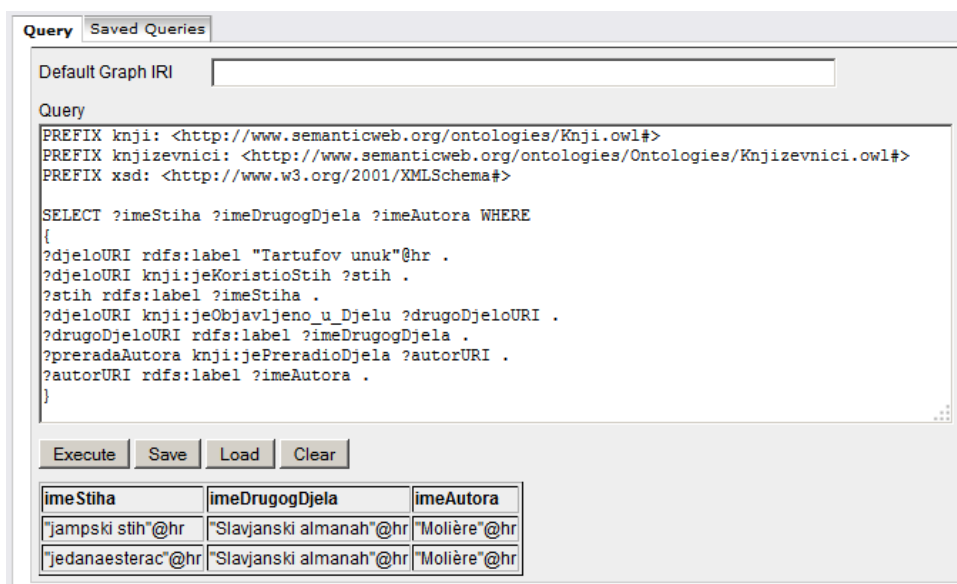
Slika 52: Odgovor na složeni upit „U kojim je sve publikacijama objavljivao A. G. Matoš?“

Primjer 6: Na Slici 53 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Na koje je djelo travestija djelo Smrt Babe Čengićkinje?“, a u donjem dijelu prozora je dobiven kao odgovor Smrt Smail-age Čengića.



Slika 53: Odgovor na složeni upit „Na koje je djelo travestija djelo Smrt Babe Čengićkinje?“

Primjer 7: Na Slici 54 je u gornjem djelu prozora je zadan SPARQL jezikom upit „Koji je stih u djelu Tartufov unuk? U kojoj je publikaciji objavljeno djelo Tartufov unuk? Kojeg je autora preradio djelo Tartufov unuk?“, a kao odgovor u donjem dijelu prozora je dobiven: jampski stih i jedanaesterac je u djelu Tartufov unuk, objavljenom u publikaciji Slavjanski almanah, a djelo je preradilo autora Molierea.



Slika 54: Odgovor na složeni upit postavljen u Primjeru 7.

7.5.2 Primjena ontologije hrvatske književnosti za pregledavanje informacija

Osim smislenog pretraživanja informacija moguće je i pregledavanje informacija s obzirom na određeni broj postavljenih kriterija. Sljedeći primjeri će prikazati upotrebu ontologije prilikom pregledavanja informacija prema odabranim kriterijima. Primjeri će navedeno demonstrirati kroz pretraživač koji je generiran na temelju SPARQL upita, kao i na direktno postavljenim SPARQL upitima. Navedeni pretraživač generiran na SPARQL upitima ukazuje na mogućnost njegove primjene za korisnike koji ne moraju znati SPARQL jezik. Upotrijebljene su sljedeće relacije za sljedeće primjere: *jeAutor*, *jeMjestoRođenja*, *jeČlan*, *jePredsjednik*, *jeUrednik*, *jePisaoNaJeziku*, *jeObjavljivaoUČasopisu*, *jeNagrađen*, *jeKnjiževnaVrsta*, *jeJezikDjela*, *godina objavljivanja*, *jeRazdobljeObjavljivanja*. Upotrijebljene su klase: *Mjesto*, *Književne vrste*, *Jezik*, *Časopisi*, *Nagrada*, *Organizacija*, *Razdoblje*. Navedene relacije, klase i konkretno navođenje njihovih pojedinih individua omogućit će postavljanje upita o dubrovačkim i splitskim književnicima, kao i njihovim književnim djelima s obzirom na postavljene kriterije.

Primjer 1: Koji su dubrovački književnici bili članovi Accademia dell'Arcadia?

Slika 55 prikazuje izbor postavljanja potrebnih kriterija. Vidljivo je kako se treba pronaći Književnik kojega određuju relacije *je rođen u* i *je član*. Za relaciju *je rođen u* odabran je grad Dubrovnik, a za relaciju *je član* odabrana je Accademia dell'Arcadia. Ukoliko bi se trebao dodati još neki kriterij, odnosno relacija, odabere se znak + i izvrši se izbor odgovarajuće relacije. Znak X služi za brisanje odabrane relacije. Odabirom gumba *Pretraži* dobivaju su odgovarajući rezultati koji mogu biti čisti podatak, ali i podatak u obliku poveznice na odgovarajući članak.

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Pretraži

- Accademia dell' Arcadia
- Accademia dell' Arcadia
- Akademija ispraznijeh
- Akademija složnih
- HAZU

Slika 55: Izbor potrebnih kriterija pregledavanja

Slike 56 i 57 prikazuju dobivene rezultate na postavljeni upit. Na Slici 56 se ne vidi SPARQL upit, dok se na Slici 57 vidi SPARQL upit. Prikazivanje, odnosno ne prikazivanje, SPARQL upita se postiže odabirom gumba *Prikaži/Sakrij generirani SPARQL*.

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Pretraži

Rezultati pretraživanja:

Knjizevnik
Albertini, Beninj
Andrović, Rafo
Galjuf, Marko Faustin
Kiša, Tomo
Kunić, Rajmund
Sorkočević, Miho
Stulli, Luko
Zlatarić, Marin

Slika 56: Rezultati pretraživanja

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Pretraži

Rezultati pretraživanja:

SPARQL:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/K

SELECT STR(?knjizevnik) as ?knjizevnik WHERE {
?var1 rdfs:label "Dubrovnik"@hr .
?var2 rdfs:label "Accademia dell' Arcadia"@hr .
?individua Knji:jeMjestoRodjenja ?var1 ;
    Knji:jeClan ?var2 ;
    rdfs:label ?knjizevnik .
}
```

Slika 57: Rezultati pretraživanja sa SPARQL

Na isti način se može umjesto Accademia dell'Arcadia odabrati Akademija složnih ili Akademija ispraznijeh (Slika 55), za što će se dobiti odgovarajući odgovori.

Primjeri 2, 3 i 4 prikazuju smanjivanje dobivenih rezultata pregledavanja povećanjem broja kriterija, odnosno uvođenjem većeg broja jezika na kojima su pisali dubrovački književnici. Slijedi pojedinačno objašnjenje svakog pojedinog primjera.

Primjer 2: Koji su dubrovčani pisali latinskim jezikom?

Slika 58 ukazuje da je potrebno pronaći književnike koji zadovoljavaju kriterije, odnosno relacije, *je rođen u Dubrovniku* i *je pisao na jeziku latinski*. Rezultati navedenih kriterija su vidljivi na Slikama 58 i 89. Na Slici se uočava da nije izabrana vidljivost generiranog SPARQL upita.

Pretraživanje baze:

Pronađi Književnik gdje:

Književnik je rođen u Dubrovnik

Književnik je pisao na jeziku latinski

Pretraži + x

Rezultati pretraživanja:

Prikaži/Sakrij generirani SPARQL

Knjizevnik
Albertini, Beninj
Aletin, Petar Frano
Andelić, Ivan Karlo
Andelić, Kristo
Andrović, Nikola
Andrović, Rafo

Slika 58: Rezultati za upit „Koji su dubrovčani pisali latinskim jezikom?“

Bašić, Đuro
Beneša, Damjan
Beneša, Matija
Betondić, Jakov
Bolić, Vlaho
Galjuf, Marko Faustin
Kriša, Tomo
Kunić, Rajmund
Petrović, Vice
Sorkočević, Miho
Stulli, Luko
Zlatarić, Marin
Lukarević, Ivan
Palmotić, Junije
Pucić Soltanović, Vice

Slika 59: Rezultati za upit „Koji su dubrovčani pisali latinskim jezikom?“ (nastavak)

Primjer 3: Koji su dubrovčani pisali grčkim i latinskim jezikom?

Slika 60 ukazuje da je potrebno pronaći književnike koji zadovoljavaju kriterije, odnosno relacije, *je rođen u Dubrovniku* i *je pisao na jeziku grčki i latinski*. Na istoj slici je vidljiv i rezultat navedenih kriterija. Uočava se da nije izabrana vidljivost generiranog SPARQL upita.

The screenshot shows a search interface with the following elements:

- Search bar: "Pronađi Književnik" with a dropdown arrow and "gdje:" label.
- Criteria 1: "Književnik je rođen u" dropdown with "Dubrovnik" selected.
- Criteria 2: "Književnik je pisao na jeziku" dropdown with "grčki" selected.
- Criteria 3: "Književnik je pisao na jeziku" dropdown with "latinski" selected.
- Buttons: "Pretraži", "+", and "x".
- Section: "Rezultati pretraživanja:"
- Toggle: "Prikaži/Sakrij generirani SPARQL".
- Table of results:

Knjizevnik
Aletin, Petar Frano
Andrović, Nikola
Beneša, Damjan

Slika 60: Rezultati za upit „Koji su dubrovčani pisali grčkim i latinskim jezikom?“

Primjer 4: Koji su dubrovčani pisali grčkim, latinskim i talijanskim jezikom?

SPARQL upit ukazuje kako je potrebno pronaći književnike koji zadovoljavaju kriterije, odnosno relacije, *je rođen u Dubrovniku* i *je pisao na jeziku grčki, latinski i talijanski*. U rezultatu je samo jedan odabrani književnik Andrović, Nikola.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?knjizevnik) as ?knjizevnik WHERE
{
  ?grad rdfs:label "Dubrovnik"@hr .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jePisaoNaJeziku Knji:grcki ;
    Knji:jePisaoNaJeziku Knji:latinski ;
    Knji:jePisaoNaJeziku Knji:talijanski ;
    rdfs:label ?knjizevnik .
}
```

Rezultati:

Andrović, Nikola

Primjeri 5, 6 i 7 ukazuju na povećanje broja zadanih kriterija pretraživanja i dobivanje točnog rezultata, iako se ne smanjuje njihov broj. Slijedi pojedinačno objašnjenje svakog pojedinog primjera.

Primjer 5: Koji su dubrovački književnici pisali drame?

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?knjizevnik) as ?knjizevnik WHERE
{
  ?grad rdfs:label "Dubrovnik"@hr .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jeAutor Knji:drama ;
    rdfs:label ?knjizevnik .
}
Rezultati:
    Ban, Matija
```

Primjer 6: Koji su dubrovački književnici pisali drame na talijanskom jeziku?

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?knjizevnik) as ?knjizevnik WHERE
{
  ?grad rdfs:label "Dubrovnik"@hr .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jeAutor Knji:drama ;
    Knji:jePisaoNaJeziku Knji:talijanski ;
    rdfs:label ?knjizevnik .
}
Rezultati:
    Ban, Matija
```

Primjer 7: Koji su dubrovački književnici pisali drame na talijanskom jeziku i uređivali časopis Dubrovnik?

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?knjizevnik) as ?knjizevnik ?casopis WHERE
{
  ?grad rdfs:label "Dubrovnik"@hr .
  ?casopis rdfs:label "Dubrovnik časopis"@hr .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jeAutor Knji:drama ;
    Knji:jePisaoNaJeziku Knji:talijanski ;
    Knji:jeUrednik ?casopis ;

```

```
        rdfs:label ?knjizevnik .
    }
```

Rezultati:

Ban, Matija

Prethodni primjeri su se odnosili na dobivanje rezultata vezanih uz određene književnike. Slijedeći primjeri 8, 9 i 10 će se odnositi na dobivanje odgovora o književnim djelima dubrovačkih književnika. Navedeni primjeri pokazuju mogućnost općenitijih pregledavanja s manjim brojem postavljenih kriterija do sve specifičnijih pregledavanja postavljanjem sve većeg broja kriterija.

Primjer 8: Koja su djela dubrovačkih književnika napisana na talijanskom jeziku?

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?naziv) as ?naziv STR(?autor) WHERE
{
    ?grad rdfs:label "Dubrovnik"@hr .
    ?individua rdf:type Knjizevnici:Knjizevnik ;
        Knji:jeMjestoRodjenja ?grad ;
        rdfs:label ?autor .
    ?djelo Knji:jeJezikDjela Knji:talijanski ;
        Knji:autorDjelaJe ?individua ;
        rdfs:label ?naziv.
}
```

Rezultati:

Moscovita (Ban, Matija)
Fingallo (Ban, Matija)
Il terremoto di Ragusa (Ban Matija)
O Kristovoj mucu (Bunić, Ivan Sarov, ml.)
Nekoliko stranica o Dubrovniku (Kaznačić, Ivan August)

Primjer 9: Izlistaj naslove drama dubrovačkih književnika.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?naziv) as ?naziv STR(?autor) WHERE
{
    ?grad rdfs:label "Dubrovnik"@hr .
    ?individua rdf:type Knjizevnici:Knjizevnik ;
        Knji:jeMjestoRodjenja ?grad ;
        rdfs:label ?autor .
    ?djelo Knji:jeKnjizevnaVrsta Knji:drama ;
        Knji:autorDjelaJe ?individua ;
        rdfs:label ?naziv.
}
```

Rezultati:

Fingallo (Ban, Matija)
Il terremoto di Ragusa (Ban, Matija)
Jan Hus (Ban, Matija)
Marojica Kaboga (Ban, Matija)
Marta Posadnica (Ban, Matija)
Mejrima ili Bošnjaci (Ban, Matija)
Moscovita (Ban, Matija)

Primjer 10: Izlistaj naslove drama dubrovačkih književnika napisanih na talijanskom jeziku.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?naziv) as ?naziv STR(?autor) WHERE
{
?grad rdfs:label "Dubrovnik"@hr .
?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    rdfs:label ?autor .
?djelo Knji:jeKnjizevnaVrsta Knji:drama ;
    Knji:jeJezikDjela Knji:talijanski ;
    Knji:autorDjelaJe ?individua ;
    rdfs:label ?naziv.
}
```

Rezultati:

Fingallo (Ban, Matija)
Il terremoto di Ragusa (Ban, Matija)
Moscovita (Ban, Matija)

Primjer 11 ukazuje i na mogućnost postavljanja upita o dobivanju nagrade za rad pojedinih književnika. Upotrebljen je osim dobivana nagrade i kriterij, odnosno relacija, *je rođen u Splitu*.

Primjer 11: Koji su splitski književnici dobili nagradu Vladimir Nazor?

Postavljeni kriteriji i dobiveni rezultati su prikazani na Slici 61.

Pretraživanje baze:

Pronađi: gdje:

Književnik:

Književnik:

Rezultati pretraživanja:

Knjizevnik
Delorko, Olinko
Fabrio, Nedjeljko
Jeličić, Živko

Slika 61: Rezultati za upit „Koji su splitski književnici dobili nagradu Vladimir Nazor?“

Primjeri 12 i 13 ukazuju kako se može postaviti upit koji se odnosi na 1988 godinu u odnosu na književnike ili djela. Primjer 12 će dati odgovor na pitanje koji su splitski književnici objavili djela 1988 godine, a Primjer 13 će pak dati odgovor na pitanje koja su djela splitskih književnika objavljena 1988 godine. Slijede odgovori na postavljene upite.

Primjer 12: Koji su splitski književnici objavili djela 1988 godine?

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT STR(?knjizevnik) as ?knjizevnik WHERE
{
  ?grad rdfs:label "Split"@hr .
  ?djelo Knji:godina_objavljivanja "1988"^^xsd:int .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jeAutor ?djelo ;
    rdfs:label ?knjizevnik .
}
```

Rezultati:

Baras, Frano
 Delorko, Olinko
 Jeličić, Živko
 Maroević, Tonko
 Marković, Tonči Petrasov
 Mrduljaš, Igor
 Tomasović, Mirko
 Zurl, Marino

Primjer 13: Koja su djela splitskih književnika objavljena 1988 godine?

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX Knji: <http://www.semanticweb.org/ontologies/Knji.owl#>
PREFIX Knjizevnici: <http://www.semanticweb.org/ontologies/Ontologies/Knjizevnici.owl#>
SELECT ?djelo WHERE
{
  ?grad rdfs:label "Split"@hr .
  ?djelo Knji:godina_objavlivanja "1988"^^xsd:int .
  ?individua rdf:type Knjizevnici:Knjizevnik ;
    Knji:jeMjestoRodjenja ?grad ;
    Knji:jeAutor ?djelo ;
    rdfs:label ?knjizevnik .
}

```

Rezultati:

```

Dalmacija anno Domini
Otimanja zaboravu
Balada o kostima
Klik!
Moći ne govoriti
Dubravko Dujšin
Tradicija i Kontekst
Rijeka i 101 njih

```

Primjeri 14 i 15 ukazuju na mogućnost pregledavanja rezultata s obzirom na razliku u značenju. Odabrana su dva kriterija za svaki upit, s tim da Primjer 14 ukazuje na splitske književnike koji su uređivali časopis Mogućnosti, a Primjer 15 na splitske književnike koji su objavljivali u časopisu Mogućnosti. Slike 62 i 63 prikazuju postavljene kriterije i rezultate za navedene primjere.

Primjer 14: Koji su splitski književnici uređivali Mogućnosti?

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Rezultati pretraživanja:

Knjizevnik
Čulić, Hrvoje
Jeličić, Živko

Slika 62: Rezultati za upit „Koji su splitski književnici uređivali Mogućnosti?“

Primjer 15: Koji su splitski književnici objavljivali u Mogućnostima?

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Rezultati pretraživanja:

Književnik
Marović, Tonči Petrasov
Zuppa, Vjeran

Slika 63: Rezultati za upit „Koji su splitski književnici objavljivali u Mogućnostima?“

Moguće je i postavljanje upita za određeno vremensko razdoblje. Tako će Primjer 16 prikazati drame dubrovačkih književnika objavljene u razdoblju 1830-1840, a postavljeni kriteriji i dobiveni rezultati su prikazani na Slici 64.

Primjer 16: Koje su drame dubrovačkih književnika objavljene u razdoblju 1830-1840?

Pretraživanje baze:

Pronađi gdje:

Književno djelo

Književno djelo

Rezultati pretraživanja:

Djelo
Dobroslavo
Fingallo
Il terremoto di Ragusa
Moscovita

Slika 64: Rezultati za upit „Koje su drame dubrovačkih književnika objavljene u razdoblju 1830-1840?“

Primjeri 17 i 18 prikazuju književnike koji su članovi HAZU-a, ali s obzirom na mjesto rođenja, što je dodatni kriterij. Slijedi prikaz primjera i dobivenih rezultata na Slikama 65 i 66.

Primjer 17: Koji su dubrovački književnici članovi HAZU-a?

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Rezultati pretraživanja:

Knjizevnik
Paljetak, Luko
Bogišić, Rafo

Slika 65: Rezultati za upit „Koji su dubrovački književnici članovi HAZU-a?“

Primjer 18: Koji su splitski književnici članovi HAZU-a?

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Rezultati pretraživanja:

Knjizevnik
Fabrio, Nedjeljko
Maroević, Tonko
Novak, Slobodan
Tomasović, Mirko

Slika 66: Rezultati za upit „Koji su splitski književnici članovi HAZU-a?“

Posljedni primjer 19 će prikazati predsjednike Društva hrvatskih književnika rođenih u Splitu. Slika 67 prikazuje postavljene kriterije i dobivene rezultate.

Primjer 19: Koji su splitski književnici bili predsjednici Društva hrvatskih književnika?

Pretraživanje baze:

Pronađi gdje:

Književnik

Književnik

Rezultati pretraživanja:

Književnik
Fabrio, Nedjeljko
Vladović, Borben

Slika 67: Rezultati za upit „Koji su splitski književnici bili predsjednici Društva hrvatskih književnika?“

7.6 Zaključak

Poglavlje *Razvoj ontologije književnosti iz enciklopedički organiziranog znanja* povezal je strukturne elemente enciklopedijskog biografskog članka s gradivnim ontologijskim elementima te je ukazalo na njihovu ulogu prilikom razvoja ontologije. Istražilo se i definiralo koji su strukturni elementi enciklopedijskog biografskog članka značajni za razvoj ontologijskih odnosa ili pojmova. Rezultat je ontologija književnosti koja je u mogućnosti opisati enciklopedijsko znanje iz navedenog područja. Rezultati istraživanja provedenog u ovom poglavlju su detaljno prikazani kroz tzv. ontologijske module sa svim mogućim ontologijskim svojstvima koji se mogu upotrijebiti u opisu određenog pojma. Također je izvršen prikaz strukture i slijeda elementa enciklopedijskog biografskog članka s razvijenim ontologijskim svojstvima kojima se mogu opisati sve vrste informacija pohranjene u članku, iz čega su utvrđeni stalni elementi koji se mogu sagledati za razvoj infookvira članaka koji su prikladni za brzi uvid u najbitnije informacije o pojedinom članku. Prikazana je i praktična primjena razvijene ontologije književnosti kroz mogućnost postavljanja složenih semantičkih pitanja nestrukturiranom enciklopedijskom biografskom članku s opširnijom ili detaljnijom razinom opisa i dobivenih odgovora. Sve navedeno je doprinijelo poboljšanju prikaza i pretraživanja enciklopedijskih članka u mrežnom okruženju.

8 FORMALNA POJMOVNA ANALIZA – FCA

8.1 Uvod

Ontologije, često definirane kao eksplicitna specifikacija konceptualizacije³⁷¹, nužne su za predstavljanje i razmjenu znanja³⁷². Eksplicitna specifikacija konceptualizacije znači da ontologija opisuje pojmove i odnose koji postoje u domeni³⁷³. Dostupne su i druge slične definicije ontologije³⁷⁴, iako nisu identične, u načelu izražavaju da se bilo koja ontologija sastoji od konceptualizacije domene (odnosno načina kako gledati ili modelirati domenu) i specifikacije ove konceptualizacije (odnosno formalnog opisa). Na razini konceptualizacije odlučuje se koji objekti i odnosi između njih će biti uključeni u ontologiju i također do koje razine detalja. Konceptualizacija se formalno specificira u nekom formalnom jeziku.

Ontologije omogućuju razmjenu znanja (dijeleći razumijevanje strukture informacija između softwarskih agenata i ljudi), ponovnu upotrebu znanja (ontologija se može ponovno upotrijebiti za druge operativne sustave u sličnoj domeni) te stvaranje pretpostavki o razgovijetnosti domene. Dakle, za razmjenu znanja nužno je opisati pojmove i odnose, a ne ih samo poredati u taksonomiju. Primjerice, pojam ne bi trebao biti opisan samo s njegovom pozicijom u taksonomijskoj (*to je* relacija, engl. *is-a*) hijerarhiji, nego bi trebao biti opisan i s relacijama koje se mogu primijeniti na pojam. Slično, relacija može biti opisana s pojmovima koji su zajedno povezani s tom relacijom. Taksonomije su važne, budući da one čine "kralježnicu" ontologije, ali nisu dovoljne za razmjenu znanja. U ovom poglavlju će se pojasniti FCA pristup (engl. Formal Concept Analysis, njem. *Formale Begriffsanalyse* – FBA) koji se primjenjuje za izgradnju ontologije kako bi se ustanovio skup pojmovnih svojstava kojima su definirani pojmovi u ontologiji te isti klasificiraju pojam s ostalima u hijerarhiju.

³⁷¹ Gruber, T. A Translation Approach to Portable Ontology Specifications. // Knowledge Acquisition. 5, 2(1993), 221.

³⁷² Obitko, M. Ontologies - Description and Applications. Report No: GL 126/01. Gerstner Laboratory for Intelligent Decision Making and Control. Series of Research Reports. 2001. Dostupno na URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.419&rep=rep1&type=pdf> (1.2.2015.).

³⁷³ Harmelen, Frank Van...et al. Web Ontology Language (OWL) Reference. // W3C Recommendation 10, February 2004. Dostupno na URL: <http://www.w3.org/TR/owl-ref/> (1.2.2015.).

³⁷⁴ Obitko, M., op. cit. (bilj. 372).

8.2 Pregled i primjena FCA

Teorija formalne analize pojmova je novija metoda za analizu i predstavljanje podataka kojom se iz zadane domene izračunava pojmovna rešetka, usmjereni aciklički graf u kojem su hijerarhijski posloženi svi pojmovi iz zadane domene. Metodu FCA je osmislio početkom 1980-ih R. Wille, njemački matematičar i profesor emeritus na TU Darmstadt. Iskoristio je filozofsko tumačenje *pojma* kao jedinice misli oblikovane generalizacijom kojom se obuhvaća neki skup objekata i skup njihovih atributa^{375,376,377,378,379}. Bernhard Ganter i Rudolf Wille su predstavili metodu za izgradnju ontologija na temelju formalne pojmovne analize³⁸⁰. Formalna pojmovna analiza je teorija analize podataka koja identificira pojmovne strukture među skupovima podataka^{381,382}. Ova metoda izgradnje ontologije omogućuje otkrivanje potrebe za novim pojmovima i odnosima u ontologiji, što vodi prema ontologiji koja ima ove entitete opisane na način pogodan za dijeljenje znanja ili pretraživanje informacija³⁸³. Ove strukture su grafički predstavljene kao pojmovne rešetke (eng. *conceptual lattices*), dopuštajući analizu kompleksnih struktura i otkrivajući ovisnosti unutar podataka. Formalna pojmovna analiza je tehnika pojmovnog grupiranja (eng. *conceptual clustering technique*) s dobro razvijenim matematičkim temeljima i uspješno se koristi za široki spektar primjene u različitim granama (medicini, psihologiji, knjižnicama, softwareskom inženjerstvu i ekologiji), metodama za analizu podataka, pretraživanju informacija, otkrivanju znanja u bazama podataka te u stvaranju taksonomija i ontologija. Ovom metodom se pronalaze i vizualiziraju svi pojmovi i njihove međuovisnosti iz ulaznih podataka. Isto tako, primjenjuje

³⁷⁵ Wille, Rudolf. Restructuring lattice theory: an approach based on hierarchies of concepts // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5548 LNAI, 2009, Str.: 315.

³⁷⁶ Formal Concept Analysis - Foundations and Applications / (Eds.) Bernhard Ganter; Gerd Stumme, Rudolf Wille. Berlin Heidelberg: Springer-Verlag, 2005. Str. 2.

³⁷⁷ Wolff, Karl Erich. A first course in Formal Concept Analysis // Advances in Statistical Software 4, 1993. Str. 431.

³⁷⁸ Ganter, Bernhard; Wille, Rudolf. Formal Concept Analysis: mathematical foundations. Berlin: Springer-Verlag, 1999. Str. 35.

³⁷⁹ Belohlavek, Radim. Introduction to Formal Concept Analysis. Olomuc, 2008. Dostupno na URL: belohlavek.inf.upol.cz/vyuka/IntroFCA.pdf (01.02.2015.).

³⁸⁰ Ganter, Bernhard; Wille, Rudolf, op. cit. (bilj. 378), str. 65.

³⁸¹ Ibid.

³⁸² Beneš, M., Snášel, V. Deducing Design Class Hierarchy from Object Properties. ISM'2002. Rožnov pod Radhoštěm. Czech Republic. 2002. Str. 11.

³⁸³ Obitko, M.; Smid, J.; Snášel, V. Using Easel Types for Document Management and Retrieval. PSMP3 Workshop within AIA2004. IASTED/ActaPress. 2004. Dostupno na URL: <http://www.actapress.com/Abstract.aspx?paperId=15896> (01.02.2015.).

se kao metoda strojnog učenja i u području istraživanja umjetne inteligencije^{384,385,386,387}. U teoriji baze podataka, FCA se opsežno koristila za stvaranje i upravljanje klasnom hijerarhijom^{388,389,390,391}.

FCA se temelji na filozofskom poimanju kako se pojam tvori od dva dijela: ekstenzije koja se sastoji od svih objekata koji pripadaju pojmu i intencije koja obuhvaća sve atribute koje dijele ti objekti. Ovo razumijevanje omogućuje izvođenje svih pojmova iz danog konteksta (tablice podataka) i uvođenje hijerarhije. Dakle, formalna pojmovna analiza predstavlja pristup za identificiranje strukture prisutne u domeni. Temelji se na potpunoj rešetki svih formalnih pojmova u domeni te analizira podatke koji opisuju odnos između određenog skupa objekata i određenog skupa svojstava. Dva su cilja konstrukcije FCA rezultata analize na podacima:

- prvi je pojmovna rešetka tj. zbirka formalnih pojmova u podacima koji su hijerarhijski poredani s podpojam-nadpojam odnosom. Formalni pojmovi, su parovi $\langle A, B \rangle$ gdje su $A \subseteq X$ skup objekata i $B \subseteq Y$ skup atributa takvih da je A je skup svih objekata koji imaju sve atribute iz B, a B je skup svih atributa koji su zajednički svim objektima iz A.
- drugi rezultat FCA analize je skup tzv. atributnih implikacija koji opisuje određene atribute ovisnosti koji vrijedi u podacima.

Ulazni podaci za metodu FCA se prikazuju u matričnom obliku – svaki redak se odnosi na jedan objekt iz određene domene, a svaki stupac na jedan od definiranih atributa. Treba

³⁸⁴ Belohlavek, Radim, op. cit. (bilj. 379)

³⁸⁵ Formal Concept Analysis: third international conference, ICFCA 2005, Lens, France, February 14-18, 2005: proceedings / Bernhard Ganter, Robert Godin, (Eds.). Berlin ; New York: Springer 2005.

³⁸⁶ Formal Concept Analysis: 5th International Conference, ICFCA 2007, Clermont-Ferrand, France, February 12-16, 2007: proceedings / S.O. Kuznetsov, S. Schmidt (Eds.). Berlin ; New York: Springer 2007.

³⁸⁷ Formal Concept Analysis: seventh international conference, ICFCA 2009 Darmstadt, Germany, May 21-24, 2009: proceedings / S. Ferre, S. Rudolph (Eds.). Berlin ; New York: Springer 2009.

³⁸⁸ Schmitt, Ingo; Saake, Gunter: Merging inheritance hierarchies for database integration. // 3rd IFCIS Intl. Conf. on Cooperative Information Systems: proceedings, August 20-22, 1998, New York City, New York, USA. Str. 122-131.

³⁸⁹ Waiyamai, Kitsana; Taouil, Rafik; Lakhal, Lotfi. Towards an object database approach for managing concept lattices. // Conceptual modeling - ER '97 : 16th International Conference on Conceptual Modeling, Los Angeles, California, USA, November 3-5, 1997 : proceedings. Berlin ; New York : Springer, 1997, Str. 299-312.

³⁹⁰ Amer-Yahia, Sihem...et al. iO2: An algorithmic method for building inheritance graphs in object database design. // Proceedings of ER'96 (15th International Conference on Conceptual Modeling) , 1157. Berlin [etc.] : Springer, 1996. Str. 422-437.

³⁹¹ Missikoff, M.; Scholl, M.: An algorithm for insertion into a lattice: application to type classification. // Proceeding International Conference, FODO 1989 on Foundations of Data Organization and Algorithms. New York: Springer -Verlag, 1989. Str. 64-82.

pripremiti matricu (i/ili ulazne podatke) tako da svaki objekt ili ima ili nema pojedini atribut. Ako neki objekt ima određeni atribut to se naznači (npr. oznakom "X") na mjestu gdje se siječe redak tog objekta sa stupcem tog atributa. U suprotnom sjecište retka tog objekta sa stupcem tog atributa ostaje prazno. Ova matrica se definira kao *formalni kontekst* nad kojim se vrši analiza. Formalna analiza pojmova rezultira s dva skupa izlaznih podataka. Prvi skup (A) daje hijerarhijski odnos svih ustanovljenih pojmova u obliku linijskog dijagrama odnosno pojmovne rešetke (engl. *Concept lattice*, izvorno njem. *Bergiffsverband*). Drugi skup (B) daje popis svih ustanovljenih međuovisnosti atributa iz formalnog konteksta.

Metoda FCA je matematički strogo definirana, a temelji se na Galoisovim vezama (en. *Galoisconnections*) i operatorima zatvaranja (en. *Closure operators*)³⁹². FCA³⁹³ je matematički formalizam koji omogućuje izvođenje pojmovnih rešetki iz formalnog konteksta $\langle X, Y, I \rangle$. Rudolf Wille definira matematički formalni kontekst kao trojku $K = \langle X, Y, I \rangle$ gdje je X skup objekata, Y je skup atributa i I je incidencija - binarna relacija između X i Y gdje xIy (što znači $(x, y) \in I$) ukazuje da objekt x ima atribut y. Formalni pojam od formalnog konteksta $K = \langle X, Y, I \rangle$ je definiran kao par $\langle A, B \rangle$ s

$$A \subseteq X, B \subseteq Y, A = \{x \in X \mid xIy \text{ za sve } y \in B\} \text{ i } B = \{y \in Y \mid xIy \text{ za sve } x \in A\};$$

A i B se zove ekstenzija i intencija formalnog pojma $\langle A, B \rangle$.

Hijerarhijski odnos podpojam-nadpojam (\leq) izražen rečenicom, formalni pojam $\langle A_1, B_1 \rangle$ je podpojam formalnog pojma $\langle A_2, B_2 \rangle$ "modeliran je Definicijom 1:

$$\langle A_1, B_1 \rangle (\leq) \langle A_2, B_2 \rangle \Leftrightarrow A_1 \subseteq A_2 \text{ (or } B_2 \subseteq B_1)$$

Skup svih formalnih pojmova formalnog konteksta $K \langle X, Y, I \rangle$ s relacijom redosljedica (eng. order relation) je potpuna rešetka nazvana *pojmovna rešetka formalnog konteksta* $K \langle X, Y, I \rangle$ i označava se s $B = (X, Y, I)$ ³⁹⁴. Pojmovna rešetka se sastoji od formalnih pojmova organiziranih u rešetku s parcijalnim uređenjem. Oslanjajući se na relaciju \leq , skup svih pojmova ekstrahiranih iz konteksta $K \langle X, Y, I \rangle$ je organiziran unutar potpune rešetke,

³⁹² Blyth, T. S. *Lattices and Ordered Algebraic Structures*. / T.S. London: Springer, 2005. Str. 45.

³⁹³ Ganter, Bernhard; Wille, Rudolf, op. cit. (bilj. 378), str. 67.

³⁹⁴ Wille, Rudolf. *Formal Concept Analysis as Applied Lattice Theory // Concept Lattices and Their Applications* / Ben Yahia, Sadok; Mephu Nguifo, Engelbert; Belohlavek, Radim (Eds.). Berlin ; New York : Springer, 2008. Str. 42-67.

nazvane pojmovna rešetke i označene s $B=(X,Y,I)$ ³⁹⁵. Slijede definicije prethodno navedenog:

Definicija 2: Formalni kontekst metode FCA je trojka $\langle X,Y,I \rangle$ gdje su X i Y neprazni skupovi, a I binarna relacija između X i Y. Za ulaznu matricu s n redaka i m stupaca formalni kontekst $\langle X,Y,I \rangle$ se sastoji od skupova $X=\{x_1, \dots, x_n\}$ i $Y=\{y_1, \dots, y_m\}$ te relacije I definirane kao $\langle x_i, y_j \rangle \in I$, onda i samo onda ako sjecište i-tog retka i j-stupca ulazne matrice nije prazno.

Definicija 3: Za formalni kontekst $\langle X,Y,I \rangle$ definirani su operatori formiranja pojmova $\uparrow: 2^X \rightarrow 2^Y$ i $\downarrow: 2^Y \rightarrow 2^X$ za svaki $A \subseteq X$ i $B \subseteq Y$ kao:

$$A^\uparrow = \{y \in Y \mid \text{za sve } x \in A: \langle x, y \rangle \in I\}$$

$$B^\downarrow = \{x \in X \mid \text{za sve } y \in B: \langle x, y \rangle \in I\}$$

A^\uparrow predstavlja skup svih atributa koji imaju svi objekti iz A, a B^\downarrow predstavlja skup svih objekata koji dijele sve atribute iz B. Formalni pojam je segment formalnog konteksta u kojem različiti objekti dijele iste atribute.

*Definicija 4: Formalni pojam u formalnom kontekstu $\langle X,Y,I \rangle$ je par $\langle A,B \rangle$ ($A \subseteq X$ i $B \subseteq Y$) za koji vrijedi $A^\uparrow = B$ i $B^\downarrow = A$. Dakle, $\langle A,B \rangle$ je formalni pojam onda i samo onda ako se A sastoji samo od objekata koji imaju sve atribute iz B (ekstenzija pojmova – engl. *extent*), a B se sastoji samo od atributa koje dijele svi objekti iz A (intencija pojmova – engl. *intent*). Formalni pojmovi se međusobno nalaze u hijerarhijskom odnosu (*podpojam* i *nadpojam*).*

³⁹⁵ Bendaoud, Rokia; Napoli, Amedeo; Toussaint, Yannick. Formal Concept Analysis: A unified framework for building and refining ontologies // Knowledge Engineering: Practice and Patterns : 16th International Conference, EKAW 2008, Acitrezza, Sicily, Italy September 29 - October 3, 2008, Proceedings. / Aldo Gangemi, Jérôme Euzenat, (Eds.). Berlin ; New York : Springer, 2008. Str. 156-171.

Svi formalni pojmovi nekog formalnog konteksta i njihovi međusobni odnosi mogu se transformirati u strukturu pojmovne rešetke koja se prikazuje u obliku Hasseovog (linijskog) dijagrama³⁹⁶.

Definicija 5: Struktura $B(X, Y, I)$ se definira kao kolekcija svih formalnih pojmova, $B = (X, Y, I) = \{ \langle A, B \rangle \in 2^X \times 2^Y \mid A^\uparrow = B, B^\downarrow = A \}$. Ova struktura s uključenim hijerarhijskim odnosom \leq čini pojmovnu rešetku formalnog konteksta $\langle X, Y, I \rangle$.

Struktura pojmovne rešetke zadovoljava matematičku definiciju rešetke jer predstavlja parcijalno uređeni skup u kojem bilo koja dva elementa (formalna pojma) imaju zajednički supremum (\uparrow , *najmanju gornju granicu* - engl. *join*, nadpojam) i zajednički infimum (\downarrow , *najveću donju granicu* - engl. *meet*, podpojam)^{397,398}. Rezultirajuće pojmovne rešetke se mogu transformirati u pojmovnu hijerarhiju predstavljenu formalizmima opisne logike kako bi se omogućilo formalno predstavljanje i zaključivanje³⁹⁹.

8.3 Proces stvaranja ontologije upotrebom FCA

Tvorba ontologije započinje stvaranjem hijerarhije relevantnih pojmova. Taksonomijski odnos, nazvan *to je* odnos (engl. *is a*) ili odnos podrazumijevanja (engl. *subsumtion relation*) između pojmova, percipira se kao temelj bilo koje ontologije. Ovaj pogled na stvaranje ontologije je uglavnom pod utjecajem procesa objektno orijentiranih sustava (engl. *object oriented systems*). Karakteristike predložene FCA metode⁴⁰⁰ za stvaranje ontologije su:

- pojmovi su opisani svojstvima
- svojstva određuju hijerarhiju pojmova; drugim riječima hijerarhija nije izričito definira od strane dizajnera
- kada su svojstva različitih pojmova ista, onda su i pojmovi također isti

³⁹⁶ Žubrnić, Darko. Diskretna matematika. Zagreb: Element, 2002. 2. izd. Str. 74-76.

³⁹⁷ Ibid.

³⁹⁸ Blyth, T.S., op. cit. (bilj. 392), str. 105.

³⁹⁹ Bendaoud, Rokia; Napoli, Amedeo; Toussaint, Yannick, op. cit. (bilj. 395), str. 169.

⁴⁰⁰ Obitko, Marek; Snašel, Vaclav; Smid, Jan. *Ontology Design with Formal Analysis*. Dostupno na URL: <http://ceur-ws.org/Vol-110/paper12.pdf> (02.02.2015.).

Značajna prednost stvaranja ontologije podržane alatom koji koristi FCA proces je mogućnost njegove upotrebe u suradničkom okruženju s više ontologijskih dizajnera koji rade na jednoj ontologiji. Proces stvaranja ontologije podržane FCA analizom može imati sljedeće korake:

1. Započeti s praznim skupom pojmova i svojstava
2. Dodati pojmove i svojstva po potrebi u pojmovnu tablicu
3. Rešetke pojmova s njihovim svojstvima vizualizirati upotrebom FCA analize, što omogućuje dizajnerima predočavanje ontologije ili njezinih dijelova
4. Na temelju vizualizacije, dizajner može preinačavati ontologiju kako slijedi:
 - a. "Direktno" uređivanje (kako direktno zahtijeva ontologijska upotreba)
 - i. Dodavanje ili uklanjanje pojma
 - ii. Dodavanje ili uklanjanje svojstva
 - iii. Dodjeljivanje svojstva pojmu ili uklanjanje svojstva pojmu
 - b. Uređivanje kako sugerira ontologijski dizajnerski alat
 - i. Kada dva pojma dolaze na isto mjesto vizualizacijom korištenjem FCA, oni bi trebali biti ili sjedinjeni u jedan ili bi razlika trebala biti dodana (u obliku svojstva kojeg jedan pojam ima, a drugi nema)
 - ii. FCA može generirati pojmove koji nastaju svojstvima i koji su nad-pojmovi definiranih pojmova, ali nisu izričito spomenuti u pojmovnoj tablici; to sugerira da ovaj pojam može biti izrađen (ontologijski dizajner može samo dodati ime pojma prema prijedlogu, svojstva novog pojma su očita iz generirane rešetke)
5. Ovaj proces se ponavlja dok ontologijski dizajner bude zadovoljan.

8.4 Primjer stvaranja taksonomije ontologije na podkorporusu hrvatske književnosti iz Hrvatske enciklopedije

FCA je učinkovit u analiziranju podataka i otkrivanju skrivene veze u podacima. Omogućuje korisnicima vizualizaciju podataka na jednostavan i razumljiv način. Podaci su predstavljeni u obliku formalnog konteksta, odnosno jednostavnog određivanja koji atributi zadovoljavaju određene objekte. Cilj ovog odlomka je prikazati polu-automatskog generiranje taksonomije ontologije primjenom formalne pojmovne analize na konkretnom primjeru.

8.4.1 Ekstakcija formalnog konteksta

FCA započinje ekstrakcijom podataka iz određenih tekstova kako bi se generirao odgovarajući formalni kontekst za odabrane tekstove i postavljene attribute. Odabrani tekstovi su 37 književnika hrvatske književnosti, od kojih je njih 36 rođeno u Splitu, a jedan u Zagrebu (August Šenoa). Ekstrakcija informacija se obavlja u pripremljenom Python programu tako da se upiše morfološki odgovarajući oblik riječi za attribute koji se traže u tekstovima te redni broj književnika čiji su potpuni tekstualni opisi pohranjeni u html dokumentu te su isti povezani s Pythonom kao bi se moglo izvršiti pretraživanje objekata prema zadanim atributima. Odabrani oblici riječi za attribute te redni brojevi pod kojim su opisani pojedini književnici su u PyScripeter dijelu zadani kao osnovni parametri:

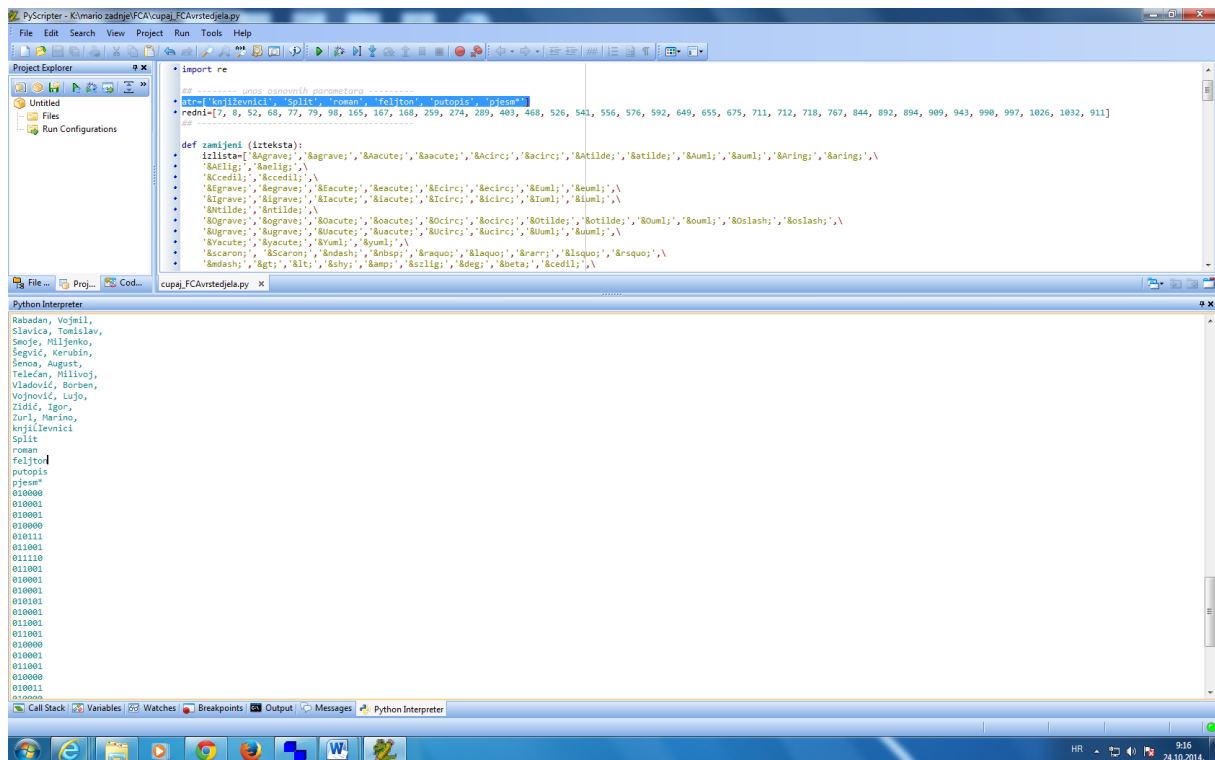
```
## ----- unos osnovnih parametara -----
```

```
atr=['književnik', 'Split', 'roman', 'feljton', 'putopis', 'pjesm*']
```

```
redni=[7, 8, 52, 68, 77, 79, 98, 165, 167, 168, 259, 274, 289, 403, 468, 526, 541, 556, 576, 592, 649, 655, 675, 711, 712, 718, 767, 844, 892, 894, 909, 943, 990, 997, 1026, 1032, 911]
```

```
## -----
```

Atributom Split se ekstrahira mjesto rođenja, a oblik riječi tj. atribut pjesm* je odabran za ekstakciju pjesme. Ostali atributi su jasni sami po sebi. Pokretanjem gumba za izvršavanje pretraživanja postavljenih atributa nad tekstovima Python automatski izvršava ekstrakciju prema zadanim atributima u obliku matrice (kao što je prikazano na Slici 68. u dijelu Python Interpreter) i izlazne con datoteke.



Slika 68: Ekstrakcija podataka Python programom

Izlazna con datoteka se učita u openfca aplikaciju Conflexplore te se automatski generira formalni kontekst iz zadanih objekata i atributa. Dobiveni rezultati se provjeravaju i utvrđuju se pogreške (trinaest neobilježenih u atributu književnik, tri u ekstrahiranju atributa roman i dva u ekstrahiranju atributa pjesme). Ručno se izvrše potrebne korekcije na dobivenom formalnom kontekstu prikazanom na Slici 69. Objekti formalnog konteksta su tekstualni enciklopedijski članci u kojima su entiteti iz područja književnosti (tj. književnici) opisani prirodnim jezikom, a atributi su imenice (odnosno oblik određene riječi) za koje smo zainteresirani. Objekti su smješteni u recima tablice, a atributi u stupcima. Pretpostavka je da tekstualni opisi tj. enciklopedijski članci upotrebljavaju rječnik specifičan za područje književnosti. Razmatranjem formalnog konteksta prikazanog na Slici 69 uočava se njegovo predstavljanje u obliku binarnog odnosa između tekstualnih opisa pojedinog književnika (objekata) i imenica (riječi) u stupcima (atributa). Primjer prikazuje formalni kontekst koji analizira podatke prema zadanim atributima i time dajući informacije o tome tko je od 37 hrvatskih književnika rođenih u Splitu pisao određenu vrstu književnih djela za koje smo zainteresirani (roman, feljton, putopis, pjesme). Primjerice, posljednji red daje informaciju da je *Zurl, Marino* književnik rođen u Splitu te da je pisao romane, putopise i pjesme.

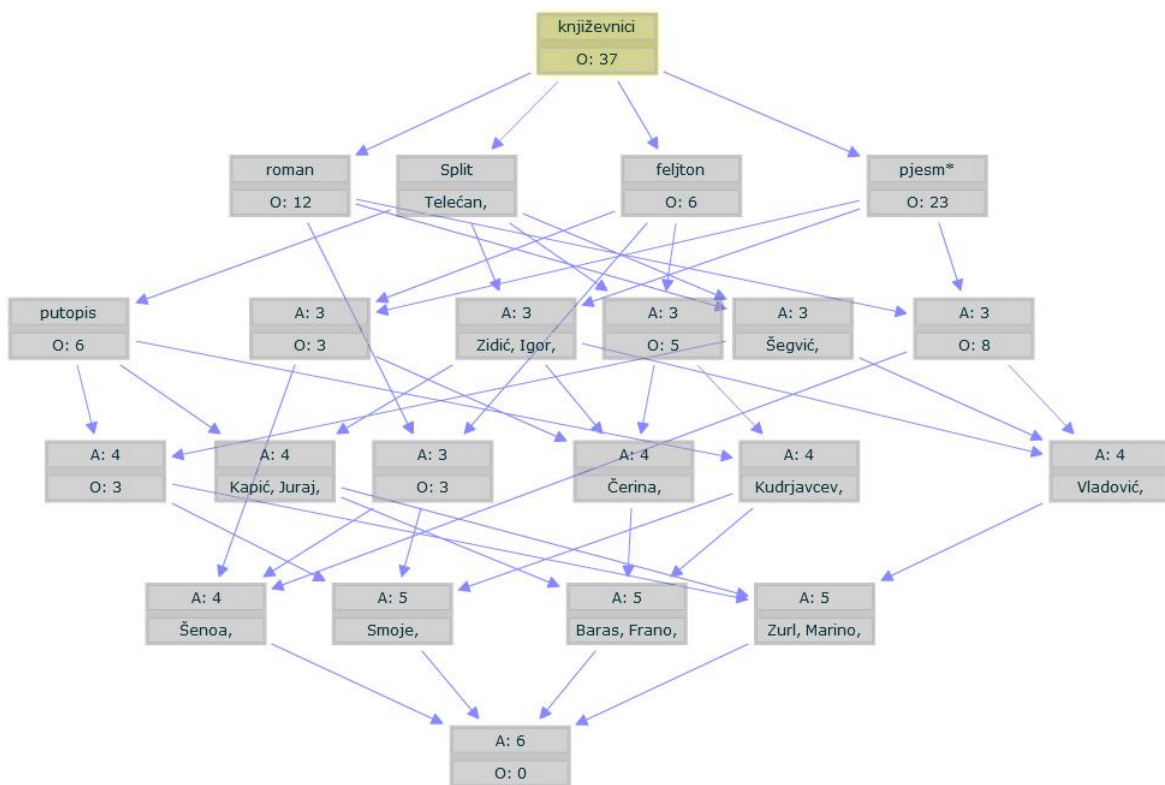
	književnici	Split	roman	feljton	putopis	pjesm*
Alberti, Matija,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alberti Matulić, N	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Arnerić, Marko,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Balistrilić, Dujam	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Baras, Frano,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Barbieri, Veljko,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Bego, Marin,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Botić, Luka,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Božičević (Božiče	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Božičević (Božiče	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Čerina, Vladimir,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Čulić, Hrvoje,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Delorko, Olinko,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Fabrio, Nedjeljko	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Grgić, Milan,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ivanišević, Đurđi	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Jeličić, Živko,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Jurišić, Šimun,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kapić, Juraj,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Kavanjin (Cavag	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Krolo, Ivan,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kudrjavcev, Anat	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Lovrić, Božo,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Maroević, Tonko	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Marović, Tonči P.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Marulić, Marko (I	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Mrduljaš, Igor,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rabadan, Vojmil	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Slavica, Tomisla	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Smoje, Miljenko,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Šegvić, Kerubin,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Šenoa, August,	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Telečan, Milivoj,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vladović, Borben	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Vojnović, Lujo,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Zidić, Igor,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Zurl, Marino,	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Slika 69: Formalni kontekst prema zadanim objektima i atributima

Prvi korak u analizi ovog formalnog konteksta je izračunavanje njegovih formalnih pojmova, što je „prirodno grupiranje“ podataka u kontekstu. Prisjetimo se, kao što je već napisano u prethodnom poglavlju, formalni pojam je par koji se sastoji od skupa objekata A i skupa atributa B, takav da objekti dijele attribute u B, a B se sastoji od točno onih atributa koji su zajednički objektima u A. Objektni skup se naziva ekstenzija, a atributni skup se naziva

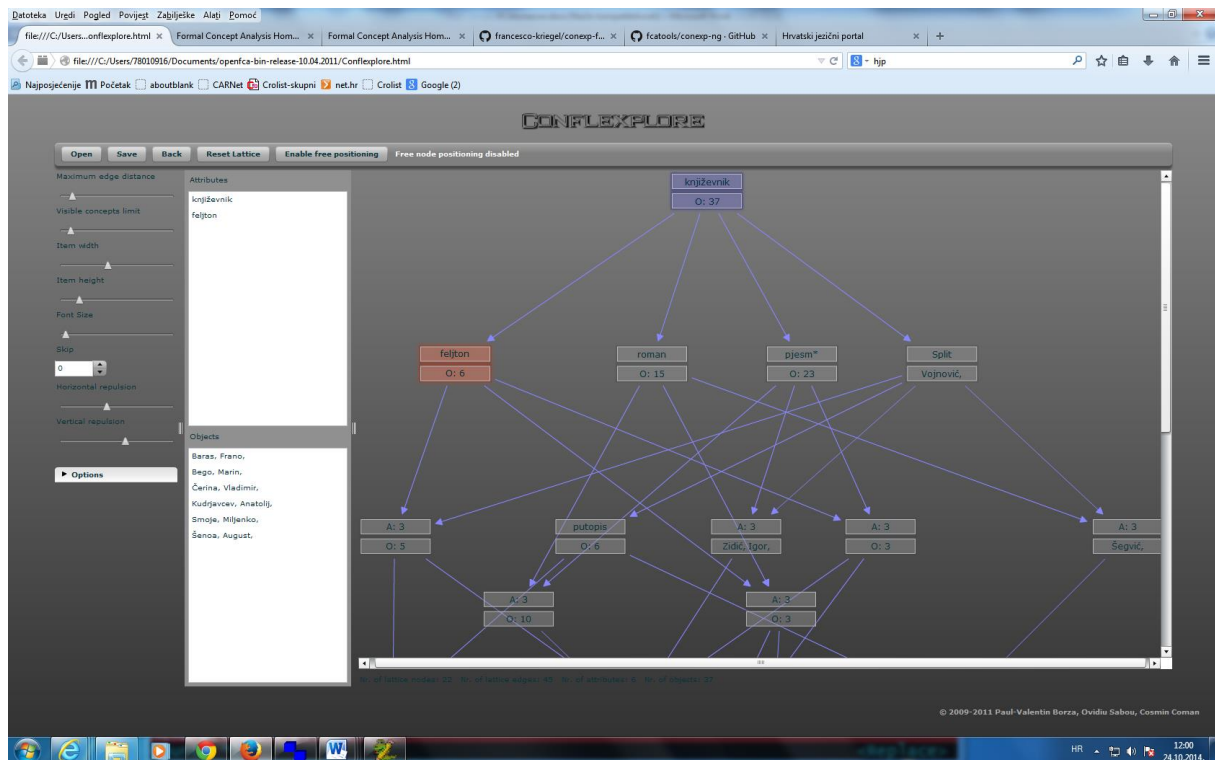
intencija formalnog pojma (A, B). Svi formalni pojmovi nekog formalnog konteksta i njihovi međusobni odnosi mogu se transformirati u strukturu pojmovne rešetke koja je grafički prikazana na Slici 70. Graf prikazuje 22 formalna čvora. Čvorovi imaju dva dijela. U gornjem se navodi atribut, a u donjem objekt. Ukoliko nije poznat atribut, onda se navodi općom oznakom A (za atribut) nakon koje se navodi broj razine. Ako za neki atribut neke razine nije poznato ime objekta onda se navodi opća oznaka O (za objekt) nakon koje piše koliko objekata ima takav (poznati ili opći) atribut.

Prema broju atributa koje sadrže pojedini formalni pojmovi grafički prikazani pojmovnom rešetkom na Slici 70, možemo ih podijeliti u šest razina. Prva razina sadrži najopćenitiji formalni pojam s jednim atributom „književnici“ (naveden u gornjem dijelu čvora) kojega sadrži svih 37 objekata. Donja razina prikazuje četiri formalna pojma koji sadrže dva atributa (npr. formalni pojam koji sadrži attribute književnici i roman). Treća razina prikazuje šest formalnih pojmova s tri atributa, dok četvrta razina prikazuje pet formalnih čvorova s četiri atributa. Na petoj razini se uočavaju tri formalna čvora s pet atributa, dok je na šestoj razini prikazan najmanje općenit formalni čvor koji je definiran sa svim atributima i ne sadrži niti jedan objekt pa je ekstenzija najnižeg formalnog pojma prazan skup. Odnosi podpojam i nadpojam između formalnih pojmova su predstavljeni strelicama u Hasse-vom dijagramu.



Slika 70: Grafički prikaz formalne rešetke

Dolaskom kursora povrh objekta (npr. feljton) s njegove lijeve strane se prikazuju odgovarajući podaci za taj formalni pojam (atributi: književnik, feljton; objekti: {(Baras, Frano; Bego, Marin; Čerina, Vladimir; Kudrjavcev, Anatolij; Smoje, Miljenko; Šenoa, August) , kao što je prikazano na Slici 71.



Slika 71: Graf pojmovne rešetke (desna strana slike) i podaci za odabrani formalni pojam (tj. objekti i atributi na lijevoj strani slike)

Potpuni prikaz svih formalnih pojmova rešetke je dan u Tablici 14 te se u njoj mogu detaljno vidjeti svi objekti i atributi svakog pojedinog pojma.

Tabela 14: Formalni pojmovi

Formalni pojmovi
Na vrhu hijerarhije se nalazi formalni pojam koji sadrži samo jedan atribut:
{(Alberti , Matija; Alberti Matulić, Nikola; Arnerić, Marko; Balistrilić, Dujam; Baras, Frano; Barbieri, Veljko; Bego, Marin; Botić, Luka; Božičević (Božičević), Božić (Nadal); Božičević (Božičević), Frano (Franciscus Natalis); Čerina, Vladimir; Čulić, Hrvoje; Delorko, Olinko; Fabio, Nedjeljko; Grgić, Milan; Ivanišević, Đurđica; Jeličić, Živko; Jurišić, Šimun; Kapić, Juraj; Kavanjin (Cavagnini), Jerolim (Girolamo); Krolo, Ivan, Kudrjavcev, Anatolij; Lovrić, Božo; Maroević, Tonko; Marović, Tonči Petrasov; Marulić, Marko; Mrduljaš, Igor; Rabadan, Vojmil; Slavica, Tomislav; Smoje, Miljenko; Šegvić, Kerubin; Šenoa, Telećan, Milivoj; Vladović, Borben; Vojnović, Lujko; Zidić, Igor; Zurl, Marino), (književnik)}
Prva razina prikazuje formalne pojmove koji sadrže dva atributa:
{(Barbieri, Veljko; Bego, Marin; Fabio, Nedjeljko; Jeličić, Živko; Lovrić, Božo; Marović, Tonči Petrasov; Slavica, Tomislav; Smoje, Miljenko; Šegvić, Kerubin; Šenoa, August; Vladović, Borben; Zurl, Marino),

(književnik, roman)}
{(Alberti , Matija; Alberti Matulić, Nikola; Arnerić, Marko; Balistrilić, Dujam; Baras, Frano; Barbieri, Veljko; Bego, Marin; Botić, Luka; Božičević (Božičević), Božić (Nadal); Božičević (Božičević), Frano (Franciscus Natalis); Čerina, Vladimir; Čulić, Hrvoje; Delorko, Olinko; Fabio, Nedjeljko; Grgić, Milan; Ivanišević, Đurđica; Jeličić, Živko; Jurišić, Šimun; Kapić, Juraj; Kavanjin (Cavagnini), Jerolim (Girolamo); Krolo, Ivan, Kudrjavcev, Anatolij; Lovrić, Božo; Maroević, Tonko; Marović, Tonči Petrasov; Marulić, Marko; Mrduljaš, Igor; Rabadan, Vojmil; Slavica, Tomislav; Smoje, Miljenko; Šegvić, Kerubin; Telečan, Milivoj; Vladović, Borben; Vojnović, Lujo; Zidić, Igor; Zurl, Marino), (književnici , Split)}
{(Baras, Frano; Bego, Marin; Čerina, Vladimir; Kudrjavcev, Anatolij; Smoje, Miljenko; Šenoa, August), (književnici , feljton)}
{(Alberti Matulić, Nikola; Arnerić, Marko; Baras, Frano; Barbieri, Veljko; Botić, Luka; Božičević (Božičević), Božić (Nadal); Božičević (Božičević), Frano (Franciscus Natalis); Čerina, Vladimir; Čulić, Hrvoje; Delorko, Olinko; Fabio, Nedjeljko; Ivanišević, Đurđica; Jeličić, Živko; Kapić, Juraj; Lovrić, Božo; Maroević, Tonko; Marović, Tonči Petrasov; Marulić, Marko; Šenoa, August; Vladović, Borben; Vojnović, Lujo; Zidić, Igor; Zurl, Marino), (književnici , pjesme)}
Sljedeća razina prikazuje formalne pojmove koji sadrže tri atributa
{(Baras, Frano; Bego, Marin; Čerina, Vladimir; Kudrjavcev, Anatolij; Smoje, Miljenko), (književnici , Split, feljton)}
{(Barbieri, Veljko; Bego, Marin; Fabio, Nedjeljko; Jeličić, Živko; Lovrić, Božo; Marović, Tonči Petrasov; Slavica, Tomislav; Smoje, Miljenko; Šegvić, Kerubin; Vladović, Borben; Zurl, Marino), (književnici , Split, roman)}
{(Baras, Frano; Bego, Marin; Kapić, Juraj; Kudrjavcev, Anatolij; Smoje, Miljenko; Zurl, Marino), (književnici , Split, putopis)}
{(Bego, Marin; Smoje, Miljenko; Šenoa, August), (književnici , roman, feljton)}
{(Alberti Matulić, Nikola; Arnerić, Marko; Baras, Frano; Barbieri, Veljko; Botić, Luka; Božičević (Božičević), Božić (Nadal); Božičević (Božičević), Frano (Franciscus Natalis); Čerina, Vladimir; Čulić, Hrvoje; Delorko, Olinko; Fabio, Nedjeljko; Ivanišević, Đurđica; Jeličić, Živko; Kapić, Juraj; Lovrić, Božo; Maroević, Tonko; Marović, Tonči Petrasov; Marulić, Marko; Vladović, Borben; Vojnović, Lujo; Zidić, Igor; Zurl, Marino), (književnici , Split, pjesme)}
{(Baras, Frano; Čerina, Vladimir; Šenoa, August), (književnici , feljton, pjesme)}
{(Barbieri, Veljko; Fabio, Nedjeljko; Jeličić, Živko; Lovrić, Božo; Marović, Tonči Petrasov; Šenoa, August; Vladović, Borben; Zurl, Marino), (književnici , roman, pjesme)}
Sljedeća razina prikazuje formalne pojmove s četiri atributa
{(Bego, Marin; Smoje, Miljenko; Zurl, Marino), (književnici , Split, roman, putopis)}
{(Baras, Frano; Čerina, Vladimir), (književnici , Split, feljton, pjesma)}
{(Baras, Frano; Kapić, Juraj; Zurl, Marino), (književnici , Split, putopis, pjesma)}
{(Barbieri, Veljko; Fabio, Nedjeljko; Jeličić, Živko; Lovrić, Božo; Marović, Tonči Petrasov; Vladović, Borben; Zurl, Marino), (književnici , Split, roman, pjesma)}
{(Baras, Frano; Bego, Marin; Kudrjavcev, Anatolij; Smoje, Miljenko), (književnici , Split, feljton, putopis)}
{(Šenoa, August), (književnici , roman, feljton, pjesma)}
Sljedeća razina prikazuje formalne pojmove s 5 atributa
{(Bego, Marin; Smoje, Miljenko), (književnici , Split, roman, feljton, putopis)}
{(Zurl, Marino), (književnici , Split, roman, putopis, pjesma)}
{(Baras, Frano), (književnici , Split, feljton, putopis, pjesma)}
Formalni pojam sa šest atributa
{(∅), (književnici , Split, feljton, roman, putopis, pjesma)}

8.4.2 Redukcija informacija

U pojmovnoj rešetki postoje suvišne informacije koje se mogu iz nje ukloniti, bez gubitka bilo kojih informacija. Dvije su vrste informacija koje se mogu reducirati:

- 1.) Višak elemenata u intenciji formalnih pojmova: za par (A, B), B (intencija) će se pojaviti u svakom potomku. Pojedini nasljedni elementi intencije se mogu ukloniti. U navedenom primjeru formalni pojam na vrhu hijerarhije koji sadrži atribut književnik i sve objekte ostaje isti. Provođenjem postupaka eliminacije utvrđeno je da se iz intencije potomaka formalnog pojma *Književnici* može ukloniti atribut književnik, a ostali su ostavljeni kako bi se lakše imenovali objekti.
- 2.) Višak objekata u ekstenziji formalnih pojmova

Primjenom postupka eliminacije pod točkama 1) i 2) na prikazanom primjeru pojmovne rešetke dobivaju se formalni pojmovi sadržani u reduciranoj rešetci intencije L_{CR} , prikazani u prvom stupcu Tablice 15.

8.4.3 Imenovanje pojmova

Formalnim pojmovima prikazanim u Tablici 15 dodijeljena su imena prema sljedećem postupku:

1. Pojmovi dobivaju jedinstveno ime koje je ime atributa formalnih pojmova, preostalog nakon postupka redukcije. Za svaki atribut c , postoji najopćenitiji pojam čija intencija uključuje c , što se naziva atributni pojam i ime ovog pojma je ime odgovarajućeg(ih) atributa, odnosno njegovog izvedenog oblika riječi koja potječe od imena atributa. Dakle, atributi formalnog pojma ukazuju na specifični pojam određene domene.
2. Nakon imenovanja provedenog pod točkom 1), mogu postojati čvorovi koji nisu dobili ime, a njihova imena trebaju osigurati stručnjaci područja ili ontologijski dizajneri. Moguće je automatski generirati formalna imena (C1, C2...) te zatim pitati savjet od stručnjaka ili ontologijski dizajner može analizom gornjeg dijela rešetke odrediti odgovarajuće ime. Ovi čvorovi zapravo označavaju nove, nepoznate pojmove jer tekstovi specifični za domenu ne uključuju bilo kakvu imensku frazu za označavanje ovih pojmova.

Tabela 15: Formalni pojmovi sadržani u reduciranoj rešetci

<i>Reducirana latica intenata</i> L_{CR}	<i>Imenovani pojmovi</i>
(književnik)	Književnici
(roman)	Romanopisci
(Split)	Splitski književnici
(feljton)	Feljtonisti
(pjesme)	Pjesnici

(Split, feljton)	Splitski feljtonisti
(Split, roman)	Splitski romanopisci
(Split, putopis)	Splitski putopisci
(roman, feljton)	C1
(Split, pjesme)	Splitski pjesnici
(feljton, pjesme)	C2
(roman, pjesme)	C3
(Split, roman, putopis)	C4
(Split, feljton, pjesma)	C5
(Split, putopis, pjesma)	C6
(Split, roman, pjesma)	C7
(Split, feljton, putopis)	C8
(roman, feljton, pjesma)	C9
(Split, roman, feljton, putopis)	C10
(Split, roman, putopis, pjesma)	C11
(Split, feljton, putopis, pjesma)	C12
(Split, feljton, roman, putopis, pjesma)	C13

8.4.4 Model logike prvog reda pojmovne rešetke

U svrhu izgradnje logičkog modela prvog reda za pojmovnu rešetku potrebno je definirati pridruživanja iz strukture rešetke na jezik logike prvog reda. Ovaj proces omogućuje logičko izražavanje ontologijske rešetke i specificiranje namjeravanog semantičkog opisa u logici prvog reda. Za tu svrhu je uvedena Horn logički temeljena formulacija pojmovne rešetke. Primjenjuju se pravila parcijalnog redoslijeda odnosa, aksioma i operacija rešetke u svrhu zaključivanja o ontologiji.

8.4.4.1 Izgradnja jezika

Upotrijebljena je standardna sintaksa za logiku prvog reda i definiran je jednostavni jezik pravila temeljen na Horn klauzuli:

1. Skup konstanti N se sastoji od skupa imena pojmova C , imena svojstava A i posebnih imena bilo kojih imena (vrh rešetke, prazan vrhu je uvijek Istina) i nula (engl. nil, dno rešetke, prazno dno je uvijek Netočno).
2. Skup imena varijabli V .
3. Skup predikatnih simbola P .

Termini su konstante ili varijable. Atom (atomska formula) je formula u obliku $p(t_1, \dots, t_n)$, gdje je p predikatni simbol, a t_1, \dots, t_n termini. Formula se zove osnovna (engl. ground formula) ako ne sadrži varijable.

Horn klauzula (pravilo) ima najviše jedan atom na svojoj glavi i postoje formule u sljedećem obliku:

$$A \leftarrow B_1, B_2, \dots, B_n,$$

gdje je B_2, \dots, B_n konjunkcija atoma B_i , $i = 0, \dots, n$ i univerzalna kvantifikacija varijabli se pretpostavlja.

Definitna klauzula ima točno jedan atom u glavi. Osnovna pravila (engl. ground rule) pravila ne sadrže varijable. Pravilo s praznom glavom se smatra upitom ili ograničenjem. Osnovno pravilo s praznim tijelom se naziva činjenica.

Pravila se koriste za definiranje pojmovnih odnosa i operacija na rešetki. Interpretacija za pravila jezika je definirana kao skup osnovnih atoma (engl. ground atom) izgrađen od predikatnih imena u P i konstanti u N .

8.4.4.2 Transformacija pojmovne rešetke u jezik pravila

Pridruživanja iz rešetke u jezik pravila su definirana kao pridruživanja pojmova i taksonomijskih odnosa. Pojmovi su predstavljeni ontologijskim imenima kao konstante u jeziku pravila, npr. književnici, romanopisci itd. Pridruživanja taksonomijskih odnosa se izvršavaju predikatom *subconcept*(Pojam1, Pojam2) kako bi se označilo da je Pojam1 neposredan podpojam od Pojam2. Predikat *to-je* (engl. *is-a*) se upotrebljava za predstavljanje odnosa parcijalnog poretka između pojmova. Primjerice, predikat *isa*(romanopisci, književnici) definira odnosa parcijalnog poretka između pojmova romanopisci i književnici navodeći da su romanopisci književnici (odnosno, romanopisci su podpojam od književnici).

8.4.4.3 Pravila za aksiome rešetke

L_{CR} je potpuna rešetka, a pravila za aksiome rešetke su:

Refleksivnost:

$\text{isa}(\text{Concept}, \text{Concept})$

Tranzitivnost:

$\text{isa}(\text{Concept1}, \text{Concept2}) \leftarrow \text{subconcept}(\text{Concept1}, \text{Concept2})$

$\text{isa}(\text{Concept1}, \text{Concept2}) \leftarrow \text{isa}(\text{Concept1}, \text{Concept3}), \text{isa}(\text{Concept3}, \text{Concept2})$

Predikatni *subconcept* označava da je *Concept1* odmah podpojam od *Concept2*.

Antisimetrija:

$\text{equal}(\text{Concept1}, \text{Concept2}) \leftarrow \text{isa}(\text{Concept1}, \text{Concept2}), \text{isa}(\text{Concept2}, \text{Concept1})$

8.4.4.4 Pravila za operacije rešetke

L_{CR} je potpuna rešetka, a za svaki skup pojmova postoji uvijek najveća donja granica (eng. greatest lower bound - glb ili najveći zajednički podpojam) i najmanja gornja granicu (eng. least upper bound - lub ili najmanji zajednički nadpojam). Podpojam rešetke se koristi za izračunavanje glb, a nadpojam je operacija za izračunavanje lub. Ove operacije su definirane upotrebom sljedećih pravila:

Operacije za izračunavanje podpojma

$\text{common_subconcept}(C, C1, \dots, Ck) \leftarrow \text{isa}(C, C1), \dots, \text{isa}(C, Ck)$

$\text{greatest_common_subconcept}(C, C1, \dots, Ck) \leftarrow \text{common_subconcept}(C, C1, \dots, Ck),$
 $\text{common_subconcept}(T, C1, \dots, Ck), \text{isa}(T, C)$

Predikat $\text{common_subconcept}(C, C1, \dots, Ck)$ znači da je pojam *C* zajednički podpojam skupa pojmova $\{C1, \dots, Ck\}$.

Predikat $\text{greatest_common_subconcept}(C, C1, \dots, Ck)$ znači da je pojam *C* najveći zajednički podpojam skupa pojmova $\{C1, \dots, Ck\}$.

Simetrično, definirani su predikati i pravila za operacije nadpojma rešetke kao što slijedi:

$\text{common_superconcept}(C, C1, \dots, Ck) \leftarrow \text{isa}(C1, C), \dots, \text{isa}(Ck, C)$

$\text{least_common_superconcept}(C, C_1, \dots, C_k) \leftarrow \text{common_superconcept}(C, C_1, \dots, C_k),$
 $\text{common_superconcept}(T, C_1, \dots, C_k), \text{isa}(C, T)$

Logički model rešetke L_{CR} se može automatski generirati na bazi gore prezentiranih pridruživanja. Ovaj proces je prikazan u slijedećem primjeru:

subconcept (romanopisci, književnici)
subconcept (Splitski književnici, književnici)
subconcept (Feljtonisti, književnici)
subconcept (Pjesnici, književnici)
subconcept (Splitski feltonisti, Splitski književnici)
subconcept (Splitski feljtonisti, feljtonisti)
subconcept (Splitski romanopisci, Splitski književnici)
subconcept (Splitski romanopisci, romanopisci)
subconcept (Splitski putopisci, Splitski književnici)
subconcept (Splitski putopisci, književnici)
subconcept (Splitski pjesnici, Splitski književnici)
subconcept (Splitski pjesnici, pjesnici)
subconcept (C1, romanopisci)
subconcept (C1, Feljtonisti)
subconcept (C2, Feljtonisti)
subconcept (C2, pjesnici)
subconcept (C3, romanopisci)
subconcept (C3, pjesnici)
subconcept (C4, Splitski putopisci)
subconcept (C4, Splitski romanopisci)
subconcept (C5, C2)
subconcept (C5, Splitski feljtonisti)
subconcept (C5, Splitski pjesnici)
subconcept (C6, Splitski pjesnici)
subconcept (C6, Splitski putopisci)
subconcept (C7, C3)
subconcept (C7, Splitski pjesnici)
subconcept (C7, Splitski romanopisci)
subconcept (C8, Splitski feljtonisti)

subconcept (C8, Splitski putopisci)
subconcept (C9, C3)
subconcept (C9, C2)
subconcept (C9, C1)
subconcept (C10, C8)
subconcept (C10, C4)
subconcept (C10, C1)
subconcept (C11, C7)
subconcept (C11, C6)
subconcept (C11, C4)
subconcept (C12, C8)
subconcept (C12, C6)
subconcept (C12, C5)
subconcept (C13, C9)
subconcept (C13, C10)
subconcept (C13, C11)
subconcept (C13, C12)

Pravila za zaključivanje aksioma i operacija rešetke se mogu upotrijebiti za odlučivanje o taksonomijskim odnosima između pojmova, kao i za izvođenje operacija rešetke. Primjerice, za pronalaženje najmanjeg zajedničkog nadpojma od pojma Splitski književnici, definira se upit `least_common_superconcept(X, Splitski književnici)`. Ukoliko se želi pronaći sve nadpojmove pojma C13, upit `isa(C13, X)` dati će popis osnovnih atoma za odgovor.

8.4.4.5 Predstavljanje ne-taksonomijskih odnosa upotrebom pravila

U svrhu definiranja ne-taksonomijskih odnosa treba definirati odgovarajući predikat. Za definiranje objektnih svojstava, može se upotrijebiti sljedeći predikat:

`hasproperty(imePojma, imeSvojstva)`

Primjerice, sljedeći osnovni atomi mogu biti dodatni na jednosavan ontologijski opis

`hasproperty(književnici, datumRođenja)`

Nasljeđivanje svojstava može biti predstavljen kao pravilo:
 $\text{hasproperty}(C1, X) \leftarrow \text{isa}(C1, C2), \text{hasproperty}(C2, X)$

Može se postaviti upit: $\text{hasproperty}(\text{književnici}, X)$ i primiti odgovor da on ima svojstvo datumRođenja .

Ontologijska svojstva poput npr. *part-of*, *related-to* itd. se mogu jednostavno predstaviti preko predikat, čime se mogu ontologijskom dodavati osnovni atomi :

$\text{partof}(C1, C2)$

$\text{related}(C1, C2)$

$\text{synonyms}(C1, C2)$, itd.

Ne-taksonomijska svojstva daju dodatne mogućnosti za zaključivanje o ontologiji.

8.4.4.6 Zaključivanje o ontologiji

Zaključivanje je važno kako bi se osigurala kvaliteta izrade ontologije. Može se upotrijebiti za pronalaženje protuslovnih pojmova, izvođenje neizravnih odnosa itd. Pravila za zaključivanje za aksiome i operacije rešetke se mogu upotrijebiti za određivanje taksonomijskih odnosa između pojmova, kao i za izvođenje operacija rešetke. Primjerice, kako bi se pronašao najmanji zajednički nadpojam skupa pojmova

{romanopisci, Splitski književnici}, definira se sljedeći upit:

$\text{least_common_superconcept}(X, \text{romanopisci}, \text{Splitski književnici})$. Odgovor je pojam književnici.

Ukoliko smo zainteresirani za pronalaženje najvećeg zajedničkog podpojma pojmova Splitski književnici i romanopisci, tada će odgovarajući upit biti:

$\text{greatest_common_subconcept}(X, \text{Splitski književnici}, \text{romanopisci})$. Odgovor je pojam Splitski romanopisci.

Možemo biti zainteresirani za sve nadpojmove od pojma Splitski pjesnici. Upit $isa(\text{Splitski pjesnici}, X)$ daje popis osnovnih atoma kao odgovor. U našem primjeru, to je popis sljedećih atoma $isa(\text{Splitski pjesnici}, \text{Splitski književnici})$, $isa(\text{Splitski pjesnici}, \text{pjesnici})$ i $isa(\text{Splitski pjesnici}, \text{književnici})$.

Zaključivanje o ne-taksonomijskim odnosima je postalo moguće zbog dodatnih pravila. Primjerice, može se dodati činjenica $hasproperty(\text{Splitski književnici}, jeMjestoRođenjaSplit)$.

Može se postaviti upit $hasproperty(X, jeMjestoRođenjaSplit)$ i primit će se odgovor $hasproperty(X, jeMjestoRođenjaSplit)$: $hasproperty(\text{Splitski pjesnici}, jeMjestoRođenjaSplit)$, $hasproperty(\text{Splitski romanopisci}, jeMjestoRođenjaSplit)$, $hasproperty(\text{Splitski feljtonisti}, jeMjestoRođenjaSplit)$, $hasproperty(\text{Splitski putopisci}, jeMjestoRođenjaSplit)$, $hasproperty(C4, jeMjestoRođenjaSplit)$, $hasproperty(C5, jeMjestoRođenjaSplit)$, $hasproperty(C6, jeMjestoRođenjaSplit)$, $hasproperty(C7, jeMjestoRođenjaSplit)$, $hasproperty(C8, jeMjestoRođenjaSplit)$, $hasproperty(C10, jeMjestoRođenjaSplit)$, $hasproperty(C11, jeMjestoRođenjaSplit)$, $hasproperty(C12, jeMjestoRođenjaSplit)$, $hasproperty(C13, jeMjestoRođenjaSplit)$.

Logička zaključivanja o ontologijskom izražavanju osiguravaju sredstvo za stvaranje konzistentnosti i točnosti ontologije.

8.5 Zaključak

Poglavlje o formalnoj pojmovnoj analizi je u prvenstveno nastojalo istaknuti bitne definicije za razumijevanje mjesta formalne pojmovne analize u metodologiji izgradnje ontologije. Osim teoretskog objašnjenja same metoda pojmovne analize i procesa njezine primjene prilikom dizajniranja ontologije, prikazan je i konkretan primjer izvršavanja FCA na 37 članaka područja hrvatska književnost iz Hrvatske enciklopedije i transformacija pojmovne rešetke u formalni jezik logike prvog reda. Prednost primjene FCA analize je generiranje novih i nepoznatih pojmova koji bi se teško ustanovili samo ručnom izradom ontologije jer tekstovi specifični za domenu ne uključuju bilo kakvu imensku frazu za označavanje ovih pojmova.

9 LATENTNA SEMANTIČKA ANALIZA

9.1 Problemi metoda automatskog indeksiranja i pretraživanja informacija

Temeljni nedostatak postojećih metoda pretraživanja informacija je nepodudarnost u riječima koje korisnici najčešće koriste i onih kojima je informacija koju traže indeksirana. Upitne su zapravo dvije stvari, tzv. sinonimija i polisemija. Sinonimija se koristi kako bi se opisala činjenica da postoji mnogo načina upućivanja na isti objekt. Korisnici će u različitim kontekstima, potrebama i jezičnim navikama opisati iste informacije koristeći različite termine. Učestalost sinonima teži smanjivanju odziva (engl. recall) sustava za pretraživanje informacija. Polisemija ukazuje na činjenicu da većina riječi ima više od jednog tj. različita značenja (homografija) koja poprima u različitim kontekstima. Polisemija je uzrok loše točnosti dobivenih rezultata pretraživanja.

Neuspjeh trenutnog automatskog indeksiranja u prevladavanju ovih problema se može opisati s tri čimbenika. Prvi čimbenik je nepotpunost načina identificiranja indeksnih termina. Termini upotrijebljeni za opis ili indeksiranje dokumenta uobičajeno sadrže samo dio termina koje će korisnici pokušati pogledati. Ovo je dijelom zato jer dokumenti ne sadrže sve termine koje će korisnici primijeniti, a ponekad i zato jer postupci za odabir termina namjerno izostave mnoge termine u dokumentu. Namjera bavljenja problemom sinonimije oslanja se na intelektualnu ili automatsku proširenost termina, odnosno izgradnju tezaurusa. To je vjerojatno prednost za savjesne i obrazovane korisnike koji mogu koristiti takve alate kako bi predložili dodatne termine za pretraživanje. Nedostatak za potpuno automatsku metodu je jer dodani termini mogu imati drugačije značenje od namjeravanog (učinak višeznačnosti), što dovodi do brze degradacije točnosti⁴⁰¹. Nije samo potencijalni problem dvosmislenost i nedostatak točnosti, nego brzo raste i problem identifikacije indeksnih termina koji nisu u tekstu dokumenta.

Drugi čimbenik je nedostatak adekvatne automatske metode za rješavanje polisema. Jedan opći pristup je upotreba kontroliranih rječnika i posrednika (čovjeka) koji će djelovati kao prevoditelj. Ne samo da je ovo rješenje veoma skupo, nego nije ni dovoljno učinkovito. Drugi pristup je Boolov presjek ili koordinacije s drugim terminima u svrhu smanjena dvosmislenosti značenja. Uspjeh je ozbiljno zapriječen korisnikovom nemogućnošću

⁴⁰¹ Sparck Jones, Karen. A statistical interpretation of term specificity and its application in retrieval. // Journal of Documentation 28, 1(1972), 11-21.

shvaćanja odgovarajućih ograničenja termina ako postoje, kao i činjenicom da se takvi termini ne moraju pojaviti u dokumentima ili ne moraju biti uključeni u indeksiranje.

Treći čimbenik je više tehničke prirode, povezan je s načinom na koje trenutno automatsko indeksiranje i sustavi za pretraživanje informacija zapravo rade. U takvim sustavima svaka vrsta riječi se tretira neovisno od ostalih (npr. van Rijsbergen⁴⁰²). Za takva podudaranja (ili nepodudaranja) dvaju termina koja se gotovo uvijek javljaju zajedno računa se težina dvaju podudaranja koja se rijetko nalaze u istom dokumentu. Takvo vrednovanje uspjeha, bilo Boolov ili koordinacijska razina pretraživanja, ne uzima u obzir redundancije, što može narušiti rezultate. Ovaj problem pogoršavaju korisnikove poteškoće u korištenju složenih termina u upitu za učinkovito proširivanje ili ograničenje pretraživanja.

Analizom i pretraživanjem informacija (eng. information retrieval) dominira pojava brzog rasta količine informacija i teksta, kao i vrijednosti tih informacija, što zahtijeva brz pristup i smislenu analizu. Konvencionalne tehnologije za pretraživanje, organiziranje i analizu informacija se nisu brzo razvile. Većina sustava za pretraživanje informacija se može nositi samo s riječima kao nizovima znakova ili ključnim riječima, iako korisnici trebaju alate koji mogu pronaći temeljne pojmove. Dizajneri su slijedili različite pristupe za integriranje nekog stupnja semantičke obrade u sustave za pretraživanje informacija: pomoćne strukture, lokalno zajedničko statističko pojavljivanje i latentno semantičko indeksiranje.

Kontrolirani rječnici ili pomoćne strukture, kao što su tezaursi, omogućuju u upite ugradnju širih, užih i srodnih termina. Jedan su od načina za prevladavanje nekih od najtežih ograničenja Boolovih upita slobodnog teksta ključnom riječju – višestrukih riječi koje imaju slična značenja (sinonime), kao i riječi koje imaju više od jednog značenja (višeznačajnice tj. poliseme). Kontrolirani rječnici mogu pridonijeti učinkovitosti i sveobuhvatnosti pretraživanja informacija i srodnih radnji analize teksta. No, ovaj pristup semantičke obrade najbolje funkcionira kada su teme usko definirane i terminologija standardizirana. Ne pristaje dobro pretraživanju informacija iz sve veće količine nestrukturiranih podataka koji sadrže tisuće jedinstvenih termina koji obuhvaćaju velik broj tema. Neki od nedostataka upotrebe pomoćnih struktura su: identificiranje korisnih kontroliranih rječnika zahtijeva puno ljudskog uključenosti i nadzora; jezik se brzo razvija, što zahtijeva stalno ažuriranje kontroliranih rječnika; kontrolirani rječnici često mogu predstavljati pogled na svijet svojih tvoraca, što dovodi do pojmovne neusklađenosti; dohvaćaju pogled na svijet u određenom vremenskom

⁴⁰² Van Rijsbergen, C. J. A theoretical basis for the use of co-occurrence data in information retrieval. *Journal of Documentation* 33, 2(1977), 106-119.

trenutku jer se teško mogu mijenjati kako se pojmovi mijenjaju u određenom tematskom području.

Tijekom godina izgrađene su dodatne pomoćne strukture općeg interesa, kao što su npr. veliki sinonimni skupovi WordNet. Dolazi do stvaranja podatakovnih modela za predstavljajne skupova pojmova unutar domene (ontologije), koji mogu uključiti odnose među terminima. Stoga je iz svega navedenog uočljivo kako je sposobnost rada s tekstom na semantičkoj osnovi temelj za moderne sustave pretraživanja informacija.

Modeli vektorskog prostora su razvijeni kako bi eliminirali mnoge probleme povezane s uparujućim leksičkim tehnikama (engl. lexical matching techniques). Konkretno, u slučaju polisema leksičkim uparujućim tehnikama je teško razlikovati dva dokumenta koji dijele određenu riječ, ali se drugačije upotrebljava tj. leksičke tehnike ne omogućuju razumijevanje konteksta u kojem je riječ korištena. Također, srodni dokumenti ne moraju koristiti istu terminologiju za opisivanje njihovih zajedničkih pojmova jer ima više načina za opis određenog pojma (sinonomi). Koristeći terminologiju jednog dokumenta u postavljenom upitu, neće se pretražiti drugi srodni dokumenti. Čak i ako zbirka sadrži srodne dokumente koristeći u upitu terminologiju različitu od one upotrijebljene u srodnim dokumentima sadržanim u zbirci neće se moći pretražiti svi dokumenti upotrebljavajući tehnike leksičkih uparivanja⁴⁰³. Vidljivo je kako je latentno semantičko indeksiranje (LSI)⁴⁰⁴ jedan od najučinkovitijih pristupa u području semantičke obrade, posebno za pretraživanje informacija iz velike količine nestrukturiranih podataka, te se zbog načina predstavljanja termina i dokumenata u termin-dokument prostoru smatra modelom vektorskog prostora pretraživanja informacija. Sljedeći odjeljci ovog poglavlja definiraju i objašnjavaju osnovne pojmove poput *indeksni termini*, *prikaz termina u dokumentu* i mjerenje *težine termina* u zbirci dokumenata (Odjeljak 9.2), mjere pronalaženja sličnih dokumenata (Odjeljak 9.3), opći model vektorskog prostora (Odjeljak 9.3.1) te se teorijski prikazuje LSI metoda (Odjeljak 9.3.1.1), kao i njezina primjena na primjeru enciklopedijskih članaka iz područja hrvatske književnosti kako bi se spoznala učinkovitost i svrha LSI metode u izgradnji ontologije određenog područja (Odjeljak 9.3.1.1.3).

⁴⁰³Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W. Using linear algebra for intelligent information retrieval. *SIAM Review* 37, 4(1995), :573–595.

⁴⁰⁴ U literaturi se često naziva i latentna semantička analiza (engl. latent semantic analysis, LSA)

9.2 Izračunavanje težine termina

9.2.1 Indeksni termini ili ključne riječi

LSI metodom se nastoje poboljšati prethodno navedeni problemi koji se susreću u sustavima automatskog pretraživanja informacija. U svrhu boljeg razumijevanja LSI metode Odjeljak 9.2 definira pojmove *indeksni termini (ključne riječi)*, objašnjava *prikaz termina u dokumentu* te prikazuje osnovne mjere izračunavanja *težine termina*. Modeli pretraživanja informacija uzimaju u obzir mogućnost opisa svakog dokumenta skupom reprezentativnih ključnih riječi tzv. indeksnim terminima. Indeksni termin je riječ ili niz riječi u dokumentu, odnosno indeksni termin je bilo koja riječ iz zbirke dokumenata. Navedeni pristup je preuzet od dizajnera tražilica za pretraživanje. Pristup knjižničara i informacijskih stručnjaka definira indeksne termine kao prethodno odabrane grupe riječi koje predstavljaju ključni pojam ili temu dokumenta. Dakle, prethodno odabrane grupe indeksnih termina mogu se upotrijebiti za sažimanje sadržaja dokumenta.

Definicija 6: Neka je t broj indeksnih termina u zbirci dokumenata i neka je k_i opći indeksni termin. $V = \{k_1, \dots, k_t\}$ je skup svih različitih indeksnih termina u zbirci i obično se naziva rječnikom zbirke. Veličina rječnika je t . Rječnik je važan sastavni dio zbirke jer on identificira sve indeksne termine, a rastom zbirke raste i veličina rječnika.

Sažimanje sadržaja dokumenata i upita kroz skup indeksnih termina može voditi prema lošoj kvaliteti pretraživanja zbog dva razloga: (1) mnogi nesrodni dokumenti mogu biti uključeni u skup dokumenata koji se dobiva kao rezultat pretraživanja na postavljeni upit i (2) relevantni dokumenti koji nisu indeksirani s bilo kojom ključnom riječi neće biti pretraženi. Glavni razlog ovih dviju posljedica je svojstvena dvosmislenost povezana s postupkom pretraživanja informacija koji se temelji na skupovima ključnih riječi.

Tekst dokumenta sadrži pojmove (predstavljaju stvari svijeta) i odnose među njima. Stoga, proces pridruživanja dokumenata na određeni upit se može temeljiti na pridruživanju pojma umjesto na pridruživanju indeksnog termina, što bi omogućilo pretraživanje dokumenata čak i kada oni nisu indeksirani indeksnim terminom u upitu. Primjerice, dokument može biti pretražen jer dijeli pojmove s drugim dokumentom koji je relevantan za određeni upit. Latentno semantičko indeksiranje je pokušaj rješavanja ovog problema. Glavna

zamisao u modelu latentnog semantičkog indeksiranja je pridružiti svaki dokumentni i upitni vektor dimenzionalnom prostoru sastavljenom od pojmova, što se postiže pridruživanjem indeksnih termina u ovaj dimenzionalni prostor pojmova te uobličavanjem dokumenata i upita u termine tih pridruživanja. Tvrdnja je da pretraživanje u reduciranom prostoru pojmova može biti superiornije u odnosu na pretraživanje u prostoru indeksnih termina.⁴⁰⁵

9.2.2 Prikaz termina u dokumentu (engl. Term-Document Representation)

Pojavljivanjem termina u dokumentu uspostavlja se odnos između njih. Ovaj termin-dokument odnos se može mjeriti, primjerice, frekvencijom termina u dokumentu. To se u oblik matrice zapisuje:

$$\begin{array}{c} k_1 \\ k_2 \\ k_3 \end{array} \begin{array}{cc} d_1 & d_2 \\ \left[\begin{array}{cc} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{array} \right] \end{array}$$

gdje svaki $f_{i,j}$ element stoji za frekvenciju termina k_i u dokumentu d_j .⁴⁰⁶

Matricom možemo predstaviti bazu podataka koja je skup dokumenata opisanih određenim terminima.

Definicija 7⁴⁰⁷: Matrica baze podataka. Neka je $\mathbf{A} \in \mathbb{R}^{p \times d}$, gdje je d ukupan broj dokumenata a p ukupan broj termina u bazi podataka. Matrica

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2d} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{id} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pj} & \dots & a_{pd} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_j \quad \dots \quad \mathbf{a}_d],$$

⁴⁰⁵ Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier. Modern information retrieval : the concepts and technology behind search. 2nd ed. Harlow [etc.] : Pearson Education, 2011. Str. 61-62.

⁴⁰⁶ Ibid.

⁴⁰⁷ Ivančević, Zvonimir; Jelić, Slobodan. Primjena SVD rastava matrice u dohvatima informacija i kompresiji slike. Dostupno na URL: http://www.mathos.unios.hr/~sjelic/papers/SVD_ir_ic.pdf (16.02.2015.).

gdje je

$$a_{ij} = \begin{cases} 1 & , \text{ ako } i\text{-ti pojam opisuje } j\text{-ti dokument} \\ 0 & , \text{ inače} \end{cases} ,$$

naziva se matrica baze podataka. Ova matrica se naziva termin-dokument matrica te sadrži punoznačne termine iz cijele zbirke kao redove i dokumente zbirke kao stupce. Dakle, broj stupaca matrice jednak je broju dokumenata, a broj redaka matrice jednak je broju termina. Opisuje pojavljivanje punoznačnih termina u svakom dokumentu zbirke. Čitajući novi tekstualni dokument, novi termini se dodaju u matrici kao redovi, a njihove frekvencije se dodaju u odgovarajuće stupce. Sustav treba provjeriti jesu li ovi novi termini gramatički oblici od nekih prethodnih termina.

Za heterogene tekstualne zbirke, odnosno zbirke koje predstavljaju mnogo različitih konteksta (ili tema), kao što su novine i enciklopedije, broj termina (p) je obično mnogo veći od broja dokumenata ($d \ll p$). Na webu, međutim, situacija je obrnuta. Dakle, zbirka od d dokumenata indeksiranih s p terminima (ključnim riječima) može biti predstavljena kao $p \times d$ termin – dokument matrica $A = [a_{ij}]$. Indeksiranje u određenom dokumentu predstavlja označavanje samo onih termina koji imaju ključnu važnost u opisu semantike tog dokumenta. Matrica A predstavlja skup *termin-dokument* povezivanja. Elementi matrice su frekvencije pojavljivanja pojedine riječi u dotičnom dokumentu $A = a_{ij}$, gdje je a_{ij} frekvencija pojavljivanja riječi i u dokumentu j . Ne-nulti element a_{ij} , gdje je

$A = [a_{ij}]$, ukazuje ne samo da se termin i pojavljuje u dokumentu j , nego također i broj puta pojavljivanja u tom dokumentu. Budući je broj termina u danom dokumentu obično daleko manji nego broj termina u cijeloj zbirci dokumenata, A je uobičajeno veoma prorijeđena⁴⁰⁸. Za računanje elemenata matrice umjesto frekvencije moguće je koristiti i neku od težinskih frekvencija. Dakle, stvarne vrijednosti dodijeljene elementima termin – dokument matrice $A = a_{ij}$ su uobičajeno težinske frekvencije kojom se termin i pojavljuje u dokumentu j ⁴⁰⁹, za razliku od sirovih brojeva terminskog pojavljivanja (unutar dokumenta ili na cijeloj zbirci). Glavna svrha frekvencije termina je poboljšanje izvođenja pretraživanja informacija, što se odnosi na sposobnost pretraživanja relevantnih informacija (odziv) i odbacivanju nerelevantnih informacija (točnost). Odziv se mjeri kao omjer broja relevantnih dokumenata

⁴⁰⁸ Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W., loc cit. (bilj. 403)

⁴⁰⁹ Ibid.

pretraženih za upit i broja dokumenata koji su relevantni za taj upit u cijeloj zbirci dokumenata, a točnost se mjeri kao omjer broja relevantnih dokumenata koje je sustav pretražio i ukupnog broja pretraženih dokumenata. Modeli pretraživanja informacija obično izražavaju učinkovitost izvedbe sustava pretraživanja informacija točnošću i odzivom. Poželjan IR sustav je onaj koji postiže visoku točnost uz visoki odziv (ako ne i sve). Jedan način za poboljšanje odziva za dani upit je koristiti riječi s visokim frekvencijama, odnosno, one koje se pojavljuju u mnogim dokumentima u zbirci. Nasuprot tome, dostizanje visoke točnosti može zahtijevati korištenje vrlo specifičnih termina ili riječi koje će odgovarati najrelevantnijim dokumentima u zbirci. Nema sumnje kako mora biti učinjena neka vrsta dogovora kako bi se postigao dovoljan odziv bez slabe točnosti. Oboje, i točnost i odziv se izražavaju kao vrijednost između 0 i 1. Optimalni sustav pretraživanja informacija će osigurati točnost i odziv na vrijednosti 1, iako se točnost smanjuje s većim odzivom u realnim sustavima⁴¹⁰. U cilju ostvarivanja pretraživanja relevantnih informacija (odziv) i odbacivanja nerelevantnih informacija (točnost), razvijeno je nekoliko različitih načina izračuna vrijednosti termina, što se postiže pridjeljivanjem težina indeksnim terminima tzv. težina termina, koje će se prikazati u Odjeljku 9.2.3.

9.2.3 Težina termina

Uočljivo je kako nisu svi termini jednako korisni za opisivanje sadržaja dokumenta. Odlučivanje o važnosti termina za sažimanje sadržaja dokumenta nije jednostavno pitanje. Unatoč navedenoj poteškoći, postoje svojstva indeksnih termina koja su lako mjerljiva i koja su korisna za vrednovanje važnosti termina. Za ilustraciju, ukoliko postoji zbirka s nekoliko tisuća dokumenata, riječ koja se pojavljuje u svakom od sto tisuća dokumenata nije korisna kao indeksni termin jer nam ništa ne govori o tome za koje dokumente korisnik može biti zainteresiran. S druge strane, riječ koja se pojavljuje u samo pet dokumenata zbirke je poprilično korisna jer znatno sužava skup dokumenata koji mogu biti interesantni korisniku. Pridruživanje težina terminima zasniva se na sljedećim pretpostavkama: izrazi koji se pojavljuju u jako puno ili jako malo dokumenata su manje važni od onih koji se pojavljuju umjereno u skupu dokumenata, dugački dokumenti nisu važnijih od kratkih i višestruka pojavljivanja izraza u dokumentu nisu manje važna od jednostrukih. Stoga, treba biti jasno

⁴¹⁰Information Retrieval: Data Structures & Algorithms. / William B. Frakes, Ricardo Baeza-Yates (editors). Englewood Cliffs, New Jersey : Prentice Hall, , 1992. Str. 102.

kako različiti indeksni termini imaju različitu važnost kada se koriste za opis sadržaja dokumenta te se za stvaranje težina indeksnih termina uzima u obzir koliko je važan termin za opisivanje dokumenta ili podskupa dokumenata zbirke. To se ilustrira jednostavnim primjerom, npr. ukoliko postoji zbirka od oko tisuću članaka o John Lennonu te ukoliko korisnik specificira upit „John Lennon“, kako se mogu razvrstati dokumenti da bi se prvo prikazali oni koji su najbliži interesu korisnika? Dakle, dokumenti se ne mogu razvrstati u nedostatku dodatnih informacija. Upit kojega je korisnik postavio jednostavno ne nosi nijednu vrijednu informaciju u kontekstu ove određene zbirke. Stoga se kaže da je vrijednost informacija povezana s upitom 0. Ovaj veoma praktičan primjer ne rasvjetljuje kako dizajnirati algoritam za rangiranje nego razjašnjava važnost u pogledu težine termina – različiti termini nisu jednako važni i trebaju se različito odvagati. Učinak je dostignut kroz dodjeljivanje brojčane težine svakom indeksnom terminu dokumenta. Težina termina je najčešće korišten pristup za poboljšanje učinka pretraživanja sustava automatskog indeksiranja, kao i za poboljšanje rangiranja u vektorskom modelu.

Definicija 8: neka je t broj indeksnih termina u zbirci, N je ukupni broj dokumenata i neka je $M = [a_{ij}]$ termin-dokument matrica s t redovima i N stupcima. Svakom elementu a_{ij} ove matrice dodijeli težinu $w_{i,j}$ povezanu s termin-dokument parom (k_i, d_j) . Ova težina $w_{i,j}$ se može izvesti upotrebom TF-IDF težina klasičnog vektorskog modela, mjerom o kojoj će više biti riječi u Odjeljku 8.2.7.

Definicija 9: Za prikazivanje važnosti termina, težina $w_{i,j}, w_{i,j} > 0$, je povezana sa svakim indeksnim terminom k_i dokumenta d_j u zbirci. Za indeksni termin k_i koji se ne pojavljuje u dokumentu težina $w_{i,j}$ ima vrijednost $w_{i,j} = 0$.

Težina $w_{i,j}$ mjeri važnost indeksnog termina za opisivanje sadržaja dokumenta. Dodjeljivanje težine indeksnim terminima dozvoljava izračunavanje brojčanog ranga za svaki dokument u zbirci s obzirom na dano pitanje, što vodi poboljšanju rezultata. Nadalje, težine su pod utjecajem dokumenata u zbirci. Kako bi se izračunala važnost termina trebaju se izračunati težine koje odražavaju važnost termina u zbirci i u svakom posebnom dokumentu. Ove težine ovise o frekvenciji pojavljivanja termina unutar dokumenata, što se definira kao što slijedi.

Definicija 10: Neka je $f_{i,j}$ frekvencija pojavljivanja indeksnog termina k_i u dokumentu d_j , odnosno, broj puta kojim se termin k_i pojavljuje u tekstu dokumenta d_j . Ukupna frekvencija F_i od termina k_i u zbirci je zbroj pojavljivanja frekvencije termina u svim dokumentima, odnosno

$$F_i = \sum_{j=1}^N f_{i,j}$$

gdje je N broj dokumenata u zbirci. Frekvencija termina k_i po dokumentu je broj dokumenata u kojima se on pojavljuje i jednostavno je naznačen kao n_i . Uočljivo je kako je $n_i \leq F_i$.⁴¹¹

9.2.4 Termin – termin korelacije

U prethodnom izlaganju se pretpostavlja da su težine indeksnih termina međusobno neovisne. Ovo znači da poznavanje težine $w_{i,j}$ povezane s parom (k_i, d_j) ništa ne govori o težini $w_{i+1,j}$ povezane s parom (k_{i+1}, d_j) . Ovo je pojednostavljeno jer pojavljivanja indeksnih termina u dokumentu nisu nekolerirajuća. Primjerice, neka se termini *računalo* i *mreža* upotrebljavaju za indeksiranje dokumenata o *računalnim mrežama*. Često, pojavljivanje jednog od ovih termina privlači i pojavljivanje drugog termina u dokumentu. Dakle, oni koreliraju te bi stoga i njihove težine trebale odražavati ove korelacije. Uzimajući u obzir termin-termin korelacije, može se izračunati korelacijska matrica. Ključni koncepti u ovom slučaju se mogu formalizirati kako slijedi.

Definicija 11: Neka $M = [a_{ij}]$ bude termin-dokument matrica s t redovima i N stupcima, gdje je $a_{ij} = w_{i,j}$, odnosno svaka stavka ij u matrici je zadana težinom povezanom s termin-dokument parom (k_i, d_j) . Budući je M^T ispremještana od M , matrica $C = M \cdot M^T$ je termin-termin korelirajuća matrica. Svaki element $c_{u,v} \in C$ izražava korelaciju između termina k_u i k_v , zadanu s

⁴¹¹ Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier. Modern information retrieval : the concepts and technology behind search. 2nd ed. Harlow [etc.] : Pearson Education, 2011. Str. 66-67.

$$c_{u,v} = \sum_{d_j} w_{u,j} \times w_{v,j}$$

Termin-termin korelacijska matrica C uspostavlja odnos između bilo koja dva termina k_u i k_v temeljeći se na njihovom zajedničkom pojavljivanju unutar dokumenata zbirke, kao što je prikazano na Slici 72. Ovaj odnos se mjeri korelacijskim faktorom $c_{u,v}$. Što je veći broj dokumenata u kojima se termini k_u i k_v međusobno pojavljuju to je jača ova korelacija.⁴¹²

$$\begin{array}{c}
 \begin{array}{cc} d_1 & d_2 \end{array} \\
 \begin{array}{c} k_1 \\ k_2 \\ k_3 \end{array} \left[\begin{array}{cc} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{array} \right] \\
 \mathbf{M}
 \end{array}
 \quad \times \quad
 \begin{array}{ccc} k_1 & k_2 & k_3 \\
 \begin{array}{c} d_1 \\ d_2 \end{array} \left[\begin{array}{ccc} w_{1,1} & w_{2,1} & w_{3,1} \\ w_{1,2} & w_{2,2} & w_{3,2} \end{array} \right] \\
 \mathbf{M}^T
 \end{array}
 \end{array}$$

↓

$$\begin{array}{ccc} k_1 & k_2 & k_3 \\
 \begin{array}{c} k_1 \\ k_2 \\ k_3 \end{array} \left[\begin{array}{ccc} w_{1,1}w_{1,1} + w_{1,2}w_{1,2} & w_{1,1}w_{2,1} + w_{1,2}w_{2,2} & w_{1,1}w_{3,1} + w_{1,2}w_{3,2} \\ w_{2,1}w_{1,1} + w_{2,2}w_{1,2} & w_{2,1}w_{2,1} + w_{2,2}w_{2,2} & w_{2,1}w_{3,1} + w_{2,2}w_{3,2} \\ w_{3,1}w_{1,1} + w_{3,2}w_{1,2} & w_{3,1}w_{2,1} + w_{3,2}w_{2,2} & w_{3,1}w_{3,1} + w_{3,2}w_{3,2} \end{array} \right]
 \end{array}$$

Slika 72: Termin-dokument matrica, njezino premještanje te termin-termin korelacijska matrica za mali primjer zbirke sa samo dva dokumenta i tri termina (preuzeto iz: (Baeza-Yates, R.; Ribeiro-Neto, B., 2011.).

9.2.5 Frekvencija termina

Frekvencija termina (engl. *Term Frequency*, *tf*) se uzima kao mjera težine termina jer se pretpostavlja kako će se riječi bitne za dokument češće spominjati u tekstu. Dakle, što je termin frekventniji u dokumentu, to se procjenjuje važnijim za dokument. Frekvencija termina prikazuje koliko puta se dani termin pojavljuje u određenom dokumentu. Ispravnost dobivenih rezultata frekvencije termina ovisi o dužini teksta. Kod kraćih tekstova riječi se pojavljuju samo nekoliko puta, što može navesti na krive zaključke. Rijedak termin će imati znatno veći utjecaj na kraće tekstove. Zato je frekvencija termina znatno učinkovitija mjera težine kod dužih tekstova.

⁴¹² Ibid.

Pretpostavimo postojanje korpusa koji sadrži tri uzorka dokumenata, kao što je prikazano u primjeru:

'a' : " Tako je Dečak Mirko osnivač informativnog dnevnika *Novosti* . \

Tomislav Golubović je isto tako prvi urednik informativnog dnevnika *Večernji vjesnik* . "

'b' : " Bio je u uredništvu zbirke *Lirski dnevnik i Poezija*."

'c' : " Pjesme, vedrih ugođaja, objavljivao je u mnogobrojnim periodicima. \

Samostalne su mu zbirke *Lirski dnevnik i Poezija*.

Frekvencija termina može se jednostavno prikazati kao broj puta koliko se termin pojavljuje u tekstu, ali sve je češće slučaj normalizacije uzimajući u obzir ukupan broj termina u tekstu, tako da ukupni rezultat izračunat za duljinu dokumenata odgovara frekvenciji termina. Primjerice, “dnevnik” se pojavljuje dva puta u korpusu ['a'] i samo jednom u korpusu ['b'], tako da će korpus ['a'] proizvesti veći rezultat, ukoliko bi frekvencija bila jedini kriterij. Međutim, ako se normalizira za duljinu dokumenata, korpus ['b'] će imati nešto veći rezultat frekvencije termina za “dnevnik” (1/9) od korpusa ['a'] (2/19), jer je korpus ['b'] kraći od korpusa ['a'] — iako se “dnevnik” pojavljuje češće u korpusu ['a'].

Uobičajena tehnika izračunavanja složenog upita, kao što je “informativni *dnevnik*” je zbroj rezultata frekvencije termina za svaki od termina u upitu u svakom dokumentu te se na taj način dobivaju dokumenti rangirani po zbroju rezultata frekvencije termina. Tablica 16 prikazuje normalizirane rezultate frekvencije termina za svaki dokument na postavljeni upit “informativni *dnevnik*” na uzorcima korpusa.

Tabela 16: Rezultati frekvencije termina za „informativni dnevnik”

Dokument	tf (informativni)	tf (dnevnik)	Suma
korpus ['a']	2/19	2/19	4/19 (0,2105)
korpus ['b']	0	1/9	1/9 (0.1111)
korpus ['c']	0	1/16	1/16 (0.0625)

Kao rezultat frekvencije termina za navedeni primjer dobiva se korpus ['a'] (dokument koji smo očekivali da će vratiti), budući je korpus ['a'] jedini koji sadrži složeni izraz “informativni *dnevnik*”. Međutim, mogu se pojaviti brojni problemi jer model izračunavanja

frekvencije termina promatra svaki dokument kao neuređenu zbirku riječi. Primjerice, upiti “*dnevnik informativni*” ili “*dnevnik informativni hrvatski*” vratili bi točno isti rezultat kao upit za “informativni dnevnik”, iako se ni jedan od tih izraza ne pojavljuje u uzorcima rečenica. Dodatno, brojni su scenariji koji bi mogli jednostavno ilustrirati prilično siromašne rezultate tehnike rangiranja frekvencije termina s obzirom da se kontekst oko izraza ne uzima u obzir izračunavanjem.

Razmatrajući samu frekvenciju termina ističe se problem izračunavanja frekvencije na razini dokumenta jer se ne uzimaju u obzir veoma frekventne tzv. stop riječi⁴¹³, koje su zajedničke mnogim dokumenatima. Drugim riječima, svim terminima je dana jednaka važnost, bez obzira na njihovu stvarnu važnost. Primjerice, “*tako lirski dnevnik*”, sadrži stop riječ “tako” koja iskrivljava rezultate ukupne frekvencije termina u korist korpusa ['a'] jer se dva puta pojavljuje u tom dokumentu, kao i “dnevnik”. Nasuprot tome, u korpusu ['c'] samo jednom se pojavljuje “lirski” i “dnevnik”.

Slijedom toga, rezultati će se razbiti kao što je prikazano u Tablici 17, korpus ['a'] je rangiran kao važniji od korpusa ['c'], premda intuicija može dovesti do vjerovanja da idealni rezultati upita vjerojatno ne bi trebali ispasti na takav način. Međutim, na sreću korpus ['b'] i dalje je najviše rangiran).

Tabela 17: Rezultati frekvencije termina za “tako lirski dnevnik”

Dokument	tf (<i>tako</i>)	tf (lirski)	tf (dnevnik)	Suma
korpus ['a']	2/19	0	2/19	4/19 (0.2105)
korpus ['b']	0	1/9	1/9	2/9 (0.2222)
korpus ['c']	0	1/16	1/16	1/8 (0.125)

9.2.6 Inverzna frekvencija dokumenta

Tekst uvijek sadrži mnoge uobičajene riječi koje ne nose nikakvo značenje za sadržaj, čak i nakon eliminacije stop riječi (čestice, usklici i sl.). Alati poput NLTK osiguravaju liste stop riječi koje se mogu upotrijebiti za filtriranje termina, ali treba imati na umu da pojedini termini mogu izbjeći i najboljim listama stop riječi te još uvijek biti zajednički specijaliziranim domenama. Dakle, što se neki termin pojavljuje u više tekstova, to je njegovo

⁴¹³ Stop riječi (engl. *Stopwords*) su riječi koje se često pojavljuju u tekstu, ali su neutralne u odnosu na temu teksta, npr. članovi, prijedlozi, konjunkcije itd.

značenje manje važno. Ukoliko se termin javlja u jako malom broju tekstova, on može vrlo dobro te tekstove odvajati od ostalih. Iako se može prilagođavati lista stop riječi sa znanjem domene, *inverzna frekvencija dokumenta* (engl. *inverse document frequency*) izračunava pojavljivanje zajedničkih termina preko skupa dokumenata promatrajući ukupan broj dokumenata u kojemu se termin upita ikada pojavio. Zato težina termina ovisi obrnuto proporcionalno o broju tekstova u kojima se termin javlja, a naziva se inverzna frekvencija dokumenata (*idf*). Inverzna frekvencija dokumenta proizvodi višu vrijednost ukoliko je termin rijedak u korpusu, što pomaže rješavanju problema sa stop riječima. Primjerice, upit za “zbirka” u prethodno navedenom jednostavnom skupu dokumenata treba vratiti niži rezultat inverzne frekvencije dokumenta nego upit za “pjesma” jer se “zbirka” pojavljuje u svakom dokumentu dok se “pjesma” pojavljuje u samo jednom. IDF rezultat za pojedini termin je logaritam od kvocijenta koji je definiran s N ukupnim brojem dokumenata u korpusu podijeljen s n_i brojem dokumenata u korpusu koji sadrže termin i :

$$idf^i = \log \left(\frac{N}{n_i} \right),$$

Tablica 18 prikazuje rezultat inverzne frekvencije dokumenta koji odgovara rezultatu frekvencije termina u Tablici 17. Prilikom usporedbe Tablice 17 i Tablice 18 treba imati na umu kako je rezultat frekvencije termina izračunat na dokument osnovi, a inverzne frekvencije dokumenta na temelju cijelog korpusa. Svrha *idf* mjere je njezino djelovanje kao normalizatora zajedničkih riječi preko cijelog korpusa razmatrajući ukupan broj dokumenata u kojima se upitni termin ikada pojavio.

Dakle, inverzna frekvencija dokumenta *idf* mjeri koliko je termin rijedak u korpusu te ukoliko je termin rijedak u cjelokupnom korpusu, smatra se vrlo relevantnim za dokument. Shodno tome, ako je termin veoma čest u cjelokupnom korpusu tj. pojavljuje se u većini dokumenata, ne podrazumijeva se da je naročito reprezentativan za dokument.

Tabela 18: Rezultati inverzne frekvencije dokumenta za termine koji se pojavljuju u izrazima „informativni dnevnik“ i “tako lirski dnevnik”

idf (informativni)	idf (dnevnik)	idf (tako)	idf (lirski)
$1 + \log (3/1) = 2.0986$	$1 + \log (3/3) = 1.0$	$1 + \log (3/1) = 2.0986$	$1 + \log (3/2) = 1.4055$

9.2.7 Frekvencija termina i inverzna frekvencija dokumenta - TF-IDF

Frekvencija termina i inverzna frekvencija dokumenta (tf-idf) je statistička mjera koja se često koristi u pretraživanju informacija i rudarenju podataka tj. postavljanju upita korpusu izračunavanjem normaliziranih rezultata koji izražavaju važnost termina u dokumentima. Dakle, TF-IDF je jedna od osnovnih tehnika koja se koristi za opisivanje i uspoređivanje dokumenata u vektorskom prostoru za pretraživanje relevantnih dokumenata iz korpusa (zbirke). Mjeri svaku komponentu vektora dokumenta, gdje je svaki od vektora povezan s jednim terminom korpusa.

Matematički TF-IDF je produkt frekvencije termina i inverzne frekvencije dokumenta, $tf-idf = tf * idf$, gdje termin tf predstavlja važnost termina u određenom dokumentu, a idf predstavlja važnost termina u odnosu na cijeli korpus. Množenjem ovih kriterija zajedno dobiva se mjera težine koja uzima u obzir oba faktora te predstavlja vrijednost koja pokazuje koliko je neki termin relevantan za pojedini dokument ili cjelokupni korpus. TF-IDF mjera računa se formulom:

$$tf \times idf^i = tf^i \cdot \log \left(\frac{N}{n_i} \right)$$

Najbolji termini (pojmovi) su oni koji se pojavljuju često unutar teksta, ali rijetko u drugim dokumentima. Termini koji se rjeđe pojavljuju u cijeloj zbirci dokumenata su bolji jer su pouzdaniji. Dakle, bolji konačni rezultat se postiže danom većom frekvencijom termina u danom dokumentu i niskom frekvencijom termina u cjelokupnom korpusu. Tako se postiže filtriranje učestalih termina.

TF-IDF mjera omogućuje računanje rezultata za višeterminske upite uzimajući u obzir frekvenciju termina pojavljivanih u dokumentu, duljinu dokumenta u kojemu se pojedini termin pojavljuje i ukupnu jedinstvenost termina preko dokumenata u cijelom korpusu. Iza frekvencije termina i inverzne frekvencije dokumenta mogu se kombinirati pojmovi u jedan račun množeći ih zajedno, tako da je $TF-IDF = TF * IDF$. Tablice 19, 20 i 21 prikazuju računske radnje uključene u izračun TF-IDF tri uzorka upita koji uključuju četiri različita termina, a dobiveni veći TF-IDF rezultati ukazuju na relevantnije upite:

- “*dnevnik*”
- “*informativni dnevnik*”

- “tako lirski dnevnik”

Tabela 19: Izračun TF i IDF koji su potrebni za izračun TF-IDF uzorak upita

Dokument	tf (informativni)	tf (dnevnik)	tf (<i>tako</i>)	tf (lirski)
korpus ['a']	0.1053	0.1053	0.1053	0
korpus ['b']	0	0.1111	0	0.111
korpus ['c']	0	0.0625	0	0.0625

idf (informativni)	idf (dnevnik)	idf (<i>tako</i>)	idf (lirski)
2.0986	1.0	2.099	1.4055

Tabela 20: Rezultati TF-IDF mjere za pojedini termin

Dokument	tf-idf (informativni)	tf-idf (dnevnik)	tf-idf (<i>tako</i>)	tf-idf (lirski)
korpus ['a']	$0.1053 * 2.0986 = 0.2209$	$0.1053 * 1.0 = 0.1053$	$0.1053 * 2.099 = 0.2209$	$0 * 1.4055 = 0$
korpus ['b']	$0 * 2.0986 = 0$	$0.1111 * 1.0 = 0.1111$	$0 * 2.099 = 0$	$0.111 * 1.4055 = 0.1562$
korpus ['c']	$0 * 2.0986 = 0$	$0.0625 * 1.0 = 0.0625$	$0 * 2.099 = 0$	$0.0625 * 1.4055 = 0.0878$

Tabela 21: Zbroj TF-IDF vrijednosti za višeterminske uzorke upita

Upit	korpus ['a']	korpus ['b']	korpus ['c']
dnevnik	0.1053	0.1111	0.0625
informativni dnevnik	$0.2209 + 0.1053 = \mathbf{0.3262}$	$0 + 0.1111 = 0.1111$	$0 + 0.0625 = 0.0625$
tako lirski dnevnik	$0.2209 + 0 + 0.1053 = \mathbf{0.3262}$	$0 + 0.1562 + 0.1111 = 0.2673$	$0 + 0.0878 + 0.0625 = 0.1503$

Korpus dokumenta ['b'] je pobjednik za upit „dnevnik“, s korpusom ['a'] koji je odmah iza njega. U ovom slučaju, odlučujući faktor je bila duljina korpusa ['b'] koji je mnogo manji od korpusa ['a'] – normalizirani TF rezultat okreće rezultate u korist korpusa ['b'] za njegovo jedno pojavljivanje „dnevnik“, iako se „dnevnik“ u korpusu ['a'] pojavljuje dva puta. Za upit „informativni dnevnik“, jasan i odgovarajući pobjednik je korpus ['a']. Međutim, on također izlazi na vrh za upit „tako lirski dnevnik“.

Postavljanje upita višestrukih termina se koristi za postizanje rezultata u kojima su dokumenti poredani po relevantnosti. Rangiranje traženog termina po relevantnosti je ogromna korist kada se analiziraju nestrukturiranih tekstualnih podataka. Prilikom TF-IDF izvedbe treba imati na umu kako apsolutne vrijednosti rezultata nisu zapravo veoma važne nego je bitna sposobnost pronalaženja i sortiranja dokumenta po relevantnosti. Rezultat postavljenog upita se prikazuje na uzorku rezultata upita za “*pjesma u prozi*,” u obliku kako slijedi:

Javio se zbirkom pjesama u prozi *Primjeri* (1965), pod utjecajem...

Link: <http://www.enciklopedija.hr/Natuknica.aspx?ID=39068>

Score: 0.0805961208217

Nakon dobivenih rangiranih rezultata, sljedeći prirodni korak je otkrivanje ostalog sličnog sadržaja koji bi zadovoljio interes korisnika, što će se objasniti u nastavku.

9.3 Pronalaženje sličnih dokumenata

Nakon postavljanja upita i otkrivanja dokumenata od interesa, sljedeće što korisnik može poželjeti je pronalaženje sličnih dokumenata. Dok TF-IDF može osigurati način za sužavanje korpusa na temelju traženih termina, kosinusna sličnost je jedna od najčešćih tehnika za usporedbu dokumenta u pretraživanju i obradi teksta. Kako bi se izračunala sličnost između dvaju dokumenata, potrebno je dobiti terminski vektor za svaki dokument i izračunati skalarni umnožak jediničnih vektora za te dokumente. Odgovarajući terminski vektori se izračunavaju za dani par dokumenata dodjeljivanjem TF-IDF rezultata svakoj komponenti u vektoru.

Predstavljanje dokumenata u modelu vektorskog prostora se često u literaturi naziva predstavljanje vrećom riječi (engl. bag of words representation). Osnovna teorija modela vektorskog prostora (engl. vector space model) je postojanje višedimenzionog prostora koji sadrži jedan vektor za svaki dokument i udaljenost između bilo koja dva vektora ukazuje na sličnost odgovarajućih dokumenata. Upit se također može predstaviti kao vektor i otkriti najrelevantnije dokumente za upit pronalazeći vektore dokumenata s najmanjom udaljenosti.

Kosinusova mjera⁴¹⁴ je mjera sličnosti između dva vektora koji se nalaze unutar n dimenzija pronalazeći kosinus kuta između njih. Unutar modela vektorskog prostora sličnost dva ili više dokumenta i upita računa se pomoću kosinusne udaljenosti dvaju vektora:

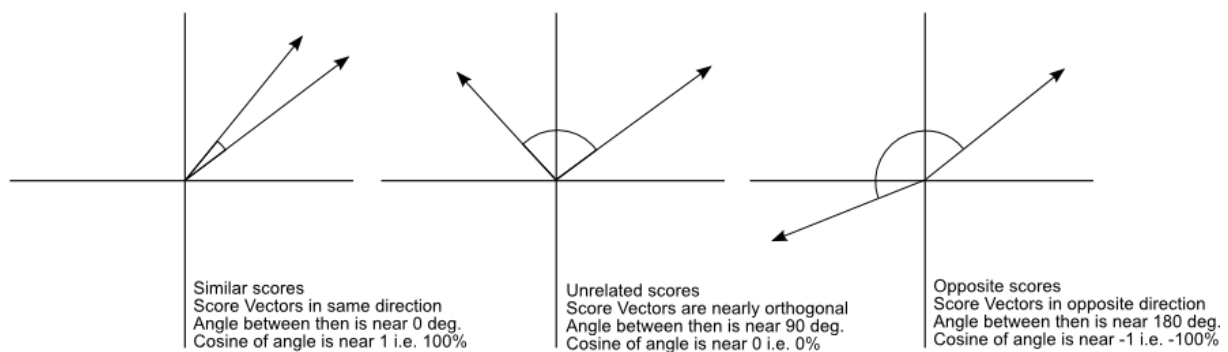
$$\text{sličnost} = \cos(\alpha) \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

gdje je brojnik skalarni umnožak dva vektora, a nazivnik umnožak njihovih duljina. Nazivnik u formuli normalizira vektore upita i dokumenta tj. skalarni produkt $A \cdot B$ dijeli s Euklidovom mjerom udaljenosti između A i B , tj. $A \cdot B / (\|A\| \|B\|)$ kompenzirajući time razlike zbog duljina tekstova unutar upita, odnosno dokumenta.

Ovaj omjer određuje kosinus kuta između dva vektora s vrijednostima između 0 i 1 , gdje 0 označava u potpunosti različit dokument, dok 1 označava u potpunosti jednak dokument, a sve između neposrednu sličnost ili različitost. Kako se kut između vektora smanjuje, kosinus se približava vrijednosti 1 , što znači da se dva vektora približavaju. Ovo je vrlo prikladan način rangiranja dokumenata mjereći blizinu njihovih vektora. U području obrade teksta vektori A i B su obično *tf-idf* vektori dokumenta. Kosinusna sličnost je korištena kao postupak normalizacije duljine dokumenata pri njihovoj usporedbi. Dakle, model vektorskog prostora indeksnim pojmovima u dokumentima i upitima pridružuje težine koje poprimaju realne vrijednosti i računa se stupanj podudarnosti između dokumenta i upita koji može poprimiti vrijednosti u intervalu $[0, 1]$. Na osnovi izračunatog stupnja podudarnosti dokumenti iz zbirke mogu se rangirati prema relevantnosti za korisnički upit. Kosinusna sličnost između dva vektora (ili dva dokumenta u vektorskom prostoru) je mjera koja izračunava kosinus kuta između njih. Ova mjera je mjera orijentacije, a ne veličine. Može se promatrati kao usporedba između dokumenata na normaliziranom prostoru jer se u obzir uzima kut između dokumenata. Međutim, prilikom računanja sličnosti neminovno je izvršiti reduciranje dimenzionalnosti kako bi se izračunavanje učinilo provodljivim. Kosinusna sličnost generira mjeru koja kazuje kako se odnose dva dokumenta gledajući kut umjesto

⁴¹⁴ Tata, Sandeep; Patel, Jignesh M. Estimating the selectivity of *tf-idf* based cosine similarity predicates // ACM SIGMOD Record 36, 2007, 75-80.

veliĉine, kao na Slici 73.



Slika 73: Odnos dvaju vektora tj. dokumenata (preuzeto iz: (Baeza-Yates, R.; Ribeiro-Neto, B., 2011.).

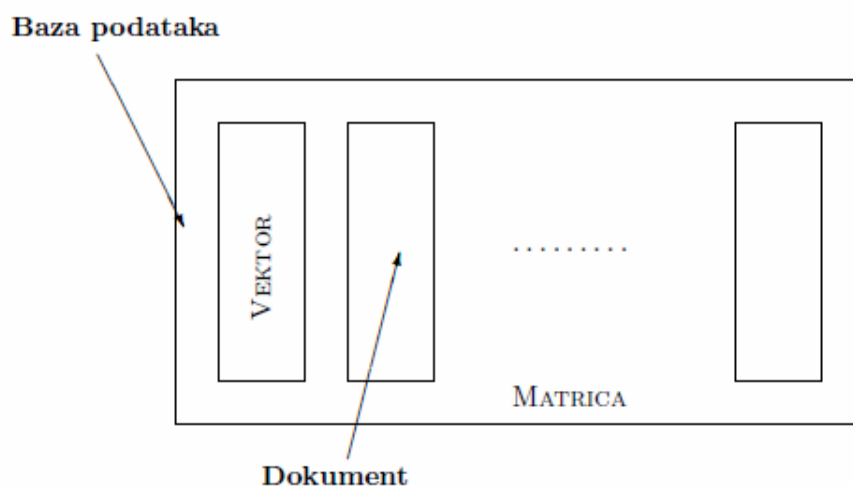
Dakle, korisno je uoĉiti ŝto su bliŝe dva vektora jedan drugom, izmeĊu njih ĉe biti manji kut te prema tome veći kosinus meĊu njima (tj. kut izmeĊu njih je blizu 0 stupnjeva, kosinus kuta je blizu 1). Dva identična vektora bi imala kut od 0 stupnjeva i mjeru sliĉnosti od 1.0, dok bi dva vektora koja su ortogonalna jedan na drugi imala kut od 90 stupnjeva i mjeru sliĉnosti od 0.0. Suprotni vektori bi imali kut izmeĊu njih blizu 180 stupnjeva, a kosinus kuta je blizu -1. Kosinus od 0 ° je 1, a to je manje od 1 za bilo koji drugi kut. To je, dakle, procjena orijentacije, a ne veliĉine: dva vektora s istom orijentacijom imaju kosinus sliĉnost 1, dva vektora na 90 ° imaju sliĉnost 0, a dva dijametralno suprotna vektora imaju sliĉnosti -1 neovisno o njihovoj veliĉini. Kosinus sliĉnost se posebno koristi u pozitivnom prostoru, gdje je ishod u blizini granica [0,1]. Pretpostavimo da postoji dokument s rijeĉi „novela“ koja se pojavljuje 200 puta i drugi dokument s s rijeĉi „novela“ koja se pojavljuje 50 puta, Euklidska distanca izmeĊu njih ĉe biti viŝa, ali ĉe kut joŝ uvijek biti mali jer oni pokazuju u istom smjeru, ŝto je zapravo ono bitno kada se usporeĊuju dokumenti.

9.3.1 Model vektorskog prostora (eng. Vector Space Model)

Tehnike koje nalikuju modelu vektorskog prostora N. Belkin i W. Croft⁴¹⁵ definiraju kao tehnike pretraŝivanja informacija koje se oslanjaju na formalni matematiĉki model za pretraŝivanje informacija, oblikujuĉi dokumente kao skupove termina koji se mogu

⁴¹⁵ Belkin, Nicholas J.; Croft, W. Bruce. Retrieval techniques. // Annual Review of Information Science and Technology (ARIST), vol. 22, chapter 4, pages 109–145. / M. Williams, editor . New York : Elsevier Science Publishers, 1987.

pojedinačno izvagati i obrađivati, izvodeći upite i uspoređujući predstavljanje upita s predstavljanjem svakog dokumenta u prostoru, čime se mogu pretraživati i dokumenti koji nužno ne sadrže neki od traženih termina. Modeli vektorskog prostora se oslanjaju na pretpostavku proizlaženja pojmovnog značenja dokumenta iz termina (riječi) sadržanih u dokumentu. Stoga, pristup vektorskog prostora opisuje svaki dokument kao skup termina te se model vektorskog prostora primjenjuje za kodiranje termina i dokumenata u tekstualnoj zbirci. U Odjeljku 9.2.2. (prikaz termina u dokumentu) bazu podataka smo predstavili matricom, koja je u modelu vektorskog prostora skup dokumenata opisanih određenim pojmovima. Vektori stupci $a_{11}, a_{12}, \dots, a_{1d}$ matrice A (Odjeljka br. 9.2.2.) predstavljaju dokumente baze podataka (vidi Sliku 74: Model vektorskog prostora). Broj stupaca matrice jednak je broju dokumenata i predstavljaju „dokument vektore ili d vektore“. Broj redaka matrice jednak je broju termina i predstavljaju „termin vektore ili p vektore“.



Slika 74: Model vektorskog prostora (preuzeto iz(Ivančević, Zvonimir; Jelić, Slobodan, loc cit. (bilj. 361.))).

Termini predstavljaju dokumente kao vektore termina $d = (p_1, p_2, \dots, p_n)$ gdje je p_i ($1 \leq i \leq n$) ne-negativna vrijednost koja označava pojedinačna ili višestruka pojavljivanja termina u dokumentu d . Dakle, svaki jedinstveni termin u zbirci dokumenata odgovara dimenziji u prostoru.⁴¹⁶ Dakle svaki vektor je identifikator jednog dokumenta, a svaka dimenzija odgovara zasebnom terminu. Stoga, svaka riječ u rječniku postaje nezavisna dimenzija visokodimenzionalnog vektorskog prostora. Dakle, ukoliko određenom dokumentu

⁴¹⁶ Salton, Gerard. Automatic text processing : the transformation, analysis, and retrieval of information by computer. Reading, Mass. : Addison-Wesley, 1989. Str. 127.

pripada neka ključna riječ, tada se u njegovom vektoru na poziciji te ključne riječi nalazi ne-nulta vrijednost. Ključne riječi po kojima se dokumenti indeksiraju čine bazu vektorskog prostora. Svaki element vektora predstavlja određenu važnost ili prisutnost pojma u opisu semantike tog dokumenta. Model vektorskog prostora postupa sa svakim terminom samostalno⁴¹⁷ jer svaki termin može biti pojedinačno izvagan (Odjeljak 9.2.3.), čime se dozvoljava da taj termin postaje više ili manje važan u dokumentu ili u cijeloj zbirci dokumenata kao cjelini.

Prilikom svakog korisničkog upita, sustav formira vektor upita i uspoređuje ga sa vektorima obilježja (predstavljenih ključnim riječima ili frazama) raspoloživih dokumenata tako što se računa njihova uzajamna sličnost. Korisniku se nudi onaj dokument koji ima najveću sličnost sa upitom primjenom mjera sličnosti (Odjeljak 9.3). Dakle, upit je predstavljen kao vektor $q = (p_1, p_2, \dots, p_m)$, gdje je termin p_i ($1 \leq i \leq m$) ne-negativna vrijednost koja označava broj pojavljivanja p_i (ili samo 1 kako bi označio pojavljivanje termina p_i u upitu)⁴¹⁸. Vektor dokumenta i vektor upita osigurava položaje objekata u termin-dokument prostoru. Izračunavanjem Euklidove udaljenosti između upita i drugih objekata u prostoru (termina ili dokumenata) bit će pretraženi objekti sličnog semantičkog sadržaja upitu. U nekim slučajevima, smjerovi vektora dobiveni izračunavanjem kosinusne mjere sličnosti su pouzdaniji pokazatelji semantičke sličnosti objekata od udaljenosti između objekata u termin-dokument prostoru⁴¹⁹. Modeli vektorskog prostora smještanjem termina, dokumenata i upita u termin-dokument prostor i izračunavanjem mjere sličnosti između njih omogućuju rangiranje rezultata upita prema upotrijebljenoj mjeri. Za razliku od tehnika leksičkog uparivanja koje ne osiguravaju rangiranje ili osiguravaju veoma sirove sheme rangiranja (primjerice, rangiranje jednog dokumenta prije drugog jer sadrži više pojavljivanja traženog termina), modeli vektorskog prostora temeljeći svoje rangiranje na Euklidskoj distanci ili na mjerenju kuta između upita i termina ili dokumenata u prostoru u mogućnosti su automatski voditi korisnika do dokumenata koji mogu biti više pojmovno slični i od veće koristi nego drugi dokumenti. Također, predstavljajući termine i dokumente u istom prostoru, modeli vektorskog prostora često osiguravaju metodu implementacije relevantne povratne informacije⁴²⁰. Relevantna povratna informacija dozvoljavajući dokumentima i terminima da

⁴¹⁷ Information Retrieval: Data Structures & Algorithms, loc cit. (bilj. 410).

⁴¹⁸ Belkin, Nicholas J.; Croft, W. Bruce., op. cit. (bilj. 415), str. 111.

⁴¹⁹ Information Retrieval: Data Structures & Algorithms, loc cit. (bilj. 410).

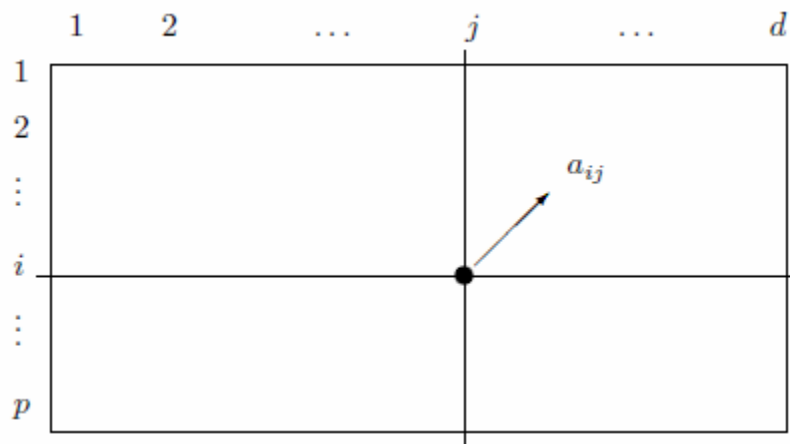
⁴²⁰ Salton, Gerard; Buckley, Chris. Improving retrieval performance by relevance feedback. // *Journal of the American Society for Information Science*, 41, 4(1990):288–297.

oblikuju upit te upotrebljavajući termine u ovim dokumentima za dopunu upita, povećavajući duljinu i preciznost upita pomaže korisniku točnije određivanje pretraživanja. Kao što je već navedeno, model vektorske reprezentacije sve riječi i dokumente preslikava u višedimenzionalni prostor. Povećanjem broja riječi i raznolikošću sadržaja raste i dimenzija tog prostora. Kod velikih skupova dokumenata dimenzija prostora može biti prevelika za računalnu obradu. Međutim, sve češće se javlja i potreba za analizom podataka velike dimenzionalnosti. Postaje teško razumjeti skrivenu strukturu takvih podataka. Nadalje, pohrana, prijenos i obrada podataka velike dimenzionalnosti postavljaju velike zahtjeve sustavu. Stoga je poželjno smanjiti dimenzionalnost podataka te istovremeno zadržati što više informacija i originalnu strukturu, što se postiže transformacijom i selekcijom obilježja čime se originalni podaci velike dimenzionalnosti preslikaju u novi prostor manje dimenzionalnosti, gdje je svaka značajka linearna ili nelinearna kombinacija značajki originalnog prostora. Cilj je u novom prostoru zadržati udaljenosti među podatkovnim vektorima. Smatra se da prostor manje dimenzionalnosti predstavlja skrivenu latentno semantičku strukturu originalnog skupa podataka. Najčešće upotrebljavane metode su *latentno semantičko indeksiranje*, *analiza nezavisnih komponenti*, *nasumična projekcija* i *kanonska korelacijska analiza*. Utvrđeno je da su ove metode manje učinkovite kada u skupu podataka postoji više nebitnih nego bitnih dimenzija⁴²¹.

Model indeksiranja skrivene semantike (LSI, Slika 75) predstavlja varijantu modela vektorskog prostora u kojemu je matrica baze podataka aproksimirana matricom nižeg ranga. Dakle, latentno semantičko indeksiranje koristi za preslikavanje originalnog prostora velike dimenzionalnosti u prostor manje dimenzionalnosti upotrebom dekompozicije singularnih vrijednosti (SVD), koja je matematička pozadina modela korištenog za redukciju dimenzionalnosti. Dimenzije novog prostora su ortogonalne matrice, što znači da su nezavisne, tj. termin-dokument matrica se rastavlja u tri matrice kao terminski vektor (U), singularna matrica (S) i dokumentski vektor (V) upotrebom latentnog semantičkog indeksiranja (LSI) temeljenog na SVD-u. Pokazalo se da se korištenjem SVD-a zadržavaju najvažnije veze između riječi i dokumenata te učinkovito odstranjuje šum, redundancija i višeznačnost unutar skupa podataka. Na taj način se određuje skupa pojmova iz termin-dokument matrice, gdje se pojam definira kao skup srodnih termina. Vektorski prostor

⁴²¹ Tang, Bin...et al. A Comparative Study of Dimension Reduction Techniques for Document Clustering: technical report. Dostupno na URL: https://www.cs.dal.ca/sites/default/files/technical_reports/CS-2004-14.pdf (17.02.2016).

stupaca matrice baze podataka zamijenjen je jednim svojim potprostorom. Smanjenje ranga interpretiramo kao uklanjanje nepotrebnih informacija iz baze podataka.



Slika 75: Model indeksiranja skrivene semantike – LSI model (preuzeto iz(Ivančević, Zvonimir; Jelić, Slobodan, loc cit. (bilj. 361.))).

9.3.1.1 Latentno semantičko indeksiranje

9.3.1.1.1 Upotreba

Područje informacijskih znanosti koje se bavi pretraživanjem informacija (engl. Information Retrieval – IR) ima osnovni zadatak razviti modele i metode za pretraživanje dokumenata u velikim bazama podataka i pronaći relevantne informacije u dokumentima na temelju zadanog upita. Kao što je navedeno u prethodnom tekstu, većina pristupa pretraživanja tekstualnih dokumenata ovise o leksičkoj podudarnosti riječi između postavljenog upita i onih koje su pridružene dokumentima. Leksičke metode su nužno nepotpune i neprecizne zbog velike raznolikosti u izboru riječi (sinonimi) koje ljudi koriste kako bi opisali isti dokument. Za razliku od leksičkih metoda, latentna semantička analiza (LSA) predstavlja metodu obrade prirodnog jezika (engl. natural language processing) koja analiziranjem odnosa između skupa dokumenata i sadržanih termina donosi skup pojmova povezanih s dokumentima i terminima. Dakle, LSI model se oslanja na sastavne termine dokumenta za predlaganje semantičkog sadržaja dokumenta te promatra termine u dokumentu kao nepouzdanе pokazatelje pojmova sadržanih u dokumentu jer je pretpostavka postojanost osnovne latentne semantičke strukture u podacima koja je djelomično prekrivena

varijabilnošću izbora riječi. Predloženi pristup pokušava prevladati nedostatke terminski podudaranog pretraživanja informacija obradom nepouzdanosti promatranih termin-dokument povezanih podataka kao statističkog problema upotrebom statističkih tehnika za procjenu ove latentne strukture te za otklanjanje šuma u podacima.⁴²² LSI statistički analizira obrasce upotrijebljenih riječi preko cijele zbirke dokumenata, smještajući dokumente sa slično upotrijebljenim obrascima riječi jednih blizu drugih u termin-dokument prostor i dozvoljavajući da semantički srodni dokumenti budu blizu jedan drugome, iako možda ne dijele iste termine⁴²³. Na taj način LSI pristup koristiti implicitne strukture visokog reda u povezivanju riječi s dokumentima tj. vezu 'riječ-dokument' ("semantička struktura") kako bi se poboljšalo otkrivanje relevantnih dokumenata na temelju termina pronađenih u upitu. Dakle, latentno semantičko indeksiranje je statistička metoda pretraživanja informacija koja može pretraživati tekst na temelju pojmova koje sadrži, a ne samo na temelju podudaranja određenih ključnih riječi. Upotrebom dekompozicije na singularne vrijednosti (SVD) iskorištava se prednost implicitne strukture visokog reda u povezivanju termina s dokumentima jer se dokumenti i termini predstavljeni kroz najveće singularne vektore podudaraju s korisničkim upitima. Dakle, ova metoda se naziva *latentno semantičko indeksiranje* (eng. *Latent Semantic Indexing - LSI*) zbog svoje sposobnosti korelacije semantički srodnih termina koji su "latentni" u zbirci teksta i jer podprostor predstavlja važan asocijativni odnos između termina i dokumenata koji nije očit u pojedinim dokumentima. Opis termina i dokumenata temeljen na latentnoj semantičkoj strukturi koristi se za indeksiranje i pretraživanje informacija te je LSI prvotno i osmišljen kao učinkovit alat za automatsko indeksiranje i pretraživanje podataka. Pronalaženjem semantičke strukture visokog reda (veza 'riječ-dokument') rješava se problem višeznačnosti govornog jezika. Stoga je LSI obećavajući način poboljšanja korisničkog pristupa mnogim vrstama tekstualnih materijala, odnosno dokumentima i uslugama za koje su dostupni tekstualni opisi.

Svrha upotrebe LSA metode u izgradnji ontologije je da se njezinom primjenom na odabranim dokumentima (npr. enciklopedijskim člancima) omogućuje određivanje pojma kao skupa srodnih termina te pripadnost pojedinih dokumenata tom pojmu. Ovaj pristup omogućuje automatsko klasificiranje članaka pojedinim ontologijskim klasama na temelju srodnih termina. Dakle, određivanje pripadnosti pojedinih termina pojedinom pojmu

⁴²² Pod "semantičke strukture" podrazumjeva se samo povezanost strukture na način na koji se pojedine riječi pojavljuju u dokumentima; "semantika" podrazumjeva samo činjenicu da termini u dokumentu mogu biti uzeti kao referenti na dokument ili na svoju temu.

⁴²³ Dumais, Susan T. Improving the retrieval of information from external sources. Dostupno na URL: <http://machinelearningtext.pbworks.com/w/file/attach/47378287/improvingRetrievalBRMIC91.pdf> (15.02.2015.).

omogućena je automatska klasifikacija članaka pojedinoj ontologijskoj klasi na temelju obrazaca riječi koje sadrže odabrani članci.

LSI postupak koristi termin-dokument matricu koja opisuje pojavljivanje termina u dokumentima. Primjenjuje dodjeljivanje lokalnih i globalnih težina termina⁴²⁴ za povećanje, odnosno smanjenje, važnosti termina u dokumentu ili između dokumenata preko cijele zbirke.⁴²⁵ Produkt lokalne i globalne težinske funkcije se primjenjuje na svaki ne-nulti element matrice A . Dakle, dokument se predstavlja vektorom, a za određivanje vrijednosti njegovih koordinata, koristi se tzv. TF-IDF mjera (engl. *Term Frequency – Inverse Document Frequency*), koja je detaljno opisana u prethodnom 9.2.7. Odjeljku. Težina pridružena i -toj ključnoj riječi u dokumentu, računa se kao produkt frekvencije pojavljivanja te ključne riječi u posmatranom dokumentu i tzv. inverzne frekvencije dokumenta. Uočljivo je da je frekvencija pojavljivanja ključne riječi lokalni parametar jer zavisi samo od sadržaja promatranog dokumenta, dok je inverzna frekvencija dokumenta globalni parametar jer zavisi od sadržaja svih dokumenata u prostoru. Latentno semantički indeksirani (LSI) model pretraživanja informacija predstavlja termine i dokumente u visoko dimenzionalnom prostoru omogućujući iskorištavanje temeljnih ("latentnih") semantičkih odnosa između termina i dokumenata tijekom pretraživanja, a dekompozicijom singularnih vrijednosti (SVD)⁴²⁶ reducira dimenzionalnost termin-dokument prostora pokušavajući riješiti probleme sinonimije i polisemije koji muče sustave automatskog pretraživanja informacija. Reduciranjem dimenzionalnosti termin-dokument prostora otkrivaju se temeljni semantički odnosi između dokumenata, a velik dio "buke" (razlika u korištenju riječi, termini koji ne pridonose razlikovanju dokumenata, itd.) je eliminiran. LSI pojmovi su bogato opisani u Deerwester, S., Dumais, ST, Furnas, GW, Landauer, T. K., Harshman, R. et al.⁴²⁷ LSI je detaljno opisan u Berry et al.⁴²⁸. Latentno semantičko indeksiranje (LSI) je predstavljeno 1990. godine radom

⁴²⁴ Ibid

⁴²⁵ Težine termina se obično sastoje od dvije komponente, komponente dokumenta i komponente zbirke: $a_{i,j} = L_{i,j} G_{i,j}$, gdje je $L_{i,j}$ lokalna težinska funkcija za termin i u dokumentu j i $G_{i,j}$ je globalna težinska funkcija za termin i .

⁴²⁶ Golub, Gene H.; Van Loan, Charles F. *Matrix Computations*. 3rd ed. Baltimore, London : Johns Hopkins University 1996. Dostupno na URL:

<http://web.mit.edu/ehliu/Public/sclark/Golub%20G.H.,%20Van%20Loan%20C.F.-%20Matrix%20Computations.pdf> (15.02.2016.).

⁴²⁷ Deerwester, Scott...et al. Indexing by latent semantic analysis. // *Journal of the American Society for Information Science* 41, 6 (1990), 391–407. Dostupno na URL:

http://www.cob.unt.edu/itds/faculty/evangelopoulos/dsci5910/LSA_Deerwester1990.pdf (15.02.2015.).

⁴²⁸ Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W., loc cit. (bilj. 403).

S. Deerwester, S. Dumas, G. Furnas, T. Landauer, R. Harsman⁴²⁹, a radom M. W. Berry, S.T. Dumas i G.W. O'Brien⁴³⁰ predstavljena je metoda dodavanja novih dokumenata i pojmova 1995. godine. Koristi se u različitim pretraživanjima informacija i aplikacijama za obradu teksta, iako je njegova primarna primjena bila pojmovno pretraživanje teksta i automatska kategorizacija dokumenata. Dodatni načini upotrebe LSI-ja su: sažimanje teksta, otkrivanje informacija, otkrivanje odnosa, automatsko označavanje slika ključnim riječima, vizualizacija informacija i sl. Upotreba LSI pristupa ne zahtijeva da tekst bude u obliku rečenice, LSI se može nositi i s popisima imena, slobodno oblikovanim bilješkama i sl. Sve dok riječ-dokument matrica može biti generirana, LSI može raditi s podacima u bilo kojem formatu ili jeziku. LSI ne upotrebljava pomoćne strukture, kao što su kontrolirani rječnici, što ga čini odgovarajućim za složenu analizu nestrukturiranih zbirki dokumenata. Automatski se prilagođava novoj i mijenjajućoj terminologiji, a to pokazuje da je izuzetno tolerantan na šum (odnosno, pogrešno napisane riječi, tiskarske pogreške, nečitljive znakovi, itd), što je osobito važno za aplikacije koje koriste tekst izveden iz optičkog prepoznavanja znakova (OCR). Budući da su postupci koji se primjenjuju u LSI pristupu pojašnjeni u prethodnim odjeljcima (8.2.2 i 8.2.3), u nastavku slijedi objašnjene SVD-a.

9.3.1.1.2 SVD

Dekompozicija na singularne vrijednosti je matematička pozadina modela korištenog za redukciju dimenzionalnosti. Preslikavanje originalnog prostora velike dimenzionalnosti u prostor manje dimenzionalnosti s minimalnom pogreškom udaljenosti vrši se dekompozicijom na singularne vrijednosti (SVD). Dimenzije novog prostora su ortogonalne matrice, što znači da su nezavisne. Korištenjem SVD-a zadržavaju se najvažnije veze između riječi i dokumenata te se učinkovito odstranjuje šum, redundancija i višeznačnost unutar skupa podataka.

Za svaku matricu A tipa $m \times n$ postoji dekompozicija singularnih vrijednosti (engl. singular value decomposition, SVD) koja otkriva strukturu matrice te omogućuje aproksimaciju nižeg ranga. Dekompozicija matrice na singularne vrijednosti (SVD) je dobro poznata metoda rastavljanja matrice koja dopušta jednostavnu strategiju optimalne

⁴²⁹ Deerwester, Scott...et al., loc cit. (bilj. 426).

⁴³⁰ Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W., loc cit. (bilj. 403).

aproksimacije manjim matricama te je prikazana je na Slici 76. Ona rastavlja matricu A , odnosno $m \times n$ termin-dokument matricu ($A = m \times n$) s m terminima i n dokumentima, kao

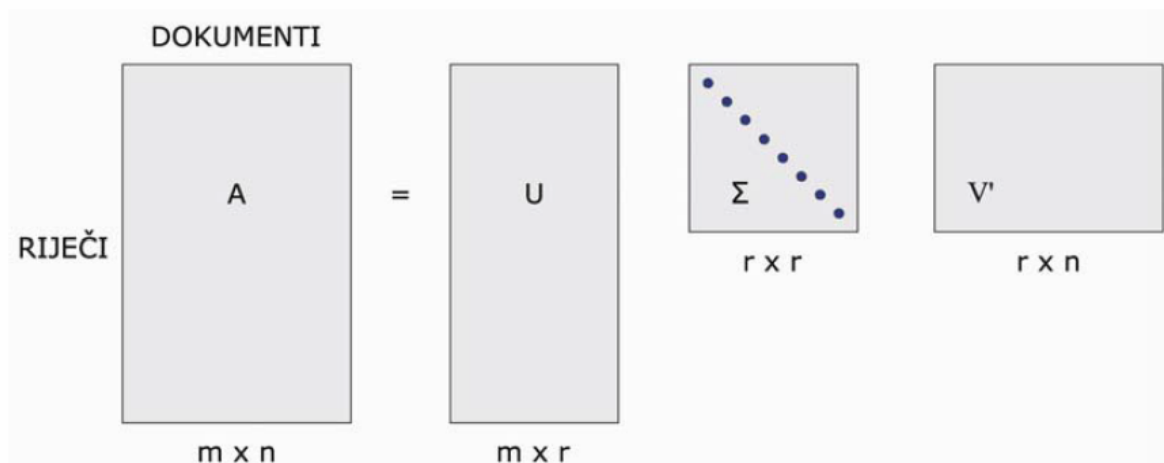
$$A = U \Sigma V^T$$

gdje je:

$U = m \times r$ ortogonalna matrica čiji su stupci lijevi singularni vektori matrice A , nazvana termin matrica

$V = r \times n$ ortogonalna matrica čiji su stupci desni singularni vektori matrice A , nazvana dokument matrica

$\Sigma = r \times r$ je dijagonalna matrica na čijoj dijagonali su singularne vrijednosti matrice A u padajućem redoslijedu.



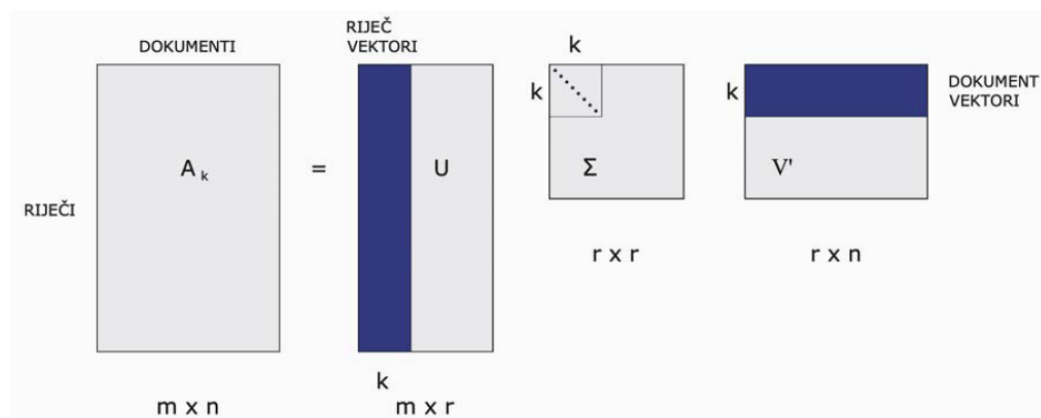
Slika 76: Dekompozicija matrice na singularne vrijednosti (preuzeto iz: (Deerwester...et al., 1990.).

Kako su singularne vrijednosti matrice Σ sortirane u padajućem poretku možemo zadržati prvih k najvećih, a ostale postaviti na nula. Za metodu LSI koristi se *krnja dekompozicija singularnih vrijednosti* (engl. *truncated SVD*). Nova matrica A_k je matrica

ranga k te je dobra aproksimacija matrice A . Kao što je raspravljeno u Nicholas i Dahlberg⁴³¹, te Derwester et al.⁴³², A_k je aproksimacija od A , s rastućom točnošću kako se k približava r .

Matricu Σ možemo pojednostavniti brisanjem redaka i stupaca koje smo postavili na nula. Na taj način dobijemo novu matricu Σ_k . Obrišemo li odgovarajuće stupce matrica U i V dobit ćemo nove matrice U_k i V_k . Kao rezultat dobivamo slijedeći model, koji je najbolja moguća aproksimacija matrice A ranga k te je prikazan Slikom 77.

$$A \approx A_k = U_k \Sigma_k V_k^T$$



Slika 77: Aproksimacija matrice A matricom A_k (preuzeto iz: (Deerwester...et al., 1990.).

$U_k = m \times k$ matrica, čije stupce čini prvih k lijevih singularnih vektora matrice A ;

$\Sigma_k = k \times k$ matrica, na čijoj je dijagonali vodećih k singularnih vrijednosti matrice A ;

$V_k = n \times k$ matrica, gdje je $k \leq r$, čije stupce čini prvih k desnih singularnih vektora matrice A ;

Redci matrice U_k = indeksni pojmovi;

Redci matrice V_k = dokumenti.

⁴³¹ Nicholas, Charles K; Dahlberg Randall. Spotting Topics with the Singular Value Decomposition // PODDP '98 Proceedings of the 4th International Workshop on Principles of Digital Document. London: Springer-Verlag, 1998. Str. 82-91.

⁴³² Deerwester, Scott...et al., loc cit. (bilj. 427).

Važno svojstvo SVD-a je da izračunata matrica A_k točno ne rekonstruira originalnu „termin-dokument matricu“ A , ali SVD zadržava većinu najvažnije skrivene strukture u povezanosti termina i dokumenata te istovremeno odbacuje šum u podacima koji uzrokuje lošije pretraživanje informacija. A_k je najbliža k -dimenzionalna aproksimacija originalnog termin-dokument prostora predstavljenog s maticom A . Posljedica LSI obrade je uspostavljanje veza između riječi koji se pojavljuju u sličnim kontekstima. Riječi koje se pojavljuju u sličnim dokumentima biti će blizu u k -dimenzionalnom prostoru čak i ako se nikad zajedno ne pojavljuju u istom dokumentu ukoliko je to konzistentno s najjačim asocijativnim uzorcima u podacima tj. ukoliko je u skladu s glavnim obrascem povezivanja podataka.

Dakle, usko povezane riječi i dokumenti postavljeni su jedni blizu drugih u semantičkom prostoru te će raspored prostora odražavati glavne asocijativne obrasce u podacima i zanemariti manje važne utjecaje. Rezultat toga je da se termini koji se zapravo ne pojavljuju u dokumentu mogu ipak smjestiti u blizini dokumenta, ako su u skladu s glavnim obrascem o povezivanju podataka. Pozicija u prostoru zatim služi kao nova vrste semantičkog indeksiranja i pretraživanje informacija se odvija upotrebom termina u upitu za identifikaciju točaka u prostoru te se dokumenti u njihovoj blizini vraćaju korisniku. Kao rezultat, upiti prema skupu dokumenata na kojima se izvršio LSI vratiti će rezultate koji su u značenju pojmovno slični upitu, čak ako i ne dijele određenu riječ ili riječi s upitom.

U LSI, lijevi i desni singularni vektori određuju lokacije termina i dokumenata u reduciranom termin-dokument prostoru. Singularne vrijednosti se često koriste za mjerenje terminskih i dokumentskih vektora, omogućujući grupiranje termina i dokumenata kako bi se lakše identificirali. U reduciranom prostoru semantički povezani termini i dokumenti će se vjerojatno nalaziti u blizini jer SVD namjerava izvući temeljnu semantičku strukturu termin-dokument prostora⁴³³.

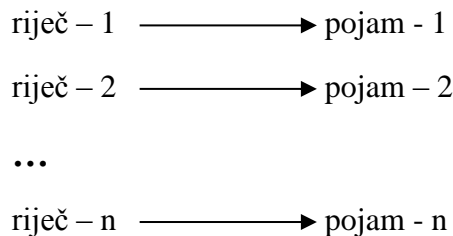
9.3.1.1.3 LSI u izgradnji ontologije hrvatske književnosti na korpusu Hrvatske enciklopedije

9.3.1.1.3.1 *Uvod*

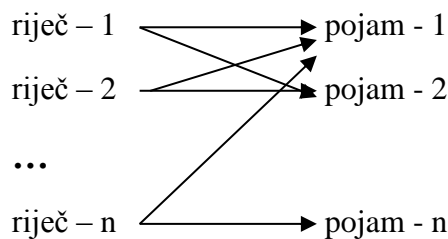
Latentna semantička analiza (LSA) doslovno označava analiziranje dokumenata kako bi se pronašlo temeljno značenje ili pojmovi tih dokumenata. Ukoliko jedna riječ uistinu znači

⁴³³ Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W., loc cit. (bilj. 403)

jedan pojam i jedan pojam je opisan samo s jednom riječju, tada će LSA biti jednostavna budući da u tom slučaju postoji samo jednostavno pridruživanje (mapiranje) s jedne riječi na jedan pojam.



Nažalost ovaj problem nije jednostavan jer različite riječi mogu imati isto značenje (sinonimi), ili biti riječi s višestrukim značenjem (polisemi) čineći nejasnim pojmove.



Podsjetimo, latentna semantička analiza je izrasla iz problema pronalaska relevantnih dokumenata riječima za pretraživanje jer temeljna poteškoća nastaje kada se uspoređuju riječi za pronalazak relevantnih dokumenata, a ne ono što se stvarno želi, usporedba *značenja ili pojmova* iza tih riječi. LSA pokušava riješiti ovaj problem pridruživanjem (mapiranjem) i riječi i dokumenata u "pojmovni" prostor i usporedbom u ovom prostoru. Kako autori imaju širok izbor riječi koje su im dostupne prilikom pisanja, pojmovi mogu biti nejasni zbog različitog izbora riječi kod različitih autora. Ovaj slučajni izbor riječi uvodi „buku“ u riječ-pojam odnos. Latentna semantička analiza filtrira nešto od ove buke i pokušava pronaći najmanji skup pojmova koji se proteže u svim dokumentima. Kako bi ovaj težak problem bio rješiv, LSA uvodi neka pojednostavljenja: (1) dokumenti se prikazuju kao "vreća riječi", gdje je redoslijed riječi u dokumentu nevažan, bitno je samo koliko se puta svaka riječ pojavljuje u dokumentu i (2) pojmovi su predstavljeni kao uzorci riječi koje se obično pojavljuju zajedno u dokumentima. Iz prethodno navedenog proizlazi da će se riječi i dokumenti sa sličnim značenjem i tematikom grupirati u istom klasteru. Na temelju formilarnih klastera se mogu izvući određeni pojmovi. Za navedeno slijedi primjer.

9.3.1.1.3.2 Primjena LSI na enciklopedijskim člancima

Odabrano je šesnaest članka Hrvatske enciklopedije, čiji su naslovi tj. natuknice članaka označeni kao D1, D2,...,D16. Ekstrahirane su natuknice (prezime i ime književnika tj. naslovi članaka koji predstavljaju dokumente) i termini iz područja književnosti koji se odnose na vrstu književnog djela koju su pojedini autori pisali. Kao ključne tj. indeksne riječi su odabrani termini koji se pojavljuju u dva ili više članaka. Navedeno je prikazano u Tablici 22.

Tabela 22: Ekstrahirani termini

NATUKNICA	TERMINI
Smoljan, Ivo = D 1	['publicist', 'kritika', 'feljton']
Zima, Zdravko = D2	['publicist', 'kritika', 'kriticar', 'esejist', 'esej']
Bartulica, Milostislav = D3	['publicist', 'kritika', 'feljton', 'esej', 'putopis']
Pavletić, Vlatko = D4	['kriticar', 'esejist', 'feljton', 'esej']
Dečak, Mirko = D5	['publicist', 'kritika', 'putopis']
Pavličić, Pavao = D6	['roman', 'kriminalisticni', 'proza', 'price', 'feljton']
Tribuson, Goran = D7	['roman', 'kriminalisticni', 'proza', 'price', 'kritika']
Klarić, Kazimir = D8	['roman', 'humoristicni']
Tomović, Robert = D9	['roman', 'humoristicni', 'kritika', 'esej']
Kumičić, Eugen (Evgenij) = D10	['roman', 'povijesni']
Mayer, Milutin = D11	['roman', 'religiozno', 'povijesni']
Ivanac, Ivica = D12	['roman', 'dramaticar', 'drama']
Novosel, Krešo = D13	['roman', 'dramaticar', 'drama', 'esej']
Božičević (Božičević), Božić (Nadal) = D14	['pjesnik', 'hagiografski', 'epilij']
Botić, Luka = D15	['epsko', 'lirska', 'pjesma']
Čerina, Vladimir = D16	['lirska', 'pjesma', 'feljton', 'esej']
KLJUČNE RIJEČI:	
['drama', 'dramaticar', 'esej', 'esejist', 'feljton', 'humoristicni', 'kriminalisticni', 'kriticar', 'kritika', 'lirska', 'pjesma', 'povijesni', 'price', 'proza', 'publicist', 'putopis', 'roman']	

Kao što je objašnjeno u prethodnom tekstu (Odjeljak 9.3.1.1.1.), prvi korak u latentnoj semantičkoj analizi je stvaranje termin-dokument matrice od odabranih članaka (dokumenata). U ovoj matrici, svaka indeksna riječ je predstavljena u recima, a svaka natuknica tj. naslov pojedinog članka (koja predstavlja pojedini članak, odnosno dokument) je prezentirana u stupcima. Svaka ćelija matrice sadrži broj puta pojavljivanja riječi u

određenom naslovu (dokumentu). Na primjer, riječ "drama" se pojavljuje jedan put u naslovu D12 i D13. Općenito, matrice izgrađene tijekom LSA znaju biti vrlo velike s većinom ćelija koje sadrže 0 jer svaki naslov ili dokument uobičajeno sadrži samo mali broj svih mogućih riječi. Dobivena je sljedeća matrica za navedeni primjer:

```
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0.]
 [0. 1. 1. 1. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 1.]
 [0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 1. 1. 0. 1. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 1. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 1. 1. 1. 1. 1. 1. 1. 0. 0. 0.]]
```

U latentnoj semantičkoj analizi prethodno dobiveni podaci matrice se obično oblikuju tako da su rijetke riječi izvagane kao teže od uobičajenih općih riječi. Na primjer, riječi koje se pojavljuju u samo 5% dokumenata trebaju vjerojatno biti izvagane kao teže nego riječi koje se pojavljuju u 90% dokumenata. Primijenjena mjera za vaganje težine riječi je TF-IDF mjera (termin frekvencija – inverzna frekvencija dokumenta), prema kojoj se broj svake čelije zamjenjuje izračunom prema formuli objašnjenom u Odjeljku 9.2.7. Nakon izračuna težine pojedinih termina izvodi se dekompozicija matrice na singularne vrijednosti. SVD pronalazi reducirani dimenzionalni prikaz matrice koji naglašava najjače odnose i odbacuje buku. Trik u korištenju SVD-a je u pronalaženje broja dimenzija ili "pojmovna" koje će se upotrijebiti kada se aproksimira matrica. SVD faktorizira matricu na tri matrice. *Termin matrica U* daje koordinate svake riječi u „pojmovnom“ prostoru, *dokument matrica Vt* daje koordinate svakog dokumenta u „pojmovnom“ prostoru i *S matrica* singularnih vrijednosti daje naslutiti koliko je dimenzija ili "pojmovna" potrebno uključiti. Kod velikih zbirka dokumenata broj upotrijebljenih dimenzija je u rasponu od 100 do 500. U ovom malom primjeru, kojega se želi grafički prikazati, koristiti će se tri dimenzije s tim da se prva dimenzija izbaciti te se grafički prikaže druga i treća dimenzija. Razlog zbog kojeg se odbacuje prva dimenzija je jer ona korelira s duljinom dokumenta, a za riječi korelira s brojem puta koliko se riječ pojavljuje u svim člancima (dokumentima), te nije tako informativan kao druga i treća dimenzija. Svaka riječ ima tri broja povezana s njom, po jedan za svaku dimenziju. Isto tako, svaki naslov (natuknica) koja predstavlja članak ima 3 broja povezana s njim, po jedan za svaku dimenziju. Prva dimenzija nije jako zanimljiva jer teži odgovarati broju riječi u naslovu.

Ovo su singularne vrijednosti i uzete su prve tri: 4.155, 3.232 i 2.528

```
[ 4.15583932e+00  3.23202649e+00  2.52879431e+00  2.35056669e+00
 1.95214899e+00  1.77621714e+00  1.57086599e+00  1.41421356e+00
 1.01966877e+00  8.36690537e-01  4.35673567e-01  3.91474581e-16
 2.41416890e-16  2.01188605e-16  9.77754350e-17  8.86672057e-33]
```

Prva tri stupca termin U matrice

```
[[ 0.09620724 -0.17673132 -0.29499616]
```

```
[ 0.09620724 -0.17673132 -0.29499616]
```


[0.45507596 0.19895469 -0.46288097]

[0.1543899 0.16863255 -0.14455899]

[0.35591191 0.16935724 0.0234811]

[0.11449729 -0.11285907 -0.08269959]

[0.14633343 -0.2210539 0.31133142]

[0.1543899 0.16863255 -0.14455899]

[0.47172184 0.18600212 0.34300878]

[0.06110978 0.0571381 -0.18348077]

[0.06110978 0.0571381 -0.18348077]

[0.05380724 -0.15843769 -0.05542368]

[0.14633343 -0.2210539 0.31133142]

[0.14633343 -0.2210539 0.31133142]

[0.31112271 0.33373584 0.2130821]

[0.15563335 0.16668111 0.15308587]

[0.41084521 -0.66908198 -0.12178802]]

Prva tri 3 retka dokument Vt matrice

[[0.27401359 0.37217519 0.42096569 0.26944441 0.22582151 0.29013572

0.31800251 0.12641069 0.34942167 0.11180713 0.11180713 0.14515953

0.25466232 -0. 0.02940911 0.2245533]

[0.2132084 0.32671692 0.32633737 0.21830794 0.2123804 -0.35980102

-0.35465104 -0.24193522 -0.12282827 -0.2560374 -0.2560374 -0.31637879

-0.25482153 -0. 0.03535745 0.14931441]

[0.22918906 -0.07747094 0.10668202 -0.28808901 0.28044067 0.33046869

0.45682443 -0.08086368 -0.12826658 -0.07007755 -0.07007755 -0.28147024
-0.46451438 0. -0.14511324 -0.31887189]]

9.3.1.1.3.3 Interpretacija i vizualizacija rezultata

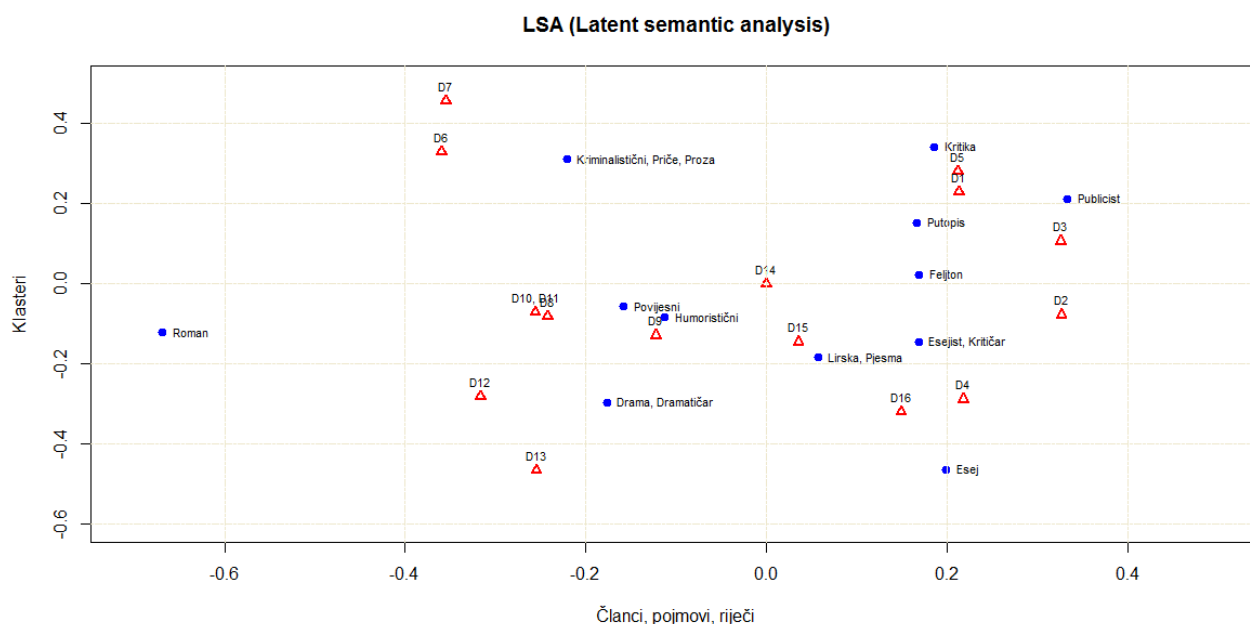
Nakon što se izvela latentna semantička analiza na primjeru, dobiveni su sljedeći rezultati koji se mogu vizualizirati u tabelarnom prikazu. Tablica 23 prikazuje prisutnost određenih riječi grupiranih po pripadnosti određenom članku (istaknuto znakom +) te pripadnost riječi i članaka određenom klasteru (znak + obojan prema pripadnosti dokumenta i riječi određenom klasteru). Svakom pojedinom klasteru je dodijeljena određena boja zbog boljeg prikaza, što ukazuje na prisutnost četiri klastera, odnosno srodni termini iz svakog klastera se mogu oblikovati pod jedan pojam koji će se imenovati.

Tabela 23: Prisutnost određenih riječi grupiranih po pripadnosti određenom članku (istaknuto znakom +) te pripadnost riječi i članaka određenom klasteru

	Smoljan,	Zima,	Bartulica,	Pavlečić,	Dečak,	Pavličić,	Tribuson,	Klarić,	Tomović,	Kumičić,	Mayer,	Ivanac,	Novosel,	Božićević	Botić,	Čerina,
	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
pjesnik														+		
epilij														+		
hagiografski														+		
epik																
epsko															+	
lirska															+	+
pjesma															+	+
roman						+	+	+	+	+	+	+	+			
humoristični								+	+							
kriminalistični						+	+									
povijesni										+	+					
religiozno											+					
proza						+	+									
priče						+	+									
dramaticar												+	+			
drama												+	+			
publicist	+	+	+		+											
kritika	+	+	+		+		+		+							
kriticar		+		+												
esejist		+		+												
feljton	+		+	+		+										+
esej		+	+	+					+				+			+
putopise			+		+											

Grafički se mogu prikazati indeksne riječi i natuknice članaka (dokumenti) na XY grafu i identificirati pojedine klasterne. Natuknice (dokumenti) su označene crvenim trokutima,

a indeksne riječi plavim krugovima. Na primjer, uočljivo je da članci D12 i D13 pišu o dramatičarima jer se isti nalaze u blizini indeksnih riječi koje karakteriziraju dramski klaster (drama, dramatičar). Članci D1-D7, D9, D13 i D16 su smješteni s indeksnim riječima (publicist, kritika, kritičar, esejist, feljton, esej i putopis) okupljajući književnike stvaraoce diskurzivnih djela te na taj način formiraju diskurzivni klaster. U XY graf se unosi druga dimenzija na X os i treća dimenzija na Y os svake riječi i dokumenta. Primjerice, riječ „drama“ ima dimenzijske vrijednosti (0.0962, -0.1767, -0.2949). Prva dimenzijska vrijednost 0.0962 se zanemaruje, a grafički prikaz riječi „drama“ se pozicionira ($x = -0.1767$, $y = -0.2949$). Na isti način se grafički prikazuju i natuknice članaka tj. dokumenti. Dimenzijske vrijednosti grafičkog prikaza svake indeksne riječi i dokumenta su prikazane u Tablici 24, kao i na Slici 78.



Slika 78: Grafički prikaz indeksnih riječi i dokumenata

Tabela 24: Vrijednosti pojedinih indeksnih riječi i dokumenata

Drama	-0.17673132	-0.29499616
dramatičar	-0.17673132	-0.29499616

Esej	0.19895469	-0.46288097
Esejist	0.16863255	-0.14455899
Feljton	0.16935724	0.0234811
humoristični	-0.11285907	-0.08269959
kriminalistični	-0.2210539	0.31133142
kritičar	0.16863255	-0.14455899
kritika	0.18600212	0.34300878
Lirska	0.0571381	-0.18348077
Pjesma	0.0571381	-0.18348077
Povijesni	-0.15843769	-0.05542368
Priče	-0.2210539	0.31133142
Proza	-0.2210539	0.31133142
Publicist	0.33373584	0.2130821
putopis	0.16668111	0.15308587
roman	-0.66908198	-0.12178802
D1	0.2132084	0.22918906

D2	0.32671692	-0.07747094
D3	0.32633737	0.10668202
D4	0.21830794	-0.28808901
D5	0.2123804	0.28044067
D6	-0.35980102	0.33046869
D7	-0.35465104	0.45682443
D8	-0.24193522	-0.08086368
D9	-0.12282827	-0.12826658
D10	-0.2560374	-0.07007755
D11	-0.2560374	-0.07007755
D12	-0.31637879	-0.28147024
D13	-0.25482153	-0.46451438
D14	-0.	0.
D15	0.03535745	-0.14511324
D16	0.14931441	-0.31887189

Prednost ove tehnike je da su indeksne riječi i natuknice članaka (prezime i ime književnika tj. dokumenti) prikazani istim grafom, što omogućuje grupiranje skupine indeksnih riječi i književnika (dokumenata) u klaster, a time i imenovanje klastera tj. pojma koji okuplja srodne termine i dokumente. Primjerice, može se uočiti klaster koji sadrži indeksne riječi kritika, publicist, putopis, feljton, esejist, kritičar, esej te članke o književnicima D1, D2, D3, D4, D5. Očito je da navedeni klaster obuhvaća diskurzivne književnike tj. književnike koji su pisali diskurzivna djela. Stoga se navedeni klaster može imenovati npr. diskurzivni književnici. Ukoliko su autori pored diskurzivnih djela pisali i npr. pjesme tada će autor tj. članak D16 ulaziti i u klaster pjesnika tj. smjestit će se bliže riječima lirska i pjesma koje definiraju klaster pjesnika. Klasteru pjesnika pripadaju, osim D16, i članci D14 i D15. Uočljivo je da se dokumenti D12 i D13 nalaze između indeksnih riječi roman, drama, dramatičar jer po vrsti književnog djela književnik predstavljen s D12 pripada prozaicima i dramatičarima. Osim toga, D13 je u blizini i riječi esej jer su i eseji prisutni u njegovom stvaralaštvu. Klasteru prozaika pripadaju D8, D10, D11. Dokumenti D6, D7 i D9 se nalaze između pojmova diskurzivni i prozaci jer su pisali djela ovih književnih vrsta. Na osnovu prethodno navedenog, razvrstavanje autora tj. natuknica članaka (dokumenata) kao instanci pojmova (klasa) i grupiranje riječi koje opisuju pojedine pojmove može se prikazati Tablicom 25.

Tabela 25: Razvrstavanje autora tj. natuknica članaka (dokumenata) kao instanci pojmova (klasa) i grupiranje riječi koje opisuju pojedine pojmove

Članci	Pojmovi	Riječi
D1, D2, D3, D4, D5, D6, D7, D9, D13, D16	Diskurzivni književnici	kritika, publicist, putopis, feljton, esejist, kritičar, esej
D6, D7, D8, D9, D10, D11, D12, D13	Prozaici	kriminalistički, priča, proza, roman, povijesni, humoristični
D12, D13	Dramatičari	drama, dramatičar
D14, D15, D16	Pjesnici	lirska, pjesma

9.4 Zaključak

Navedeno poglavlje i primjer prikazuju svrhu upotrebe LSA metode u izgradnji ontologije. Nakon provedene LSA metode na odabranim dokumentima uočava se njezina korisnost u određivanju pojma kao skupa srodnih termina te pripadnost pojedinih natuknica članaka (dokumenata) tom pojmu, čime se omogućuje automatska klasifikacija članaka pojedinim ontologijskim klasama. Ovim pristupom se vrši automatsko klasificiranje članaka na temelju srodnih termina. Dakle, određivanje pripadnosti pojedinih termina pojedinom pojmu omogućena je automatska klasifikacija članaka pojedinoj ontologijskoj klasi na temelju obrazaca riječi koje sadrže odabrani članci, što je prikazano na odabranom primjeru.

10 ZAKLJUČAK

Rezultati dobiveni u teorijskom i istraživačkom dijelu ovog rada ponovit će se u zaključnom razmatranju i staviti u kontekst s procesima koji su utjecali na promjene u objavljivanju referentnih djela pa tako i samih enciklopedija.

Cilj ove doktorske disertacije je istražiti koji elementi enciklopedičke organizacije znanja mogu pružiti podršku za razvoj ontologije te razviti metodu kojom će se generirati ontologija na osnovi enciklopedički organiziranog znanja. Na temelju korpusa odabranih 1170 biografskih članaka Hrvatske enciklopedije iz područja hrvatske književnosti i utvrđenih strukturnih elemenata biografskih članaka razvila se ontologija za opis i predstavljanje znanja obrađenog u odabranim člancima. Razvijena ontologija poslužiti će za organizaciju, pretraživanje i pregledavanje znanja pohranjenog u mrežnoj online Hrvatskoj enciklopediji, kao i za dobivanje preciznih odgovora na postavljena pitanja.

Proučavanjem literature, analizom sličnih ontologijskih modela odabranih svjetskih projekata i znanja pohranjenog u biografskim člancima Hrvatske enciklopedije iz područja hrvatske književnosti predložio se ontologijski model koji na djelotvoran način opisuje enciklopedičko znanje navedenog područja.

Moderna tehnologija utječe na enciklopedijska djela u procesu njihove izrade, rasparčavanja, korištenja, struktriranosti podataka i pohrane; te istovremeno autorima, stvarateljima i nakladnicima pruža nepojmljive mogućnosti oblikovanja i prikaza sadržaja. Globalna mreža, internet, se ubrzano transformira u semantičku mrežu prelazeći od povezivanja dokumenata na povezivanje podataka, odnosno dosadašnji WEB portali s klasičnim bazama informacija i znanja, postaju povezani podaci (engl. linked data) globalnog oblaka (engl. cloud computing).

Usporedbom organizacijske strukture tradicionalnih enciklopedija na papirnatom mediju s elektroničkim enciklopedijama u mrežnom okruženju uočavaju se određene razlike proizašle iz različitih vrsta medija koje omogućavaju nove funkcionalnosti pretraživanja.

U svrhu očuvanja korisnosti enciklopedije u današnjoj informacijskoj eksploziji potrebno je poboljšati sposobnost predstavljanja njihovog sadržaja na smisleni (semantički) način u mrežnom okruženju. Dinamična priroda i povećanje razmjera digitalnih referentnih radova čine tradicionalne uredničke metode okupljanja i organiziranja izvora metasadržaja (indeksi, unakrsne uputnice, tablice sadržaja) toliko intenzivnim i neučinkovitim da su praktički neprimjenjivi na elektroničke enciklopedije u mrežnom okruženju. Korisniku koji

proučava izvor može se ponuditi *vođeno razgledavanje* kroz sadržaj ostvarivanjem pojmovnog i semantičkog navigacijskog sučelja, što je demonstrirano u ovom radu. Kako bi se riješili navedeni problemi na odgovarajući način, potrebne su sofisticiranije tehnike generiranja metasadržaja velikih i neistovremeno ažuriranih tekstualnih korpusa, što zahtijeva tehnološke eksperte za dizajniranje formalnog prikaza i eksperta domene za dohvaćanje semantičkog sadržaja korpusa. Organizacija mrežnih enciklopedija u počecima svojega razvoja počiva na hipertekstu, čijom se pretečom na određeni način mogu smatrati upravo tradicionalna enciklopedička i leksikografska djela jer se sadržaj u njima organizira podjelom na dijelove koji su međusobno povezani uputnicama što sasvim nalikuje hipertekstu. U prvim hipertekstovnim enciklopedijama sredinom 1990-ih do željenog članka se pristupalo odabirom sve užega područja kojem članak pripada, odnosno pregledavanjem informacija (engl. *browsing*). Ubrzo se elektroničkim enciklopedijama dodaje mogućnost pretraživanja čitavoga teksta, popisa članaka ili digitaliziranoga indeksa prema upitu korisnika uspostavljanjem sustava za pretraživanje podataka (engl. *information retrieval*), odnosno tražilica. Dakle, usporedbom tradicionalnih enciklopedičkih djela izdanih na papirnatom mediju s elektroničkim enciklopedijama u mrežnom okruženju uočava se slijedeće:

hipertekstovnost je omogućila zamjenu uputnica elektroničkim poveznicama i olakšala kretanje među sadržajem, *pretraživost* je zamijenila abecedni slijed mogućnošću pretrage naslova članaka ili čitavoga teksta, *neograničenost opsega* je omogućila širenje broja i duljine članaka, *poveznice* su omogućile povezivanje pojedinih članaka na vanjske izvore (npr. digitalizirane sadržaje arhiva, knjižnica, muzeja), *kolaborativnost* je omogućila izravno sudjelovanje korisnika u stvaranju, dopunjavanju i ispravljanju enciklopedičkog djela. Enciklopedijskim djelima tradicionalnoga tipa bili su potrebni abecedariji kao svojevrsni kontrolirani rječnici utvrđeni prije početka posla na samim člancima, dok se kod enciklopedijskih djela mrežnog tipa ističe potreba za klasifikacijskim sustavom složenije razine (taksonomijom, tezaurusom i ontologijom).

Glavna trenutna prepreka bolje podrške korisnicima je ta što značenje sadržaja nije računalno razumljivo. Semantička interoperabilnost podrazumijeva postojanje infrastrukture koja će omogućiti strojnu interpretaciju i zaključivanje o sadržajima na webu. Stoga, ključni pojam semantičkog web-a predstavlja ontologija, temeljna komponenta u omogućavanju semantičke interoperabilnosti. Pojam ontologije vezuju se za mogućnost ponovne upotrebe i dijeljenja znanja u nekoj domeni. U svrhu predstavljanja određene domena ontologijom potrebno je obuhvatiti pojmove koji je opisuju i razviti njihovu taksonomiju, attribute pojmova i odnose među pojmovima.

Navedene promjene kroz koje prolazi enciklopedičko djelo zahtijeva uspostavljanje načina modeliranja organizacije enciklopedičkog znanja kako bi se osigurala njihova dostupnost i korištenje iz čega proizlazi i potreba uspostavljanja novog načina modeliranja enciklopedičke organizacije koje će svoje utemeljenje pronaći na analizi specifičnosti strukture enciklopedičkog članka uvažavanjem temeljnih postavki semantičkog weba, principa ontologijskog oblikovanja i potreba korisnika.

Rad je na temelju strukturalne organizacije biografskih enciklopedičkih članaka primijenjenih u enciklopedičkoj organizaciji znanja dao smjernice za poboljšanje i daljnji razvoj organizacije enciklopedičkog znanja u mrežnom okruženju imajući u vidu njegove specifičnosti i daljnji razvoj prema semantičkom webu.

U teoretskom dijelu rada donosi se pregled aktualne literature o semantičkom webu i ontologijama kako bi se mogle primijeniti temeljne postavke u razvoju ontologijskog modela Hrvatske enciklopedije za područje hrvatske književnosti.

Semantički web nije globalni informacijski put paralelan s postojećim World Wide Webom. Upravo suprotno, semantički web je postupni razvoj postojećeg weba. Problem trenutnog WWW je informacijska prezasićenost. Semantički web nudi postojanje metapodataka za strukturiranje. Trenutni web se promatra kao skup dokumenata u kojemu postoji nešto metapodataka pridruženih dokumentima (title, keywords, itd.) te poveznica među dokumentima, a semantički web dodaje semantičke metapodatke na razini dokumenta (pridružene cijelom dokumentu) i na razini sadržaja pridružujući ih određenom dijelu i mjestu u dokumentu. Tim Berners-Lee je zamisao semantičkog weba predstavio modelom „*Slojeviti kolač*“ (engl. „Layer Cake“), koji je podijeljen na slojeve ukazujući na arhitekturu semantičkog weba. Model ima zadatak standardizirati tehnologiju semantičkog web-a. Na najosnovnijoj razini semantički web ovisi od Unicode (standard za kodiranje niza znakova, čijom upotrebom svi ljudski jezici mogu biti napisani i čitani na webu) i Uniform Resource Identifier (URI, standardni oblik za identifikaciju izvora), što je jednostavno rečeno web identifikator. Dakle, taj sloj ima zadaću jedinstveno definirati pojmove u globalnoj mreži podataka. Drugi sloj arhitekture semantičkog weba je XML (engl. Extensible Markup Language) jezik za označivanje. XML je proširiv te tagovi mogu biti definirani i iskoristivi kako to zahtijeva određena aplikacija. U semantičkoj web arhitekturi, XML sloj s XML namespace i XML shemom objavljuje da se koristi zajednička sintaksa u semantičkom webu. Dakle, on predstavlja temeljnu strukturu zapisa podataka. Druga važna tehnologija za razvoj semantičkog weba je RDF (engl. Resource Description Framework), okvir za opisivanje informacija i resursa na webu. Stavljajući informacije u RDF datoteke računalnim

programima za pretraživanje se čini izvedivim otkrivanje, prihvaćanje, prikupljanje, analiziranje i obrada informacija s interneta jer su RDF dokumenti pisani u XML-u, a oni mogu lako biti razmijenjeni među različitim vrstama operativnih sustava i aplikacijskih jezika (engl. application language). RDF se temelji na izjavama koje su poznate kao trojke (tripleti, engl. triple). Svaka trojka sadrži subjekt, predikat i objekt; poput rečenice u prirodnim jezicima. Stoga se svaka trojka može oblikovati kao graf s dva čvorišta (subjekt i objekt), koji su povezani predikatom. RDFS Shema pruža RDF rječnik kako bi se omogućio opis taksonomije klasa i svojstava, dajući glavne elemente opisa ontologija. Dakle, *RDF* je model koji samo povezuje podatke, a *RDF Shema* daje osnovni smisao i značenje vezama te omogućuje hijerarhiju. Sljedeći temeljni sloj arhitekture semantičkog weba je OWL web ontologijski jezik (engl. Web Ontology Language), koji se može usporediti s gramatičkim sustavom semantičkog weba. Korijeni OWL-a su u deskriptivnoj logici i njegov temeljni cilj je donijeti snagu zaključivanja deskriptivne logike semantičkom webu. OWL je izgrađen na vrhu RDF-a te pruža dodatni standardizirani rječnik, vrhunsku strojnu interoperabilnost i dobro definiranu sintaksu koja je puno jača od RDF-a. Spomenute specifikacije čine OWL odgovarajućim jezikom za razvoj ontologije, odnosno prikazivanje pojmova i njihovih odnosa. Uloga logičkog sloja (engl. Logic layer) je podržati snažni logični jezik za donošenje zaključaka, koji čine semantički web dovoljno izražajnim. Dakle, sloj logike je zadužen za razumljivo izvlačenje informacija i donošenje odluka od strane aplikacija. Sva semantika i pravila se izvode na slojevima ispod sloja dokaza (engl. Proof) i rezultati će se koristiti za potvrdu zaključivanja. Gornji slojevi arhitekture semantičkog weba izražavaju zajedno i dokaze i pouzdanost podataka kako bi prikazali pouzdanost rezultata. Kriptografija je sredstvo digitalnog potpisa koji se može koristiti za vjerodostojnost. Korisničko sučelje aplikacije se može izgraditi na vrhu svih ovih slojeva. Dakle, *sloj dokaza* traži dokaz kako je pronađeni podatak točan, a *sloj povjerenja* je zadužen za provjeru mogućnosti vjerovanja resursu da su podaci točni. Ovo je najviši stupanj te će web doseći njegov potpuni potencijal kada korisnici budu imali povjerenje u njegovo djelovanje i u dostavljene informacije. Ontologija modelira znanje na semantičkoj razini, stoga je ontologija značajna za prikaz određene domene jer omogućuje automatsko zaključivanje i interpretaciju s primijenjivim semantičkim kontekstom. Ontologija je sposobna integrirati i razmjenjivati podatke između heterogenih izvora informacija te definirati sučelja za nezavisne i na znanju utemeljene usluge, kao što je odgovaranje na upite i ponovno korištenje izvora znanja. Drugim riječima, ontologija se može koristiti kao način komunikacije između čovjeka i sustava ili sustava sa sustavom, čime je omogućeno semantičko pretraživanje informacija (IR) i upravljanje znanjem.

Potrebno je spomenuti i SKOS (engl. Simple Knowledge Organization System), koji ima cilj pružiti jednostavan, ali moćan okvir za izražavanje sustava za organizaciju znanja na strojno razumljiv način. SKOS osigurava model za izražavanje osnovne strukture i sadržaja konceptualnih shema kao što su tezaursi, klasifikacijske sheme, popisi predmetnih odrednica, taksonomije, folksonomije i ostale vrste kontroliranih rječnika te također konceptualne sheme ugrađene u glosarije i terminologiju. SKOS temeljni rječnik (engl. SKOS Core Vocabulary) je skup RDF svojstava i RDFS klasa, koji se može koristiti za izražavanje sadržaja i strukture konceptualne sheme kao RDF grafa. Osmišljen je kako bi se omogućilo stvaranje usklađene organizacije znanja koja se može ponovo koristiti i na koje se može upućivat preko web-a. Upotrijebljen je u razvoju ontologije Hrvatske enciklopedije za područje hrvatske književnosti te se njime razriješuju pitanja deskriptora, sinonima, opisa termina i sl.

Poznavanje iznesenih teoretskih znanja daje dobru osnovu za dobro razumijevanje primjene ontologija u informacijskim sustavima. Prednost semantičkog weba i ontologijskog pristupa u organizaciji enciklopedičkog znanja je već prepoznata i primijenjena u velikim enciklopedijskim projektima. Uvid u primjere dobre prakse daju proučeni primjeri temeljnih ontologija na semantičkom webu i ontologija temeljenih isključivo na enciklopedijskom znanju. Prepoznajući potrebu za velikim ontologijama i otvorenim pristupom stvaranja slobodnog i javnog standarda pristupilo se suradničkoj izradi standardne ontologije gornje razine (engl. Standard Upper Ontology - SUO), koja će osigurati definicije za termine opće namjene i biti će temelj za specifičnije domenske ontologije. Radna grupa je započela rad SUO ontologije na temelju nekoliko ontologija gornje razine (npr. SUMO, CYC, ontologija gornje razine opisana od strane Russell i Norvig-a i John Sowa ontologije gornje razine) pokušavajući prilagoditi većinu njihovih pojmova i odnosa u zajedničku ontologiju. Rezultat rada Radne grupe je ontologija ograničena na metakoncepte, opće, apstraktne i filozofske pojmove koji su dovoljno općeniti za usmjeravanje na široki opseg domenskih područja.

SUMO (engl. Suggested Upper Merged Ontology) je opsežan ontologijski projekt otvorenog pristupa. Izvorno se bavi konceptima meta-razine tj. općenitim entitetima koji ne pripadaju određenoj problemskoj domeni. Proširen je na ontologiju srednje razine (engl. Mid-Level Ontology, MILO) i dvanaest domenskih ontologija uz potpunu podešenost skupova poveznica na WordNet. Nastao je sjedinjenjem javno dostupnih ontoloških sadržaja u sveobuhvatnu i jedinstvenu strukturu.

UMBEL je stvoren kako bi pridonio sadržajnoj interoperabilnosti na Web-u. Predstavlja ontologiju gornje razine s oko 28.000 referentnih pojmova izvučenih iz Cyc baze znanja, koja će osigurati opća pridruživanja u svrhu povezivanja različitih ontologija ili shema. UMBEL referentna struktura je široko primjenjiva kao usmjeravajući čvorovi za bilo koju domenu znanja za pridruživanje relevantnih web dostupnih sadržaja sa specifičnim ciljem promicanja interoperabilnosti i razumijevanja jedinstvene referentne strukture i njegovih povezanih izvora. Ova referentna struktura može osigurati početne čvorove za pridruživanje dostupnih relevantnih web sadržaja te pružiti snažan skup konstelacija za vođeno otkrivanja povezanih skupova podataka i sadržaja. Predstavlja rječnik za pomaganje ontologijskih pridruživanja.

Primarni cilj Cyc projekta je izgraditi veliku bazu znanja koja će sadržati formalizirano znanje prikladno za različita zaključivanja i rješavanje problemskih zadataka u različitim domenama. Postoji i neznatno šira verzija za istraživače, nazvana ResearchCyc. Cyc ontologija ima svoj vlastiti, svrhovito građeni program zaključivanja, koji se pojavljuje u paketu zajedno s OpenCyc i ResearchCyc. Znanje u OpenCyc-u je mali podskup znanja u ResearchCyc bazi znanja, koja je verzija Cyc-a usmjerena na istraživačku zajednicu. Podskup Cyc ontologije je besplatno izdan pod nazivom OpenCyc te predstavlja opću ontologiju i bazu znanja otvorenog pristupa s namjerom uspostavljanja općeg rječnika za automatsko zaključivanje. Cyc baza znanja se tradicionalno dijeli na ontologije gornje, srednje i niže razine. Ove razlike su namijenjene pretraživanju razine općenitosti informacije sadržane u njima.

BabelNet (engl. Building a Very Large Multilingual Semantic Network) projekt je velika višejezična semantička mreža, koja automatski stvara resurs pomoću metodologije integriranja leksičkog i enciklopedijskog znanja iz WordNet-a i Wikipedije, kao i integriranjem leksičkih praznina jezično siromašnih resursa uz pomoć strojnog prevođenja u svrhu obogaćivanja resursa leksičkim informacijama. Rezultat je „enciklopedijski rječnik“, koji osigurava pojmove i imenuje entitete leksikalizirane na mnogim jezicima te je povezan s velikom količinom semantičkih odnosa.

CIDOC (engl. Conceptual Reference Model, CRM) je proširiva ontologija za koncepte i informacije u kulturnoj baštini i muzejskoj dokumentaciji. Predstavlja standard za kontroliranu razmjenu informacija o kulturnoj baštini. Svrha CIDOC CRM je osigurati referentni model i informacijski standard koji bi muzeji i ostale institucije kulturne baštine mogle koristiti za opis svojih zbirki u svrhu poboljšanja razmjene informacija.

WordNet projekt predstavlja računalni resurs na semantičkoj razini koji daje podatke o leksiku engleskog jezika. Svrha WordNet-a je stvoriti kombinaciju rječnika i tezaurusa koji je intuitivnije upotrebljiv te podržati automatsku analizu teksta. Značajka WordNet-a je njegov pokušaj organizacije leksičkih informacija u obliku značenja riječi, a ne oblicima riječi. Skupina srodnih riječi je međusobno povezana putem konceptualno semantičkih i leksičkih odnosa, što rezultira stvaranjem mreže povezanih riječi i pojmova. Semantički i leksički odnosi između leksema i sinonimskih grozdova se razlikuju za pojedine vrste riječi. WordNet kodira sinonimiju, hiperonimiju i hiponimiju (engl. hypernymy / hyponymy, tj. odnos između nadpojma i podpojma), holonimiju i meronimiju (engl. holonymy/meronymy, tj. odnos između dijela i cjeline), antonimiju (posebno važne za imenice), polisemiju (za pridjeve), troponimiju, implikaciju, kauzalnost (za glagole). Ovi odnosi, posebice hiponimija i hiperonimija, omogućuju stvaranje leksičkih hijerarhija.

Neophodno se dotaknuti Schema.org, važnog projekta za razvoj semantičkog weba. Trenutno postoji mnogo standarda i shema za označavanje različitih vrsta informacija na web stranicama. Kao rezultat toga, webmasterima se teško odlučiti koji standard koristiti kao najrelevantniji i najpodržavaniji za obilježavanje. Stvaranje sheme podržane od strane svih glavnih tražilica čini webmasterima lakšim označavanje, a tražilicama lakše stvaranje bogatih pretraživačkih obilježja za korisnike. Schema.org je rezultat suradnje između tražilica Bing, Google, Yahoo! i Yandex te se one na nju oslanjaju kako bi se poboljšao prikaz rezultata pretraživanja čineći ljudima lakše pronalaženje odgovarajuće web stranice te kako bi se webmasterima olakšalo osiguranje boljeg direktnog usmjeravanja korisnika na njihove stranice i podatke. Schema.org je zajednički napor poboljšanja weba stvaranjem strukturirane sheme za obilježavanje podataka (engl. structured data markup schema) podržane od strane glavnih tražilica. Stoga, Schema.org osigurava zbirku schema koje webmasteri mogu koristiti za označavanje HTML stranica na načine prepoznate od strane glavnih tražilica (engl. search providers), kao i za strukturiranu interoperabilnost podataka (npr. u JSON).

PROTON (PROTo ONtology) je ontologija gornje razine te služi kao modelirajuća osnova za brojne zadatke u različitim domenama. Sadrži oko 300 klasa i 100 svojstava, osiguravajući pokrivenost općih pojmova potrebnih za široki raspon zadataka, uključujući semantičko označavanje, indeksiranje i pretraživanje dokumenata. Principi dizajniranja ontologije su: (1) domenski neovisna, (2) lagane logičke definicije, (3) poravnata je i kompatibilna s popularnim standardima i ontologijama, (4) dobra pokrivenost imenovanih entiteta i konkretnih domena (npr. ljudi, organizacije, lokacije, brojevi, datumi, adrese).

Poznati svjetski ontologijski projekti nastali isključivo na enciklopedijskom znanju su DBpedia, Freebase, YAGO i InPhO.

DBpedia projekt vrši ekstrakcije strukturiranih informacija iz Wikipedije s ciljem da ih se učini dostupnim na webu. Rezultat je DBpedia baza znanja, koja opisuje veliki broj entiteta (osobe, mjesta, glazbene radove, filmove, organizacije) te sadrži RDF triplete ekstrahirane iz različitih višejezičnih izdanja Wikipedije. Pokriva mnoge domene, višejezična je i dostupna na Webu. DBpedia ontologija je višedomenska ontologija (geografske informacije, osobe, ustanove, filmovi, muzika, knjige...). Izrađena je ručno na temelju najčešće korištenih predložaka infookvira unutar Wikipedije. Postupak mapiranja provodi se zbog prevladavanja problema sinonimnih naziva svojstava i višestrukih predložaka te ekstrakcije informacija u formi RDF tripleta. Autorica ovog rada je postupkom mapiranja pridružila 36 predložaka infookvira hrvatskog izdanja Wikipedije na 32 klase DBpedia ontologije i 188 infookvirnih svojstva na 173 ontologijska svojstva DBpedije. Uočljivo je pridruživanje svojstava s istim značenjem iz različitih infookvira na jedno ontologijsko svojstvo (npr. infookvirna svojstva URL, web, webstranice koje navode web adresu unutar različitih predložaka infookvira pridružena su na jedno ontologijsko svojstvo foaf:homepage). Odabrani su predlošci iz domene književnosti, filma, glazbenika, knjiga, medija, kulture itd. Time je Hrvatska postala četvrta zemlja uključena u postupak mapiranja infookvira u DBpedia projektu (prethodila su joj mapiranja za englesko, njemačko i grčko izdanje Wikipedije). Rezultat načinjenog postupka mapiranja su ekstrahirani RDF tripleti koji opisuju trenutni broj prisutnih entiteta u hrvatskom izdanju Wikipedije: npr. 2000 glazbenika, 500 glazbenih instrumenata, 10 muzeja, 2000 filmskih umjetnika, 56 glazbenih žanrova, 19 umjetničkih kipova, 215 likovnih umjetnika, 267 papa, 163 skladatelja, 19 svećenika, 7 web stranica, 13 astronauta, 2585 muzičkih albuma, 171 biografija, 46 filozofa, 196 knjiga, 58 slika, 89 televizijskih epizoda, 118 znanstvenika, 1159 filma, 9 glazbenih festivala, 132 hrvatska grada, 117 jezika, 443 književnika, 232 monarha, 653 singla, 30 svetaca, 141 vojne osobe, 45 časopisa.

Freebase je suradnički izgrađena i uređena baza znanja općih strukturiranih podataka kojoj pridonose korisnici prikupljajući podatke iz različitih izvora, kao što su Wikipedia i MusicBranz. Namijenjena je širokoj publici te predstavlja sustav baze podataka stvoren kao otvoreni web repozitorij svjetskog znanja, a njezino stvaranje je motivirano širokom upotrebom informacijskih zajednica, kao što su Semantic Web i Wikipedia. Sadrži objekte gotovo svih vrsta (filmovi, knjige, slavne osobe, lokacije, kompanije itd) i činjenice. Vrste i svojstva može stvoriti bilo koji korisnik kako bi dodao strukturirane podatke na teme u Freebase-u.

YAGO je velika semantička baza znanja proizašla automatskom ekstrakcijom činjenica iz Wikipedije, WordNet-a i GeoNames-a. Nastao je kao potreba stvaranja veće ontologije koristeći trenutno postojeće ontologije s glavnim ciljem ujedinjenja Wikipedije i WordNet-a te iskoristivosti bogatih struktura i informacija osiguravajući vjerodostojnost činjenica. Yago ekstrahira činjenice iz kategorija i infookvira Wikipedije u kombinaciji s taksonomskim odnosima iz WordNet-a. YAGO je zamišljen kao detaljan i izražajan model podataka jer je sposoban izraziti entitete, činjenice, odnose između činjenica i svojstva odnosa. YAGO2 je novo izdanje YAGO baze znanja s naglaskom na vremensko i prostorno znanje. Razvijen je kao prošireni pristup za ekstrakciju činjenica integrirajući vremenske i prostorne semantičke informacije Wikipedije, WordNet-a, i GeoNames-a. Poboljšan prikaz podataka uvodi vrijeme i mjesto kako bi se bogatstvo prostorno-vremenske informacije u YAGO-u moglo istražiti grafički ili putem posebnog prostorno-vremenski svjesnog upitnog jezika. YAGO3 je proširenje YAGO baze znanja koja kombinira podatke iz višejezičnih Wikipedija s engleskim WordNet-om kako bi se izgradila jedna suvisla baza znanja. Iskoristile su se kategorije, infookviri, Wikidata, kao i značenje infookvirnih atributa na više jezika. Metoda je pokrenuta na deset različitih jezika, čime je YAGO povećao broj novih entiteta i činjenica. Ulazni podaci su popis Wikipedijinih članaka. Cilj je stvoriti bazu podataka od ovih višejezičnih Wikipedija, što uključuje tri izazova: (1) odrediti skup entiteta, (2) ekstrahirati činjenice o tim entitetima iz višejezičnih Wikipedija, (3) stvoriti taksonomiju. Ključna prednost je mogućnost ostavljanja postojećeg engleskog YAGO-a kao referentne baze znanja i glavnog taksonomijskog oslonca. Dok u YAGO-u svaki Wikipedijski članak postaje entitet, kod YAGO3 treba paziti kako se entiteti ne bi duplirali jer isti entitet može biti opisan u različitim jezičnim Wikipedijama.

InPhO (engl. Indiana Philosophy Ontology) je projekt pokrenut na Indiana Sveučilištu, upotrebljava kombinaciju automatskih metoda i povratne informacije stručnjaka za stvaranje dinamične računalne ontologije iz filozofske discipline. InPho projektom se razvija ontologija za dinamično referentno djelo online otvorenog pristupa Stanford Encyclopedia of Philosophy (SEP). SEP enciklopedija je prvi pokušaj akademske zajednice organiziranja vlastite profesionalne predmetne građe kolaborativnim pisanjem, objavljivanjem i upravljanjem dinamično ažuriranim referentnim djelom u cijelosti na webu. Sve što autori napišu pregledavaju volonteri pojedinih predmeta prije objave na webu, što SEP pristup čini različitim od Wikipedije. Razvoj ontologije za područje filozofije je središte uspjeha u ovom projektu. Umjesto traženja potpuno automatiziranog rješenja za SEP-ove potrebe

metapodataka, nastoji se iskoristiti pristup stručnjacima domene filozofije. Pristup dinamičkoj InPho ontologiji počinje s malom količinom početne ručno izgrađene ontologije, nakon čega se koriste razne automatizirane metode za strukturiranje oblika povratne informacije i njezino razvijanje u valjane podatke, popunjavanje ontologije te polu-automatsko proširivanje ontologijske taksonomijske strukture. InPho ontologija se sastoji od šest osnovnih kategorija: *Filozof*, *Dokument*, *Organizacija*, *Nacionalnost*, *Profesija* i *Misao*. Pojmovi gornje razine su pridruženi u standardne ontologije gdje je to moguće u svrhu povećanja interoperabilnosti podataka (*FOAF*, *AKT Reference Ontology* i sl.). Za početni nacrt ontologije izravno je prikupljena taksonomijska shema od urednika SEP-a te je tako početno građenje kategorija oblikovano traženjem poddisciplinske taksonomije od strane stručnjaka područja, kojima se daje više slobode za osiguranje najbolje konceptualizacije ciljane domene.

Prikaz velikih ontologijskih enciklopedijskih projekata je ukazao da su navedene ontologije razvijanje na temelju strukturiranih podataka enciklopedijskog članka (Wikipedijini infookviri u DBpedia projektu), doprinosa stvaranja strukturiranih podataka od strane samih korisnika (Freebase) iz različitih izvora (Wikipedia, MuzicBranz) te kombinacija automatskih metoda i povratne informacije autora i urednika domene (InPho). Proučeni projekti su ukazali na to kako ni jedan od prethodno navedenih projekata nije proučavao isključivo strukturu određene vrste enciklopedičkog članka kao osnovu razvoja ontologije, što je predmet ovog rada jer se utvrdila postojanost poveznice između strukture enciklopedičkog biografskog članka i gradivnih ontologijskih elemenata. Istražio se značaj pojedinih strukturnih elemenata enciklopedijskog članka za razvoj ontologije, a ne samo značaj strukturiranih podataka enciklopedijskog članka. Upravo se s tom neistraženom vezom nestrukturiranih podataka unutar strukture enciklopedijskog članka i ontologije bavi ovaj rad koji je potvrdio osnovnu hipotezu prema kojoj enciklopedička organizacija znanja može kvalitetno poslužiti razvoju ontologije određene domene definirajući dovoljan broj pojmova, njihovih atributa i međusobnih odnosa koji čine rječnik (ontologiju) za predstavljanje znanja odabranog područja.

Istraživanjem se razvila *ontologija domene*, odnosno vrsta ontologije koja opisuje područje književnosti u okviru nacionalnih književnosti tj. konkretno hrvatske književnosti. Ontologija je obuhvatila enciklopedijsko znanje vezano uz pojedinu nacionalnu književnost tj. u ovom slučaju hrvatsku književnost. Ne postoje ontologije koje pokrivaju predmetnu domenu, npr. DBpedia sadrži određene književne pojmove, ali to nije ontologija koja bi mogla opisati znanje književnosti određenog naroda. Dobiveni ontologijski pojmovi i odnosi

se mogu upotrijebiti za opis bilo koje nacionalne književnosti. Konkretno, dobivena ontologija se planira upotrijebiti za opis enciklopedičkog znanja hrvatske književnosti, njegovu bolju organizaciju te dobivanje preciznih odgovora na pojedina pitanja o hrvatskim književnicima i hrvatskoj književnosti, kao i za semantičko online pregledavanje odabranog područja. Kao polazna ontologija višeg reda odabrana je PROTON ontologija.

Primijenjen je Protégé računalni program otvorenog koda (engl. open source) nastao na Stanford University s jednostavnim grafičkim korisničkim sučeljem za jednostavan i brz način izrade, preoblikovanja, provjere i vizualizacije ontologije. Izradu i verifikaciju ontologije nadgleda i provjerava sam program, koji javlja pogreške u slučaju nedefiniranih elemenata ontologije. Rad je ukazao na mogućnost postizanja interoperabilnosti kako pojedinačnih ontologijskih resursa, tako i cjelokupne ontologije sa postojećim semantičkim ontologijskim projektima (DBpedia, Yago, Freebase itd.). Uspostavljanje interoperabilnosti ontologije s ostalim projektima na semantičkoj mreži omogućit će povezivanje relevantnog i bogatog znanja *Hrvatske enciklopedije* u „globalnu mrežu znanja“ koja nastaje i razvija se kroz projekte semantičke mreže.

Ontologija opisuje znanje o književnim djelima i književnicima specifično za određeni narod te tako dobivena ontologija omogućuje proučavanje nacionalnih književnosti jer sadrži sve potrebne odnose za opis književnosti bilo kojeg naroda. Obuhvaćen je veliki broj odnosa potreban za opis produkcije književnika pojedinih nacionalnih književnosti, njihovih životopisa, međusobnih odnosa i cjelokupnog znanja koje se nalazi u biografskim enciklopedijskim člancima iz područja književnosti čime se omogućuje opis književnosti bilo kojeg naroda. Kako ontologija osim književnih pojmova obuhvaća i opis znanja vezan uz konkretne osobe tj. u ovom slučaju književnike, sadrži pojmove i odnose koji nisu usko vezani uz književnost (npr. lokalitete, ustanove, zanimanja, životopis itd.). Taj dio ontologije je primjenjiv na sve enciklopedijske biografske članke, bez obzira na odabrano područje. Dakle, ontologija omogućuje uvid u detaljne informacije o književnim djelima i književnicima određenog naroda. Ontologija uključuje informacije o životopisu književnika, produkcijskoj povezanosti pojedinih književnika, odnose između pojedinih književnih djela, informacije s aspekta teorije i povijesti književnosti o pojedinim književnicima i djelima, kao i informacije o cjelokupnoj književnoj produkciji određenog naroda.

Biografski članci su u enciklopedičkom djelu posvećeni pojedinim osobama. Analiza takvih članaka u Hrvatskoj enciklopediji LZMK je pokazala njihovu ujednačenost u pogledu strukture i slijeda podataka o životu i djelovanju osobe, opsegu kojega zauzimaju imajući u vidu značaj osobe koja se u njima obrađuje. Članak počinje imenom (prezimenom i „danim“

imenom, imenom i nadimkom, imenom i brojem itd.) u poludebelom kurentu. U obliku zagradama slijedi mjesto i datum rođenja te mjesto i datum smrti. Od činjenica iz životopisa donosi se ono što je relevantno za djelovanje i opus; navodi se školska formacija i profesija; doktorat, članstvo (djelovanje) u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama; prestižne nagrade i priznanja. Djelatnost se ovisno o važnosti osobe iznosi u svim strukovnim razdjelima ili pak samo u onoj struci u kojoj je osoba ostavila vidan trag. Djela se navode u reprezentativnom izboru.

Osim opisa strukture i redoslijeda elemenata enciklopedijskog biografskog članka, analizirana je vrsta i oblik podataka koji se pojavljuju u biografskom enciklopedijskom članku Hrvatske enciklopedije za područje hrvatska književnost. Navedena analiza vrste podataka biografskog enciklopedijskog članka doprinijela je izgradnji ontologije jer je ukazala na vrste podataka koji su uključeni u njezinu izgradnju. Analizom dobivenog opisa vrste i oblika podataka dobiva se uvid u vrste informacija koje se nalaze u biografskim enciklopedijskim člancima hrvatske književnosti u Hrvatskoj enciklopediji. Ovako dobiveni uvid u vrstu informacija omogućuje postavljanje temeljnih aspekata informacija (*životopis, osobne informacije, školovanje, posao, djelo, autorstvo stvaralaštvo*), koji mogu poslužiti za postavljanje temeljnih faceta ontologijskih slojeva omogućujući facetno pretraživanje tj. sužavanje skupa rezultata koji se dobiju prvotnim pretraživanjem/pregledavanjem s obzirom na kriterije koji su omogućeni facetnim pretraživanjem. Dobiveni oblik podataka za određenu vrstu podataka se primijenio prilikom razvoja ontologije književnosti.

Izgradnja ontologije iz enciklopedijskih biografskih članaka hrvatske književnosti Hrvatske enciklopedije provela se iz opisanih strukturnih elemenata enciklopedijskog članka. Gradivni ontologijski elementi ekstrahirani su iz odabranog korpusa enciklopedijskih članaka. Istraživanje je ukazalo da natuknice članka predstavljaju *individue* ontologije s opisnim ontologijskim odnosom *ime*. Ukoliko određena osoba npr. posjeduje pseudonim ili nadimak to će se naznačiti u opisu *individue* kao *altLabel* opisni odnos. *Individue* generirane od natuknica predstavljat će članove klase *Književnik*, koja je u biti viši rodni pojam enciklopedičke definicije, kao i klasa pojedinih zanimanja npr. *Učitelj, Pedagog* itd. Dodatne obavijesti o natuknici (uglavnom lingvističke naravi, ukoliko su prisutne u članku) ručno su ekstrahirane istoznačnice i istovrijednice te dobiveni podaci predstavljaju opisna svojstva *individua* ontologije. Izneseno navodi na zaključak kako viši rodni pojmovi (nadređeni pojam, *genus proximum*) iz enciklopedičke definicije predstavlja *ontologijske klase* kojima pripadaju *individue* tj. natuknice članka (npr. *Književnik, Leksikograf, Pedagog* i sl.), a specifične

razlike tj. ograničavajuće svojstvo (*differentia specifica*) kao što su nacionalna pripadnost (hrvatski) predstavljaju individue ontologijske klase *Nacionalnost* (npr. Hrvat, Poljak i sl.) te *objektno* ontologijskog svojstvo o narodnosti. Podaci o mjestu i datumu rođenja te mjestu i datumu smrti predstavljaju *objektne, podatkovne* i *opisne* odnose ontologije. Podaci o godini rođenja i smrti također su zasebno izdvojeni kao *individue* klase *Godina* (podklasa klase *Vrijeme*) te su iz istih proizašli *podatkovni odnosi* za godine rođenja i smrti.

Tekst članka donosi činjenice iz životopisa relevantne za djelovanje (informacije koje se odnose na školsku formaciju i profesiju, doktorat, zvanje, članstvo u prestižnim znanstvenim, strukovnim i dr. društvima i udrugama, prestižne nagrade i priznanja, djelatnost) i činjenice o književnikovom opusu (djela navedena u reprezentativnom izboru i godine njihovog objavljivanja). Ručno su ekstrahirani podaci koji predstavljaju *klase* te *objektne, podatkovne* i *opisne ontologijske svojstva*. Podaci poput naslova djela i godine njihovog objavljivanja predstavljaju *individue* klase *Književno djelo*, dok su se ontologijski odnosi koji opisuju književno djelo i njegov odnos s književnikom načinili ručno. Dakle, navedena analiza je dokazala kako svaki strukturni element biografskog enciklopedičkog članka predstavlja određeni ontologijski gradivni element.

Sastavljen je popis svih termina za koje se žele načiniti tvrdnje ili koje se žele objasniti korisniku bez brige o međusobnom preklapanju pojmova koje predstavljaju, odnosa između termina, opisnih odnosa koje pojmovi mogu imati te se utvrdilo jesu li pojmovi klase ili svojstva. Označavanje važnih termina je u zavisnoj vezi s postavljanim dosegom (granicom) domene. Provedeno je izgrađivanje rječnika termina na biografskim enciklopedijskim člancima Hrvatske enciklopedije iz područja hrvatska književnost. Ovim postupkom je utvrđen skup termina koji su se uključili u ontologiju tj. skup termina uključuje sve relevantne termine domene (pojmove, instance, attribute koji predstavljaju svojstva pojmova, odnose između pojmova itd.), njihov opis prirodnim jezikom, sinonime itd.

Usljedili su postupci razvoja klasne hijerarhije i definiranje odnosa. Prvi korak u izgradnji ontologije je određivanje najopćenitijih ontologijskih pojmova, koji su klasificirani u skladu s podijelom proizašlom iz najviše razine temeljne PROTON ontologije koja započinje s temeljnim pojmom Entity, a isti se dalje dijeli na hijerarhijske podpojmove tj. ontologijske klase *Object*, *Abstract*, *Happening*. Podklase za klasu Object preuzete iz PROTON ontologije su: *Agent (Group, Person)*, *Location*, *PieceOfArt*, *Statement (InformationResource)*. Podklase za klasu *Abstract* preuzete iz PROTON ontologije su: *SocialAbstraction (ResearchArea)*,

Science). Podklase za klasu *Happening* preuzete iz PROTON ontologije su: *TimeInterval* (*CalendarYear*), *Situation*, *Event*.

Prilikom izgradnje pojmovne taksonomije, odnosno ontologijske klase i klasne hijerarhije proveo se odabir termina iz rječnika koji će predstavljati pojmove u pojmovnoj taksonomiji. U pojmovnoj taksonomiji su utvrđeni skupovi razdvojenih (engl. disjoint) pojmova tj. onih koji nemaju zajedničke instance. U razvoju pojmovne taksonomije, osim njezinog ručnog razvoja, primijenjen je i FCA pristup (formalna pojmovna analiza, engl. Formal Concept Analysis, njem. *Formale Begriffsanalyse* – FBA) za izgradnju ontologije. FCA se pokazao učinkovitim u analiziranju podataka i otkrivanju skrivene veze u podacima. Podaci su predstavljeni u obliku formalnog konteksta, odnosno jednostavnog određivanja koji atributi zadovoljavaju određene objekte. Detaljno je prikazan primjer polu-automatskog generiranja taksonomije ontologije primjenom formalne pojmovne analize. Izvršena je ekstrakcija podataka iz određenih tekstova kako bi se generirao odgovarajući formalni kontekst za odabrane tekstove i postavljene atribute. Prikazan je primjer na 37 odabranih tekstova o književnicima hrvatske književnosti iz Hrvatske enciklopedije, od kojih je njih 36 rođeno u Splitu, a jedan u Zagrebu (August Šenoa). Ekstrakcija informacija se obavila u pripremljenom Python programu u kojemu se izvršilo pretraživanje postavljenih atributa nad tekstovima te se automatski izvršila ekstrakcija prema zadanim atributima u obliku matrice automatski generirajući formalni kontekst iz zadanih objekata i atributa. Izvršena je provjera dobivenih rezultata. Objekti formalnog konteksta su tekstualni enciklopedijski članci u kojima su entiteti iz područja književnosti (tj. književnici) opisani prirodnim jezikom, a atributi su imenice (odnosno oblik određene riječi) za koje smo zainteresirani. Pretpostavka je da tekstualni opisi tj. enciklopedijski članci upotrebljavaju rječnik specifičan za područje književnosti koji se temelji na prethodno izgrađenoj ručnoj ontologiji. FCA pristupom se ustanovio skup pojmovnih svojstava kojima su definirani pojmovi u ontologiji te se izvršila klasifikacija pojmova u hijerarhiju. Isto tako se istaknula još jedna prednost primjene FCA analize, a to je generiranje novih i nepoznatih pojmova koje bi bilo teško ustanoviti samo ručnom izradom ontologije jer tekstovi specifični za domenu ne uključuju bilo kakvu imensku frazu za označavanje ovih pojmova.

Temeljna taksonomija pojmova ontologije hrvatske književnosti je zbog preglednosti prikazana kroz tzv. ontologijske module: *Agent*, *Književno djelo*, *Književni rodovi i vrste*, *Periodizacija književnosti* i *Dokument*.

Modulom *Agent* predstavljeni su pojmovi zanimanja i profesija kojima se osoba bavila i obavljala tijekom života, a njegov cilj je omogućiti razvrstavanje književnika iz enciklopedijskih članaka Hrvatske enciklopedije prema profesiji, zanimanjima i sl. kako bi se npr. znalo koji su hrvatski književnici ujedno pravници, arhitekti, lingvisti, slikari i sl. Ontologijski *Agent* modul uključuje *foaf: Agent* klasu. Podklasa od *foaf: Agent* je *foaf: Person*, koja predstavlja osobe. Ostale vrste agenata uključuju *foaf: Organization* i *foaf: Group*. Ontologijski *Agent* modul uključuje *foaf: Agent* kao primarnu nadklasu te uvodi *protont: Person* i *protont: Group* kao podklase. Nadalje, *protont: Person* je nadklasa definiranim podklasama dobivenim iz biografskih enciklopedijskih članak Hrvatske enciklopedije. S obzirom na vrstu entiteta načinjene su podklase koje idu direktno pod klasu *Person* te one koje se mogu grupirati kao podklase određene nadklase, kao što su *Književnik*, *Umjetnik*, *Zanimanja* i *Znanstvenik*. Podklase klase *Književnik* su dio pojmovne taksonomije koji ima za cilj razvrstavanje književnika iz enciklopedijskih članaka Hrvatske enciklopedije na pjesnike, esejiste, dramatičare, publiciste i sl. te su stoga izgrađene i njezine istoimene podklase.

Modul *Književno djelo* uključuje klasu *protont: PieceOfArt* (podklasa *protont: Object*), koja kao podklasu sadrži *Književno djelo*, a ova pak okuplja sva književna djela hrvatske književnosti u vidu ontologijskih individua.

Modul *Književni rodovi i vrste* razvrstava pojmove o književnim rodovima i vrstama, ekstrahirane iz enciklopedijskih članaka Hrvatske enciklopedije, u dio ontologijske hijerarhije koja se odnosi na književnosti. Ovim modulom su predstavljeni pojmovi teorije književnosti, a njegov cilj je omogućiti razvrstavanje književnih djela iz enciklopedijskih članaka Hrvatske enciklopedije prema književnim rodovima i vrstama. Ontologijski modul *Književni rodovi i vrste* uključuje *protont: Science* klasu kao primarnu nadklasu te uvodi klasu *Znanost o književnosti*, koja ima podklasu *Teorija književnosti*, a ova je pak nadklasa od *Književni rodovi i vrste*. Nadalje, klasa *Književni rodovi i vrste* sadrži podklase: *Didatičke vrste*, *Diskurzivni književni oblici*, *Drama*, *Epika*, *Lirika*, *Poezija*, *Proza* te navedene podklase također sadrže svoje podklase.

Modul *Periodizacija književnosti* enciklopedijske članke Hrvatske enciklopedije tj. hrvatske književnike i književna djela razvrstava u pojedina razdoblja i pravce, književne pokrete i generacije te dodjeljuje književnim djelima stilsko obilježje pojedinog razdoblja. Ontologijski modul *Periodizacija književnosti* uključuje *protont: Science* klasu kao primarnu nadklasu te uvodi njezinu podklasu *Znanost o književnosti*, koja ima podklasu *Povijest književnosti*, a ova je pak nadklasa od klase *Periodizacija književnosti*.

Modul *Dokument* razvrstava književna djela iz enciklopedijskih članaka Hrvatske enciklopedije u vrste publikacija, dokumenata i članaka. Ekstrahirani su pojmovi koji se odnose na vrste publikacija i dokumenata iz enciklopedijskih članaka te su isti poredani u hijerarhiju modula *Dokument*. Ontologijski modul *Dokument* uključuje *protont: Statement* klasu kao primarnu nadklasu te uvodi njezinu podklasu *protont:InformationResource*, koja ima podklasu *Dokument*, a ona pak svoje podklase detaljno prikazane u radu.

Nakon izbora klasa s popisa termina iz terminološkog rječnika, većina preostalih termina će biti svojstva klasa tj. činit će ontologijske odnose u ontologiji. Ovi termini su primjerice *mjesto (rođenja, smrti)*, *datum (rođenja, smrti)*, *godina (rođenja, smrti, izdanja)*, *književna vrsta*, *prepjev*, *autor*, *urednik* itd. Oblik navedenih svojstava je preinačen, stoga se u konačnici od prethodno navedenog javljaju npr. sljedeća svojstva: *mjesto rođenja*, *mjesto smrti*, *datum rođenja*, *datum smrti*, *godina rođenja*, *godina smrti*, *godina izdanja*, *jeKnjiževnaVrsta*, *jePrepjev*, *jeAutor*, *jeUrednik*. Odredilo se koju klasu opisuje svako svojstvo. Prilikom izrade ontologije prisutne su tri vrste odnosa: (1) opisni odnosi (engl. annotation svojstva) kojima opisujemo npr. ime klase ili individue (label), stavljamo komentar za određenu klasu ili individuu, dodajemo alternativno ime ili sinonim ukoliko ga klasa ili individua ima (altLabel) itd., (2) objektni odnosi (engl. object property) uspostavljaju odnos između dva pojma tj. individua koje pripadaju različitim klasama i (3) podatkovni odnosi (engl. data property) uspostavljaju odnos između individue i vrste podatka.

Analizom članaka uspostavljeni su *objektni odnosi* koji se mogu podijeliti u dvije veće skupine: prvu skupinu čine odnosi ekstrahirani iz samog teksta enciklopedijskih članka (svojstva izdavača, svojstva autora, svojstva djela i svojstva osobe), a drugu skupinu čine odnosi potrebni da bi ontologija mogla izvesti automatsko zaključivanje nad unesenim činjenicama (*imaOblik*, *imaZnačajke*, *jePoVeličini*, *seIzvodi*).

Svojstva osobe opisuju osobne podatke o nekoj osobi uspostavljajući npr. odnos između osobe i pojedine ustanove (npr. ustanove na kojima je osoba se školovala), odnose između dvije osobe, odnose između osobe i karakteristika njezinog književnog stvaralaštva. Svojstva osobe su podijeljena na podsvojstva *OsobaOsobniPodaci*, *OsobaDjelatnost*, *OsobaOsoba*, *OsobaStvaralaštvo*. Svako od ovih svojstava sadrži podsvojstva kojima se opisuje ono na što navedeno nadsvojstvo upućuje. Tako, primjerice svojstvo *OsobaOsobniPodaci* sadrži podsvojstva koja opisuju osnovne podatke o osobi (*jeMjestoRođenja*, *jeStudirao* itd.). Podsvojstvo *OsobaDjelatnost* opisuju poslove i djelatnosti za koje su se pojedini književnici obrazovali ili koje su obavljali tijekom svog djelovanja.

Podsvojstvo *OsobaOsoba* opisuje odnos između dvije osobe (mentorstvo, utjecaj, uzor, parafraziranje, posveta itd.). Svojstvo *OsobaOsoba* sadrži podsvojstvo *OsobaOdnosi* i podsvojstvo *ProučavanjeOsoba*. Podsvojstvo *OsobaOdnosi* definira odnos između dvije osobe, primjerice kći, kum, suprug. Podsvojstvo *ProučavanjeOsoba* definira proučavanje određene osobe od strane književnika (bibliografska obrada neke osobe, biografska obrada nekog književnika i sl.). Podsvojstvo *OsobaStvaralaštvo* omogućuje opis stvaralaštva određenog književnika, primjerice tematiku, jezik, oblik strofe, stih i sl. parametre koje je književnik koristio u svom stvaralaštvu.

Svojstva autora proizlaze iz odnosa klasa *Književnik* i *Književno djelo* te definiraju autora, prevoditelja, urednika, ilustratora i slična svojstva za neko književno djelo.

Svojstva djela proizlaze iz odnosa koje uspostavlja književno djelo s klasama *Person*, *Književnik*, podklasama klase *Periodizacija književnosti*, *Strofa*, *Književni rodovi i vrste*, *Izdavač*, *Organization*, *Figure*, *Stih*, *Društva*, *Kazalište*, *Jezik*, *Teme*, *Narječje*, sa svim podklasama klase *Dokument*, *Pismo*. *Svojstva djela* sadrži podsvojstva *DjeloOsoba* i *Opis djela*. *DjeloOsoba* povezuje određena djela s pojedinim osobama, sadrži podsvojstva koja ukazuju npr. na posvetu ili darovanje nekog djela određenoj osobi, pronalazak nekog djela od strane neke osobe i slične odnose. *Opis djela* uspostavlja odnose koje opisuju obilježja djela, npr. jezik, pismo, narječje djela i slične odnose.

Svojstva Izdavača sadrže podsvojstva *izdanoJeOdIzdavača* i *jeIzdao*. Svojstvo *izdanoJeOdIzdavača* povezuje individue klase *Književni rodovi i vrste* te klase *Književno djelo* s individuama klase *Književnici*, *Izdavač*, *Person*; a svojstvo *jeIzdao* je inverzno.

Skupinu *svojstva za ontologijsko zaključivanje* čine svojstva koja omogućuju zaključivanje o pojedinim ontologijskim klasama tj. entitetima i njihovu klasifikaciju na temelju njihovih pripadajućih svojstava. Svojstvo *imaOblik* omogućuje automatsko zaključivanje pripadaju li određene književne vrste definiranim klasama *Poeziji* ili *Prozi* tj. definira književno djelo koji može biti nestihovano ili u obliku stiha. Svojstvo *imaZnačajke* definira značajke književnih vrsta kao književnoumjetničke ili znanstvene značajke te omogućuje npr. klasifikaciju *Diskurzivnih književnih oblika*. Svojstvo *jePoVeličini* također omogućuje automatsko zaključivanje jer definira male i velike književne vrste te na taj način zajedno sa svojstvom *imaOblik* omogućuje njihovu klasifikaciju kao npr. podklase definirane klase *Poezija*. Svojstvo *seIzvodi* omogućuje automatsko zaključivanje da se pojedino djelo namijenjeno izvođenju na pozornici i klasifikaciju određenih ontologijskih entiteta kao podklasa definirane klase *Dramski* koja označava pripadnost dramskom rodu.

Podatkovni odnosi (engl. *Data Properties*) uspostavljaju odnos između individue sa doslovnim vrijednostima (npr. godina, datum, literal itd.). Razvijena je hijerarhija pododnosa, a vrh hijerarhije predstavlja relacija *owl:topDataProperty*. Razvijeni su podatkovni odnosi za opis književnika (*Književnik podaci*) i književnog djela (*Književno djelo podaci*), *godina utemeljenja* (odnosi se na organizaciju, društvo i sl.) i *vrijeme izvođenja predstave*.

Uspostavljeni podatkovni odnosi koji se odnose na književnika su smješteni u podhijerarhiju odnosa *Književnik podaci* i odnose se podatke tipa godina učlanjenja u neko društvo, godina dobivanja određene nagrade, razdoblje uređivanja nekog časopisa i slične odnose.

Uspostavljeni podatkovni odnosi koji se odnose na književno djelo su smješteni u podhijerarhiju odnosa *Književno djelo podaci* te se odnose se podatke tipa godina tiska, pretiska, izdanja, adaptacije djela i slične odnose.

Opisna svojstva opisuju individue, a u ontologiji za opis književnosti pojedinog naroda su nazvana *životopis*, *književnik*, *djelo*, *enciklopedijski članak*, *izvor znanja* te je svako od njih organizirano u svoju hijerarhiju podsvojstava.

Svojstvo *životopis* opisuje cijeli životopis neke osobe tako što prikazuje cijeli tekstualni dio enciklopedijskog članka koji donosi informacije o svim osobnim podacima, poslu i školovanju neke osobe. Time se dobiva opširniji odgovor o životopisu neke osobe. Međutim, kako bi se mogla dobiti detaljna i kratka informacija iz životopisa određenog književnika uvedena su i podsvojstva svojstva životopis: *osobni podaci*, *posao*, *školovanje*. Podsvojstvo *osobni podaci* sadrži primjerice podsvojstva *ime*, *nadimak*, *datum rođenja*, *datum smrti*, *ime člana Arkadije*, *prebivalište*, *boravište* i sl. Podsvojstvo *posao* posjeduje hijerarhiju podsvojstava koja daju informacije o npr. mjestu rada, vremenu predavanja, godini umirovljenja i slična svojstva. Podsvojstvo *školovanje* posjeduje hijerarhiju podsvojstava koja daju informacije o npr. mjestu studiranja, godini diplomiranja, naslovu doktorata, ustanovi doktorata, području doktorata, razdoblju studiranja i sl.

Opisno svojstvo *književnik* sadrži podsvojstva kojima će se opisati neki književnik, donoseći informacije o posebnom značaju književnika, temama koje su ga posebno nadahnjivale i slične odnose.

Opisno svojstvo *književno djelo* sadrži podsvojstva kojima će se opisati neko književno djelo, donoseći informacije o mjestu radnje djela, mjestu tiskanja djela, vremenu radnje djela, proširenom naslovu djela i sl.

Opisno svojstvo *enciklopedijski članak* sadrži podsvojstvo *sažetak* koje korisniku donosi sažetak o odabranom enciklopedijskom članku, prije njegovog potpunog čitanja.

Opisno svojstvo *izvor znanja* sadrži podsvojstva *wasDerivedFrom* kako bi se naveo URL web stranice iz koje su ekstrahirani podaci o određenoj individui i podsvojstvo *web stranica* kako bi se naveo URL stranice o nekoj individui.

Nakon izgradnje ontologije uslijedio je posljednji postupak stvaranja individualnih instanci klasa u hijerarhiji i njihov opis. Definiranje individualne instance klase zahtijevalo je izbor klase, stvaranje individualne instance te klase i popunjavanje vrijednosti svojstava.

Prikazana je sama struktura i slijed informacija biografskog enciklopedijskog članka kao skup ontologijskih odnosa poštujući temeljne dijelove enciklopedijskog članka (tj. glavu i tijelo članka). Primjećuje se kako se svi dobiveni odnosi neće primijeniti u opisu svih članaka nego će njihova primjena biti u skladu s potrebama opisa određene individue. Rad je potvrdio hipotezu da ipak postoji onaj nužan i dovoljan skup ontologijskih gradivnih elemenata koji čine rječnik za predstavljanje znanja o odabranom području. Osim prethodno navedenih ontologijskih modula, sljedeći ontologijski odnosi se pojavljuju u svakom enciklopedijskom biografskom članku s obzirom na njegovu strukturu:

- Glava članka (Natuknica članka: *label*, *altlabel*, *ime*),
- Tijelo članka iz definicija članka (*jePoNarodnosti*, *jeMjestoRođenja*, *datumRođenja*, *jeMjestoSmrti*, *datumSmrti*, *jeGodinaRođenja* i *jeGodinaSmrti*) i teksta članka (*životopis*, *boravište*, *školovanje*, *posao*, *jeRadio*, *naslov djela*).

Predlaže se sagledavanje navedenih obaveznih elementa članka prikazanih u vidu ontologijskih svojstava kao mogućih gradivnih svojstava za izgradnju infookvira književnih biografskih enciklopedijskih infookvira Hrvatske enciklopedije. Preporučuje se uvođenje u infookvir i svojstava bitnih za isticanje i opis posebnosti određenog književnika, iako se radi o neobveznim elementima članka: *pseudonim*, *jePisaoKnjiževneVrste*, *jeNagrađen*, *jeTema književnika*.

Rad je prikazao svrhu upotrebe LSA metode u izgradnji ontologije. Latentna semantička analiza (LSA) doslovno označava analiziranje dokumenata kako bi se pronašlo temeljno značenje ili pojmovi tih dokumenata. Latentna semantička analiza je izrasla iz problema pronalaska relevantnih dokumenata riječima za pretraživanje jer temeljna poteškoća nastaje kada se uspoređuju riječi za pronalazak relevantnih dokumenata, a ne ono što se stvarno želi, usporedba *značenja ili pojmova* iza tih riječi. LSA pokušava riješiti ovaj problem pridruživanjem (mapiranjem) i riječi i dokumenata u *pojmovni prostor* i usporedbom u ovom

prostoru. LSA uvodi neka pojednostavljena: (1) dokumenti se prikazuju kao "vreća riječi", gdje je redoslijed riječi u dokumentu nevažan, bitno je samo koliko se puta svaka riječ pojavljuje u dokumentu i (2) pojmovi su predstavljeni kao uzorci riječi koje se obično pojavljuju zajedno u dokumentima. Iz prethodno navedenog proizlazi da će se riječi i dokumenti sa sličnim značenjem i tematikom grupirati u istom klasteru. Na temelju formilarnih klastera se mogu izvući određeni pojmovi. Nakon provedene LSA metode na odabranim dokumentima uočava se njezina korisnost u određivanju pojma kao skupa srodnih termina te pripadnost pojedinih natuknica članaka (dokumenata) tom pojmu, čime se omogućuje automatska klasifikacija članaka pojedinim ontologijskim klasama. Ovim pristupom se vrši automatsko klasificiranje članaka određenoj klasi na temelju srodnih termina. Dakle, određivanje pripadnosti pojedinih termina pojedinom pojmu omogućena je automatska klasifikacija članaka pojedinoj ontologijskoj klasi na temelju obrazaca riječi koje sadrže odabrani članci, što je prikazano na odabranom primjeru. Odabrano je šesnaest članka Hrvatske enciklopedije, čiji su naslovi tj. natuknice članaka označeni kao D1, D2, ..., D16. Ekstrahirane su natuknice (prezime i ime književnika tj. naslovi članaka koji predstavljaju dokumente) i termini iz područja književnosti koji se odnose na vrstu književnog djela koju su pisali pojedini autori, a sami termini vrste književnog djela se nalaze u već prethodno izgrađenoj ontologiji. Kao ključne tj. indeksne riječi prilikom provedbe LSA metode odabrani su termini koji se pojavljuju u dva ili više članaka. Stvorena je termin-dokument matrice od odabranih članaka (dokumenata) u kojoj je svaka indeksna riječ predstavljena u recima, a svaka natuknica tj. naslov pojedinog članka (koja predstavlja pojedini članak, odnosno dokument) je prezentirana u stupcima. Svaka ćelija matrice sadrži broj puta pojavljivanja riječi u određenom naslovu (dokumentu). Primijenjena mjera za vaganje težine riječi je TF-IDF mjera (termin frekvencija – inverzna frekvencija dokumenta). Nakon izračuna težine pojedinih termina izvela se dekompozicija matrice na singularne vrijednosti (SVD), koja pronalazi reducirani dimenzionalni prikaz matrice koji naglašava najjače odnose i odbacuje buku. Trik u korištenju SVD-a je u pronalaženje broja dimenzija ili "pojmovi" koje će se upotrijebiti kada se aproksimira matrica. SVD faktorizira matricu na tri matrice. *Termin matrica* U daje koordinate svake riječi u *pojmovnom prostoru*, *dokument matrica* V daje koordinate svakog dokumenta u *pojmovnom prostoru* i S matrica singularnih vrijednosti daje naslutiti koliko je dimenzija ili pojmova potrebno uključiti. U ovom primjeru su se koristile tri dimenzije s tim da se prva dimenzija izbacila te se grafički prikazala druga i treća dimenzija. Razlog zbog kojeg se odbacuje prva dimenzija je jer ona korelira s duljinom dokumenta, a za riječi korelira s brojem puta koliko se riječ pojavljuje u svim člancima

(dokumentima), te nije tako informativna kao druga i treća dimenzija. Prednost ove tehnike je da su indeksne riječi i natuknice članaka (prezime i ime književnika tj. dokumenti) prikazani istim grafom, što omogućuje grupiranje skupine indeksnih riječi i književnika (dokumenata) u klaster, a time i imenovanje klastera tj. pojma koji okuplja srodne termine i dokumente. Prema rezultatima za provedeni primjer dobivena su četiri klastera koja su imenovana *Diskurzivni književnici*, *Prozaici*, *Dramatičari* i *Pjesnici*. Svaki klaster obuhvaća određene termine i članke o književnicima. Tako primjerice, klaster *Diskurzivni književnici* obuhvaća termine kritika, publicist, putopis, feljton, esejist, kritičar, esej te članke o književnicima koji su u svom stvaralaštvu stvarali djela koja pripadaju diskurzivnim književnim oblicima. Dakle, srodni termini iz svakog klastera su se oblikovali pod jedan pojam koji se imenovao kao *Diskurzivni književnici*, *Prozaici*, *Dramatičari* i *Pjesnici*. Iz grafičkog prikaza se uočava da npr. književnik (članak D16) koji je pored diskurzivnih djela pisao i npr. pjesme ulazi i u klaster *Pjesnika* tj. navedeni članak D16 je smješten bliže riječima lirska i pjesma koje definiraju klaster pjesnika.

Rezultat istraživanja je ontologija za opis književnosti određenog naroda predstavljena kroz prikaz ontologijskih modula i pripadajućih im ontologijskih odnosa. Rezultati dobivenog ontologijskog modela, odnosno temeljnih ontologijskih klasa, tabelarno su prikazani sa pripadajućim im ontologijskim, podatkovnim, i opisnim svojstvima za svaku od njih u svrhu lakšeg upravljanja ontologijom književnosti prilikom njezine upotrebe. Konačni rezultati dobivenih ontologijskih odnosa su prikazani kroz primjenu tako dobivene ontologije u opisu individue, kao i kroz mogućnost postavljanja složenih semantičkih pitanja i pregledavanja nestrukturiranih enciklopedijskih biografskih članak, što do sada nije bilo moguće. Dobivena ontologija omogućuje opis individua tj. članaka kroz strukturirane podatke do različitih razina detaljnosti, što omogućuje prilikom pretraživanja dobivanje većeg dijela teksta enciklopedijskog biografskog članka ili dobivanje ciljanih i kratkih odgovora iz enciklopedijskog biografskog članka. Razvijena ontologija književnosti također omogućuje i postavljanje složenih upita s više zadanih svojstava (uvjeta upita) te dobivanje preciznih i smislenih odgovora, što do sada nije bilo moguće u trenutno dostupnoj mrežnoj online elektroničkoj Hrvatskoj enciklopediji.

Rad je području enciklopedistike donio višestruke dobrobiti: utvrđeno je koji elementi enciklopedičke organizacije znanja u biografskim člancima mogu biti podrška razvoju ontologija, uspostavljena je metoda za generiranje ontologije i strukturiranih podataka iz nestrukturiranih tekstova enciklopedijskih članaka Hrvatske enciklopedije, razvijena je

ontologija za opis književnosti određenog naroda koja u budućnosti može predstavljati okosnicu semantičke interoperabilnosti i integracije informacija svih elektroničkih izdanja iz navedenog područja, razvijen je konceptualni model za enciklopedije koji će omogućiti stvaranje svih semantičkih odnosa između pojmova i instanci odabranog područja te će na taj način u potpunosti omogućiti opisivanje znanja pohranjenog unutar članaka područja nacionalne književnosti nekog enciklopedijskog izdanja i osigurati će se generiranje strukturiranih podataka za enciklopedijske članke temeljene na tako razvijenom modelu, omogućeno je uspostavljanje interoperabilnosti i povezivanje *Linked Data* koncepcijom hrvatske enciklopedistike s ostalim strukturiranim izvorima enciklopedijskog znanja na mreži (npr. DBpedia, Freebase, sl.), poboljšana je organizacija enciklopedijskog znanja u mrežnom okruženju i mogućnosti semantičkog pretraživanja samog izvora, izvršena je revalorizacija uloge enciklopedičkih djela u suvremenom informacijskom prostoru od tradicionalnog tercijarnog izvora informacija prema alatima za povezivanje informacija prezentiranih u tom prostoru i snalaženje u njemu, uspostavljeni su temelji za proširenje bilo kakve vrste odnosa ili pridruživanja pojedinih termina u odvojenim ontologijama čime će se osigurati mnogostruke mogućnosti visokog stupnja upravljanja enciklopedijskim znanjem, uspostavljeni su temelji za mogućnost razvoja semantičkog označavanja tekstova za hrvatski jezik na temelju ontologije i strojnog učenja, doprinijelo se razvoju i izgradnji boljeg informacijskog sustava temeljenog na ontologiji tj. formalnoj reprezentaciji semantike iz navedenog područja te je omogućeno Leksikografskom zavodu Miroslavu Krležu (LZMK) povezivanje normiranog znanja sa enciklopedijskim znanjem u otvorenom pristupu u onoj mjeri koju LZMK u danom trenutku želi, a koja će pridonijeti vidljivosti njegovih enciklopedijskih izdanja.

Navedene višestruke dobrobiti i mogućnosti daljnjeg razvoja su razlozi za primjenu i provođenje daljnjih ontologijskih istraživanja u okviru enciklopedijskog znanja.

11 LITERATURA

About Schema.org. Dostupno na URL: <http://schema.org/docs/about.html> (01.04.2015).

About the Stanford Encyclopedia of Philosophy. Dostupno na URL: <http://plato.stanford.edu/about.html> (01.04.2015).

ACE 2003 Evaluation Plan. Dostupno na URL: http://jaguar.ncsl.nist.gov/ace/doc/ace_evalplan-2003.v1.pdf (18.09.2015.).

Ahmed, S.; Kim, S.; Wallace, K. A methodology for creating ontologies for engineering design / Journal of Computing and Information Science in Engineering 7, 2(2007), 132-140. Dostupno na URL: <http://computingengineering.asmedigitalcollection.asme.org/article.aspx?articleid=140> (8.2.2016).

Aktors. Dostupno na URL: <http://www.aktors.org/publications/ontology/> (15.06.2015.).

Allemang, Dean; Hendler, James A. Semantic web for the working ontologist : modeling in RDF, RDFS and OWL. Amsterdam ; Boston : Morgan Kaufmann Publishers/Elsevier, 2008.

Allemang, Dean; Hendler, Jim. Semantic Web for the working Ontologist: Effective modeling in RDFS and OWL. *2nd ed.* Waltham, MA : Morgan Kaufmann/Elsevier, 2011.

Allen, James F. Towards a general theory of action and time // *Artificial Intelligence*, 23(1984), 123-154.

Amer-Yahia, Sihem...et al. iO2: An algorithmic method for building inheritance graphs in object database design. // Proceedings of ER'96 (15th International Conference on Conceptual Modeling) , 1157. Berlin [etc.] : Springer, 1996.

Antoniou, Grigoris; Van Harmelen, Frank. A Semantic Web Primer. 2. ed. Massachusetts London, England: The MIT Press Cambridge, 2008.

Antoniou, Grigoris; Van Harmelen, Frank. Web Ontology Language: OWL. Dostupno na URL: <http://www.cs.vu.nl/~frankh/postscript/OntoHandbook03OWL.pdf> (16.09.2015.).

- A Platform for Scalable, Collaborative, Structured Information Integration / Kurt Bollacker...[et al.] Dostupno na URL: <http://aaaipress.org/Papers/Workshops/2007/WS-07-14/WS07-14-004.pdf> (18.09.2015.).
- BabelNet. Dostupno na URL: <http://www.babelnet.org> (3.06.2016).
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier. Modern information retrieval : the concepts and technology behind search. 2nd ed. Harlow [etc.] : Pearson Education, 2011.
- Barker, Phil; Campbell, Lorna M. What is schema.org? A Cetus Briefing Paper for LRMI Dostupno na URL: <http://publications.cetis.ac.uk/wp-content/uploads/2014/06/schemaBriefing.pdf> (1.04.2015.).
- Barzun, Jacques. Od osvita do dekadencije: 500 godina zapadne kulture. Zagreb: Masmmedia, 2003.
- Basic Geo Vocabulary. Dostupno na URL: <http://www.w3.org/2003/01/geo/> (18.09.2015.).
- Bateman, John A...et al. A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model. // USC/Information Sciences Institute. Marina del Rey, 1990. Dostupno na URL: <http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/um89/UM89-penman.pdf> (8.2.2016.).
- Becker, Christian; Christian Bizer. DBpedia Mobile: A Location-Enabled Linked Data Browser. Dostupno na URL: <http://www4.wiwiss.fu-berlin.de/bizer/pub/BeckerBizer-DBpediaMobile-Ldow2008.pdf> (18.09.2015.).
- Bekavac, B.; Tadić, M.; Šojat, K.: *Zašto nam treba Hrvatski WordNet? // Semantika prirodno jezik i metajezik semantike: zbornik.* / ur. Jagoda Granić. Zagreb: Hrvatsko društvo za primijenjenu lingvistiku, 2005. Dostupno na URL: <http://www.hnk.ffzg.hr/txts/ksbbmt4HDPL2004.pdf> (3.04.2015.).
- Belkin, Nicholas J.; Croft, W. Bruce. Retrieval techniques. // Annual Review of Information Science and Technology (ARIST), vol. 22, chapter 4, pages 109–145. / M. Williams, editor . New York : Elsevier Science Publishers, 1987.

- Belohlavek, Radim. Introduction to Formal Concept Analysis. Olomuc, 2008. Dostupno na URL: belohlavek.inf.upol.cz/vyuka/IntroFCA.pdf (01.02.2015.).
- Bendaoud, Rokia; Napoli, Amedeo; Toussaint, Yannick. Formal Concept Analysis: A unified framework for building and refining ontologies // Knowledge Engineering: Practice and Patterns : 16th International Conference, EKAW 2008, Acitrezza, Sicily, Italy September 29 - October 3, 2008, Proceedings. / Aldo Gangemi, Jérôme Euzenat, (Eds.). Berlin ; New York : Springer, 2008.
- Beneš, M., Snášel, V. Deducing Design Class Hierarchy from Object Properties. ISM'2002. Rožnov pod Radhoštěm. Czech Republic. 2002.
- Bergman, M.K. UMBEL: A Subject Concepts Reference Layer for the Web. Dostupno na URL: <http://www.slideshare.net/mkbergman/umbel-a-subject-concepts-reference-layer-for-the-web> (26.10.2015.).
- Bernaras, A.; Laresgoiti, I.; Corera, J. Building and Reusing Ontologies for Electrical Network Applications // ECAI 96 : 12th European Conference on Artificial Intelligence, August 11-16, 1996, Budapest, Hungary : proceedings / edited by Wolfgang Wahlster. Chichester ; New York : Wiley, 1996.
- Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. // Scientific American Magazine 284, 5 (2001), 28–37.
- Berners-Lee, T., Fielding, R. RFC3986 - Uniform Resource Identifier (URI): Generic Syntax. // W3C/MIT, 2005. Dostupno na URL: <https://tools.ietf.org/html/rfc3986> (10.9.2015.).
- Bizer, Christian; Heath, T.; Berners-Lee, T. Linked Data - the story so far. // International Journal on Semantic Web and Information Systems 5, 3(2009), 1–22. Dostupno na URL: <http://www.igi-global.com/article/linked-data-story-far/37496> (27.03.2015.).
- Berry, Michael W.; Dumais, Susan T.; O'Brien, Gavin W. Using linear algebra for intelligent information retrieval. *SIAM Review* 37, 4(1995), :573–595.
- Bizer, Christian et al. DBpedia - A crystallization point for the Web of Data // Web Semantics : Science, Services and Agents on the World Wide Web 7, 3(2009), 154-165.
- Blázquez, M...et al. Building Ontologies at the Knowledge Level using the Ontology Design Environment. // Proceedings of the 11th Banff Knowledge Acquisition for

- Knowledge-Based Systems Workshop (KAW'98), 2, 1998. Dostupno na URL: [http://oa.upm.es/6457/1/Building Ontologies at the K.pdf](http://oa.upm.es/6457/1/Building_Ontologies_at_the_K.pdf) (8.2.2016.).
- Blyth, T. S. *Lattices and Ordered Algebraic Structures.* / T.S. London: Springer, 2005.
- Bogutovac, Antonijela. Razvoj definicija pojmova astronomija, fizika i matematika u hrvatskim jednojezičnim rječnicima. // *Studia lexicographica* 2, 1(2), 2008., 51-75.
- Borgo, Stefano; Guarino, Nicola; Masolo, Claudio. An Ontological Theory of Physical Objects // *Proceedings of Eleventh International Workshop on Qualitative Reasoning (QR'97)* / ed. by L. Ironi. Cortona (Italia), 3-6 Giugno, 1997.
- Borgo, Stefano; Guarino, Nicola; Masolo, Claudio. A Pointless Theory of Space Based on Strong Connection and Congruence // *Principles of Knowledge Representation and Reasoning (KR96)* / ed. by L. C. Aiello and J. Doyle, Morgan Kaufmann, 1996. Dostupno na URL: <http://www.loa.istc.cnr.it/old/Papers/KR96.pdf> (27.11.2015.).
- Burke, Peter. *A Social History of Knowledge: From Gutenberg to Diderot.* London: Polity Press u suradnji s Blackwell Publishers, 2000.
- Casati, R.; Varzi, A. *Holes and Other Superficialities.* Cambridge, Mass. : MIT Press, 1994.
- Chena, Rung Ching; Baub, Cho Tsan; Huang, Yun H. Development of Anti-diabetic Drugs Ontology for Guideline-based Clinical Drugs Recommend System Using OWL and SWRL. // In 2010 IEEE World Congress on Computational Intelligence, WCCI 2010 (2010), 5-10.
- CIDOC Conceptual Reference Model (CRM). Dostupno na URL: http://www.cidoc-crm.org/cidoc_core_graphical_representation/graphical.html (30.10.2015.).
- CIDOC CRM Applications. Dostupno na URL: http://www.cidoc-crm.org/uses_applications.html (30.10.2015.).
- Citizendium. Dostupno na URL: <http://www.citizendium.org/> (15.06.2015.).
- Classification & ontology: formal approaches and access to knowledge : proceedings of the International UDC Seminar, 19 - 20 September 2011, The Hague, The Netherlands / edited by Aida Slavić and Edgardo Civallero. Wuerzburg : [Ergon Verlag](http://www.ergon-verlag.de/), 2011.

- Collison, Robert L. *Encyclopaedias: Their History Throughout the Ages*. 2. izdanje. New York: Hafner Publishing Co., 1966.
- Conesa, Jordi; Storey, Veda C.; Sugumaran, Vijayan. Usability of upper level ontologies: The case of ResearchCyc. // *Data and Knowledge Engineering* 69, 4(2010), 343-356.
Dostupno na URL: <http://www.sciencedirect.com/science/article/pii/S0169023X09001177> (15.06.2015.).
- Crnković, Kristijan. INDIGO – platforma za upravljanje i objavu digitalnih zbirki, vođenje primarne i sekundarne muzejske dokumentacije // *Dokumentacija i korisnici*, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.
- Cruse, D. A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.
- Cycorp. Dostupno na URL: <http://www.cyc.com/> (27.8.2015.).
- Davies, J.; Boncheva, K.; Manov, D. D5.0.1 Ontology Engineering in SEKT (informal project report), 2004. Dostupno na URL: [http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20\(BULO\)%20Guidance.pdf](http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20(BULO)%20Guidance.pdf) (10.03.2015.).
- Davies, J.; Lytras, M.; Sheth, A. P. *Semantic Web Based Knowledge Management*. // *IEEE Internet Computing* 11, 5(2007), 14-16.
- DBpedia. Dostupno na URL: <http://wiki.dbpedia.org/> (26.10.2015.).
- DBpedia. Dostupno na: <http://dbpedia.org/About> (18.09.2015.).
- Dbpedia: A nucleus for a web of open data / Sören Auer...[et al.] // *The Semantic Web – ISWC 2007 + ASWC 2007, 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference*, volume 4825, 2007, 715-728. Dostupno na URL: <http://iswc2007.semanticweb.org/papers/715.pdf> (17.06.2015.).
- DBpedia Mappings. Dostupno na URL: http://mappings.dbpedia.org/index.php/Main_Page#Tutorials (18.09.2015.).
- DBpedia ontology. Dostupno na URL: <http://wiki.dbpedia.org/Ontology?v=zj4> (18.09.2015.).

- Deerwester, Scott... et al. Indexing by latent semantic analysis. // Journal of the American Society for Information Science 41, 6 (1990), 391–407. Dostupno na URL: http://www.cob.unt.edu/itds/faculty/evangelopoulos/dsci5910/LSA_Deerwester1990.pdf (15.02.2015.).
- Dejanović, Antonija; Svrtan, Zoran. Muzejski standardi u Hrvatskoj (promišljanje, stvaranje, izrada) // Dokumentacija i korisnici, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.
- Doerr, Martin; LeBoeuf, Patrick. Modelling Intellectual Processes: The FRBR-CRM Harmonization // Digital Libraries: Research and Development 4877, 2007, 114-123. Dostupno na URL: http://www.cidoc-crm.org/docs/doer_le_boeuf.pdf (30.10.2015.).
- Dublin Core Metadata Initiative (DCMI). DCMI Metadata Terms. Dostupno na URL: <http://dublincore.org/documents/dcmi-terms/> (26.11.2015.)
- Dublin Core Metadata Initiative (DCMI). *Library application profile*. 10 September 2004. Dostupno na URL: <http://dublincore.org/documents/library-application-profile/> (26.10.2015.).
- Dumais, Susan T. Improving the retrieval of information from external sources. Dostupno na URL: <http://machinelearningtext.pbworks.com/w/file/47378287/improvingRetrievalBRMIC91.pdf> (26.10.2015.).
- Duval, Erik...[et al.]. Metadata principles and practicalities. // D-Lib Magazine 8, 4(2002). Dostupno na URL: <http://www.dlib.org/dlib/april02/weibel/04weibel.html> (26.10.2015.).
- D5: Databases and Information Systems: FaQ. Dostupno na URL: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/faq/> (17.06.2015.).
- Enciklopedija // Hrvatska enciklopedija, mrežno izdanje / glavni urednik Slaven Ravlić. Zagreb: Leksikografski zavod Miroslav Krleža, 2013-2015. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=17879> (18.11.2015.).
- Enciklopedija: opća i nacionalna: u 20 knjiga / [glavni urednik Antun Vujić]. Zagreb: PRO LEKSIS: Večernji list, 2005-2007.

Enciklopedija [Elektronička građa]: opća i nacionalna u 20 knjiga / [glavni urednik Antun Vujić]. Zagreb: PRO LEKSIS: Večernji list, 2009.

Engelsing, Rolf. Der Bürger als Leser. Lesergeschichte in Deutschland, 1500–1800. Stuttgart: J. B. Metzlersche i A. E. Poeschel, 1974.

EUROWORDNET: A multilingual database with lexical semantic networks / ed. by Piek Vossen. Dordrecht, Boston, London : Kluwer Academic Publishers, 1998.

Faceted Wikipedia Search / Hahn, Rasmus...[et al.]. Dostupno na URL:
<http://www4.wiwiss.fu-berlin.de/bizer/pub/hahn-et-al-faceted-wikipedia-search-BIS2010.pdf> (18.09.2015.).

Falbo, Ricardo De Almeida. Experiences in Using a Method for Building Domain Ontologies. // SEKE 2004 : the 16th International Conference on Software Engineering & Knowledge Engineering : proceedings, technical program, July 20-24, 2004, Banff Alberta, Canada. Str. 474-477. Skokie, Il. : Knowledge Systems Institute, 2004.

Fellbaum, Christiane. A Semantic Network of English: The Mother of All WordNets. // EuroWordNet: A Multilingual Database with Lexical Semantic Networks. / ed. by Piek Vossen. Dordrecht: Boston: London: Kluwer Academic Publishers, 1998.

Fernández-Lopez, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. // AAAI-97 Spring Symposium Series SS-97-06, 1997, 33-40. Dostupno na URL:
http://oa.upm.es/5484/1/METHONTOLOGY_.pdf (8.2.2016.).

Fernández-López, M. ...et al. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. IEEE Intelligent Systems and Their Applications 14, 1(1999), 37-46. Dostupno na URL:
<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=747904> (8.2.2016.).

FOAF Vocabulary Specification. Dostupno na URL:
http://xmlns.com/foaf/spec/#term_Person (15.06.2015.).

Formal Concept Analysis: 5th International Conference, ICFCA 2007, Clermont-Ferrand, France, February 12-16, 2007: proceedings / S.O. Kuznetsov, S. Schmidt (Eds.). Berlin ; New York: Springer 2007.

Formal Concept Analysis - Foundations and Applications / (Eds.) Bernhard Ganter; Gerd Stumme, Rudolf Wille. Berlin Heidelberg: Springer-Verlag, 2005.

Formal Concept Analysis: seventh international conference, ICFCA 2009 Darmstadt, Germany, May 21-24, 2009: proceedings / S. Ferre, S. Rudolph (Eds.). Berlin ; New York: Springer 2009.

Formal Concept Analysis: third international conference, ICFCA 2005, Lens, France, February 14-18, 2005: proceedings / Bernhard Ganter, Robert Godin, (Eds.). Berlin ; New York: Springer 2005.

FRBRoo Introduction. Dostupno na URL: http://www.cidoc-crm.org/frbr_inro.html (30.10.2015.).

Freebase. Dostupno na URL: <https://www.freebase.com/> (17.06.2015.).

Freebase API. Dostupno na URL: <https://developers.google.com/freebase/> (3.06.2016.).

Friend of a Friend (FOAF). Dostupno na URL: <http://www.foaf-project.org/> (18.09.2015.).

Ganter, Bernhard; Wille, Rudolf. Formal Concept Analysis: mathematical foundations. Berlin: Springer-Verlag, 1999.

Gašević, Dragan; Djurić, Dragan; Devedžić, Vladan. Model Driven Architecture and Ontology Development. Berlin: Springer-Verlag, 2006.

GeoNames. Dostupno na URL: <http://www.geonames.org/> (17.06.2015.).

GeoNames Ontology. Dostupno na URL: <http://www.geonames.org/ontology/documentation.html> (26.10.2015.).

GeoWordNet: A Resource for Geo-spatial Applications. / by Fausto Giunchiglia...[et al.]. // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6088 LNCS, 2010, 121-136. Dostupno na URL: http://link.springer.com/chapter/10.1007%2F978-3-642-13486-9_9 (17.06.2015.).

Giles, Jim. Internet encyclopaedias go head to head: special report // Nature 438, 15(2005),

900–901. Dostupno na URL:

<http://www2.stat.unibo.it/mazzocchi/macroeconomia/documenti/Nature%202006%20wikipedia.pdf> (15.06.2015.).

Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval. Dostupno na URL:

<http://www.willpowerinfo.co.uk/glossary.htm> (26.11.2015.).

Golub, Gene H.; Van Loan, Charles F. Matrix Computations. 3rd ed. Baltimore, London : Johns Hopkins University 1996. Dostupno na URL:

<http://web.mit.edu/ehliu/Public/sclark/Golub%20G.H.,%20Van%20Loan%20C.F.-%20Matrix%20Computations.pdf> (15.02.2016.).

Gomes, P. Software Design Retrieval Using Bayesian Networks and WordNet. // Advances in Case-Based Reasoning. 7th European Conference, ECCBR. Proceedings./ ed. by P. Funk, P. González Calero. Berlin: Springer, 2004.

Gómez-Pérez, A. A Framework to Verify Knowledge Sharing Technology. Expert Systems with Application 11, 4(1996), 519-529.

Gomez-Perez, A. Knowledge Sharing and Reuse: ontologies and applications / A Tutorial on Ontologies Engineering, IJCAI, 1999.

Gómez-Pérez A.; Fernández López, M.; Corcho, O. Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the semantic web. London ; New York: Springer, 2004.

Gruber, Thomas R. A Translation Approach to Portable Ontology Specifications. // Knowledge Acquisition 5, 2 (1993), 199–220.

Gruber, T. Toward principles for the design of ontologies used for knowledge sharing. // International Journal of Human-Computer Studies 43, 5-6(1995), 907-928.

Gruber, Thomas; Liu, Ling; Özsu, M. Tamer. Ontology // In Encyclopedia of Database Systems / ed. by Ling Liu and M. Tamer Özsu. New York: Springer-Verlag, 2009. Dostupno na URL: <http://www.dainf.ct.utfpr.edu.br/~tacla/Ontologias/2013/a01-020-Ontology-definitionTomGruber.pdf> (19.09.2015.).

Gruninger, Michael; Fox, Mark S. Methodology for the Design and Evaluation of Ontologies // International Joint Conference on Artificial Intelligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal, Canada, 1995. Dostupno na URL: <http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf> (29.01.2016.)

Van Harmelen, Frank... et al. Web Ontology Language (OWL) Reference. // W3C Recommendation 10, February 2004. Dostupno na URL: <http://www.w3.org/TR/owl-ref/> (1.2.2015.).

Hodge, Gail. Systems of Knowledge Organization for Digital libraries: beyond traditional authority files. Washington, DC: the Council on Library and Information Resources. Dostupno na : <http://www.clir.org/pubs/reports/pub91/contents.html> (14.05.2015.).

Hrvatska enciklopedija. Zagreb: Leksikografski zavod Miroslav Krleža, 1999- 2009.

Hrvatska enciklopedija: uredništvo, organizacija, upute, Zagreb: Leksikografski zavod Miroslav Krleža, 1998.

Hrvatska kulturna baština. Dostupno na URL: <http://www.kultura.hr> (6.06.2016.).

Hrvatski WordNet (CroWN). Dostupno na URL: <http://crown.ffzg.hr> (3.06.2016.).

IEEE – About. Dostupno na URL: <http://www.ieee.org/about/index.html> (3.06.2016.).

IEEE P1600.1 Standard Upper Ontology (SUO) Working Group. Dostupno na URL: <http://suo.ieee.org/> (3.06.2016.).

IETF (Internet Engineering Task Force). RFC 3066: Tags for the Identification of Languages / ed. H. Alvestrand. 2001. Dostupno na URL: <http://www.ietf.org/rfc/rfc3066.txt> (26.11.2015.).

Information Retrieval: Data Structures & Algorithms. / William B. Frakes, Ricardo Baeza-Yates (editors). Englewood Cliffs, New Jersey : Prentice Hall, , 1992.

Integrating descriptions of knowledge management learning activities into large ontological

structures : a case study / Sicilia, Miguel-Ángel...[et al.] // Data & Knowledge Engineering 57, 2(2006), 111-121.

InPhO Project. Dostupno na URL: <http://inpho.cogs.indiana.edu/> (27.03.2015.).

ISO 21127:2006 (Information and documentation — A reference ontology for the interchange of cultural heritage information) Dostupno na URL: http://www.iso.org/iso/catalogue_detail?csnumber=34424 (21.08.2015.).

Ivančević, Zvonimir; Jelić, Slobodan. Primjena SVD rastava matrice u dohvatu informacija i kompresiji slike. Dostupno na URL: http://www.mathos.unios.hr/~sjelic/papers/SVD_ir_ic.pdf (16.02.2015.)

Jaschik, Scott. A stand against *Wikipedia*, *Inside Higher Ed* (January), 2007. Dostupno na URL: <http://www.insidehighered.com/news/2007/01/26/wiki> (15.06.2015.).

Jovic, Alan; Prcela, Marin; Gamberger, Dragan. Ontologies in Medical Knowledge Representation. // Proceedings of the *ITI 2007* 29th International Conference on Information Technology Interfaces, IEEE 2007, Cavtat. Dostupno na URL: <https://bib.irb.hr/datoteka/301358.iti2007jovicprcelagamberger.pdf> (27.11.2015.).

Kinzler, Daniel. Wikisense - mining the wiki // In Proceedings of Wikimania 2005. The First International Wikimedia Conference, 2005.

Knight, Kevin; Luk, Steve K. Building an Large Knowledge Base for Machine Translation. // Proceedings of the American Association of Artificial Intelligence Conference (AAAI-94). Seattle (USA). 1994. Dostupno na URL: <https://www.aaai.org/Papers/AAAI/1994/AAAI94-118.pdf> (8.2.2016.).

Knight, Kevin...et al. Filling Knowledge Gaps in a Broad-Coverage MT System. Proceedings of the 14th IJCAI Conference. Montreal (Canada). 1995. Dostupno na URL: <http://arxiv.org/pdf/cmp-lg/9506009.pdf> (8.2.2016.).

KSL Ontology Server Projects. Dostupno na URL: <http://www.ksl.stanford.edu/software/ontolingua/ontology-server-projects.html> (27.8.2015.).

Legg, Catherine. Ontologies on the semantic web // Annual Review of Information Science and Technology 41(2007), Chapter 9, 407-452.

Leksička semantika u izgradnji Hrvatskog WordNeta. Dostupno na URL:
<http://hnk.ffzg.hr/rmjt/p3html> (3.06.2016.).

Leksikografski zavod Miroslav Krleža // Hrvatska enciklopedija: mrežna verzija. Dostupno na URL: <http://www.enciklopedija.hr/Natuknica.aspx?ID=35937> (31.05.2016.).

Lenat, D. B. Cyc : a large-scale investment in knowledge infrastructure. // Communications of the ACM, 38, 11(1995), 32-59.

Lenat, D. B.; Guha, R. V. Building large knowledge-based systems : representation and inference in the Cyc project. Reading (MA) : Addison Wesley, 1990.

Liyang, Yu. A Developer's Guide to the Semantic Web. Heidelberg ; New York : Springer, 2011.

Maštrović, Mikica. Steve istraživački projekt u službi korisnika ili kako nam druga iskustva mogu pomoći u razumijevanju onoga što radimo. // Dokumentacija i korisnici, Zadar, 28.-30. listopada 2015.: knjiga sažetaka /2. skup muzejskih dokumentarista Hrvatske. <s.l.>: Hrvatsko muzejsko društvo, 2015.

Mate Ujević (1901.-1967.): utemeljitelj i glavni urednik Hrvatske enciklopedije. Dostupno na URL: www.croatianhistory.net/etf/ujevic.html (26.05.2016).

McGuinness, Deborah L.; Van Harmelen, Frank. OWL Web Ontology Language Overview. Dostupno na URL: <http://www.w3.org/TR/owl-features/> (16.09.2015.).

Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections / Kobilarov, Georgi...[et al.]. Dostupno na URL:
<http://www.georgikobilarov.com/publications/2009/eswc2009-bbc-dbpedia.pdf>
(18.09.2015.).

Melo, Gerard De; Weikum, Gerhard. Towards a Universal Wordnet by Learning from Combined Evidence. // In CIKM, 2009, 0-9. Dostupno na URL:
<http://dl.acm.org/citation.cfm?doid=1645953.1646020> (27.03.2015.).

Michel, P.; Herren, M.. Unvorgreifliche Gedanken zu einer Theorie des Enzyklopadischen –

Enzyklopadien als Indikatoren für Veränderungen bei der Organisation und der gesellschaftlichen Bedeutung von Wissen // Allgemeinwissen und Gesellschaft, Akten des internationalen Kongresses über Wissenstransfer und enzyklopadische Ordnungssysteme, Prangins, 2003. Dostupno na URL: <http://www.enzyklopaedie.ch/kongress/aufsaeetze/vorwort.pdf> (5.01.2015.).

Mihaljević, Milica. Terminološki priručnik. Zagreb: Hrvatska sveučilišna naklada, 1998.

Miller, G. A. Dictionaries in the Mind. // Language and Cognitive Processes 1, 3(1986), 171-185.

Miller, G. A. Introduction to wordnet: An on-line lexical database. // International Journal of Lexicography 3, 4 (1990), Pages: 235-244.

Miller, G. A. Nouns in WordNet. // Wordnet: An electronic lexical database./ ed. by Christiane Fellbaum. Cambridge, Mass : MIT Press, 1998.

MindPapers. Dostupno na URL: <http://consc.net/biblio.html> (15.06.2015.).

Missikoff, M.; Scholl, M.: An algorithm for insertion into a lattice: application to type classification. // Proceeding International Conference, FODO 1989 on Foundations of Data Organization and Algorithms. New York: Springer -Verlag, 1989.

MULTI-Source Entity recognition – an Information Extraction System for Diverse Text Types / Maynard, D. ... [et al.] Technical report CS—02—03, University of Sheffield, Department of Computer Science, 2003. Dostupno na URL: <http://gate.ac.uk/gate/doc/papers.html> (30.10.2015.).

Nandigam, Jagadeesh; Gudivada, Venkat N; Kalavala, Mrunalini. Semantic Web services. // J. Comput. Small Coll. 21, 1(2005), 50-63.

Neches, Robert...et al. Enabling Technology For Knowledge Sharing. AI Magazine 12 , 3 (1991), 36 – 56.

Nicholas, Charles K; Dahlberg Randall. Spotting Topics with the Singular Value Decomposition // PODDP '98 Proceedings of the 4th International Workshop on Principles of Digital Document. London: Springer-Verlag, 1998.

Nielsen, Jakob. Usability engineering. Boston : Academic Press, 1993.

Niepert, Mathias; Buckner, Cameron; Allen; Colin. A dynamic ontology for a dynamic reference work. // *Proceedings of JCDL 2007*, 2007. Dostupno na URL: <https://inpho.cogs.indiana.edu/papers/2007NiepertBucknerAllen.pdf> (16.06.2015.).

Niepert, Mathias; Buckner, Cameron; Allen, Colin. Answer set programming on expert feedback to populate and extend dynamic ontologies. // *Proceedings of 21st FLAIRS Conference*, 2008. Dostupno na URL: <http://www.aaai.org/Papers/FLAIRS/2008/FLAIRS08-116.pdf> (16.06.2015.).

Nikolić-Hoyt, A. Konceptualna leksikografija: prema tezaurusu hrvatskoga jezika. Zagreb: Hrvatska sveučilišna naklada, 2004.

Niles, Ian; Allan, Terry. The MILO : A General-purpose, Mid-level Ontology. // *Proceedings of the International Conference on Information and Knowledge Engineering*, Las Vegas: CSREA Press, 2004.

Niles, Ian; Pease, Adam. Origins of the IEEE Standard Upper Ontology. // *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*. Seattle, 2001, str. 37- 42. Dostupno na URL: <http://www.adampease.org/OP/pubs/IJCAI2001.pdf> (27.8.2015.).

Niles, Ian; Pease, Adam. Towards a Standard Upper Ontology // *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems 2(2001)*. Dostupno na URL: <http://projects.tecknowledge.com/HPKB/Publications/FOIS.pdf> (27.8.2015.).

Noy, Natalya F.; McGuinness, Deborah L. *Ontology Development 101: A Guide to Creating Your First Ontology* // Citeseer; 2001. Dostupno na URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.5085> (17.09.2015.)

Obitko, M. *Ontologies - Description and Applications*. Report No: GL 126/01. Gerstner Laboratory for Intelligent Decision Making and Control. Series of Research Reports. 2001. Dostupno na URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.419&rep=rep1&type=pdf> (1.2.2015.).

- Obitko, Marek. Semantic Web Architecture. // Ontologies and Semantic Web, 2007.
Dostupno na URL:
<http://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html> (10.09.2015.).
- Obitko, M.; Smid, J.; Snašel, V. Using Easel Types for Document Management and Retrieval. PSMP3 Workshop within AIA2004. IASTED/ActaPress. 2004. Dostupno na URL:
<http://www.actapress.com/Abstract.aspx?paperId=15896> (01.02.2015.).
- Obitko, Marek; Snašel, Vaclav; Smid, Jan. Ontology Design with Formal Analysis. Dostupno na URL: <http://ceur-ws.org/Vol-110/paper12.pdf> (02.02.2015.).
- Ontologija // Hrvatska enciklopedija, mrežno izdanje / glavni urednik Slaven Ravlić. Zagreb: Leksikografski zavod Miroslav Krleža, 2013-2015. Dostupno na URL:
<http://www.enciklopedija.hr/Natuknica.aspx?ID=45185> (18.11.2015.).
- Ontotext. Dostupno na URL: <http://www.ontotext.com/kim> (10.03.2015.).
- Ontotext: KIM platform. Dostupno na URL:
<http://ontotext.com/semantic-solutions/dynamic-semantic-publishing-platform/>
(1.04.2015.).
- OpenCyc. Dostupno na URL: <http://www.opencyc.org/doc/toc> (30.10.2015.).
- O Leksikografskom zavodu. Dostupno na URL:
<http://www.lzmk.hr/hr/leksikografski-zavod/o-leksikografskom-zavodu> (31.05.2016.).
- O Leksikografskom zavodu. Misija i vizija. Dostupno na URL:
<http://www.lzmk.hr/hr/leksikografski-zavod/o-leksikografskom-zavodu/misija-i-vizija>
(31.05.2016.).
- O portalu. Dostupno na URL: http://enciklopedija.lzmk.hr/o_portalu.aspx (03.6.2016.).
- Organizacija natuknice. // Hrvatska enciklopedija: 1 : A – Bd / < glavni urednik Dalibor Brozović>. Zagreb : Leksikografski zavod Miroslav Krleža, 1999 - .
- O Proleksis enciklopediji. Dostupno na URL: <http://proleksis.lzmk.hr/naslovnica/>
(3.06.2016.).

- Palmer, [Sean B.](#) The Semantic Web: An Introduction. Dostupno na URL: <http://infomesh.net/2001/swintro/> (16.09.2015.).
- Pease, Adam R.; Carrico, Todd M. JTF-ATD Core Plan Representation: A Progress Report // Proceedings of the AAAI-97 Spring Symposium on Ontological Engineering. 1997. Dostupno na URL: <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-012.pdf> (27.11.2015.).
- Pease, Adam; Niles, Ian; Li, John. The suggested upper merged ontology: A large ontology for the semantic web and its applications. Dostupno na URL: <http://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf> (27.11.2015.).
- Philpapers. Dostupno na URL: <http://consc.net/taxonomy.html>). (15.06.2015.).
- PhilSci Archive. Dostupno na URL: <http://philsci-archive.pitt.edu/view/subjects/> (15.06.2015.).
- Pianta, Emanuele; Bentivogli, Luisa; Girardi, Christian. MultiWordNet: developing an aligned multilingual database. // Proceedings of the First International Conference on Global WordNet, 2002. Dostupno na URL: <http://multiwordnet.fbk.eu/paper/MWN-India-published.pdf> (15.06.2015.).
- Pollock, Jeffrey T. Semantic Web for Dummies. Indianapolis: Hoboken, N.J. : Wiley, 2009.
- Portal znanja. Dostupno na URL: http://enciklopedija.lzmk.hr/o_portalu.aspx (03.6.2016.).
- Predgovor mrežnom izdanju Hrvatske enciklopedije. Dostupno na URL: <http://www.enciklopedija.hr/predgovor.aspx> (31.05.2016.).
- Garcia, A. ...et al. Preliminary Results in Tag Disambiguation using DBpedia. Dostupno na URL: <http://eprints.ecs.soton.ac.uk/17792/1/CKCaR09-final.pdf> (18.09.2015.).
- Proleksis enciklopedija. Dostupno na URL: <http://proleksis.lzmk.hr> (31.05.2016.).
- Protégé. Dostupno na URL: <http://protege.stanford.edu/about.php> (3.06.2016.).

- Protégé ontology library. Dostupno na URL:
http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library (15.06.2015.).
- PROTON (PROTo ONtology). Dostupno na URL: <http://ontotext.com/proton-ontology/>
(26.10.2015.).
- Računalnolingvistički modeli i jezične tehnologije za hrvatski jezik. Dostupno na URL:
<http://rmjt.ffzg.hr> (3.06.2016).
- Ramachandran, Deepak; Reagan, Pace; Goolsbey, Keith. First-orderized ResearchCyc :
expressivity and efficiency in a common-sense ontology. // AAI Workshop on
Contexts and Ontologies: Theory, Practice and Applications Dostupno na:
<http://www.cyc.com/doc/white-papers/folification.pdf> (30.10.2015.).
- Ravlić, Slaven. Prosvjetiteljstvo i enciklopedija: dva prosvjetiteljstva i dvije enciklopedije //
Studia lexicographica, 7 (2013), 2(13), 35-55.
- RDF 1.1 Primer. / ed by Frank Manola, Eric Miller, Brian McBride. World Wide Web
Consortium, February 2014. Dostupno na URL: <http://www.w3.org/TR/rdf-primer/>
(10.9.2015.).
- Reed, Stephen L; Lenat, Douglas B. Mapping ontologies into Cyc. // AAI 2002 Conference
workshop on ontologies for the semantic web (2002). Dostupno na:
http://www.cyc.com/doc/white_papers/mapping-ontologes-into-cyc_v31.pdf
(30.10.2015.).
- Resource Description Framework Model and Syntax Specification./ ed. by Ora Lassila, Ralph
R. Swick. World Wide Web Consortium, 1999. Dostupno na URL:
<http://www.w3.org/TR/PR-rdf-syntax/> (10.9.2015.)
- Russell, Stuart; Norvig, Peter. Artificial Intelligence: A Modern Approach // Artificial
Intelligence 1-2(1995), 932.
- Salton, Gerard. Automatic text processing : the transformation, analysis, and retrieval of
information by computer. Reading, Mass. : Addison-Wesley, 1989.
- Salton, Gerard; Buckley, Chris. Improving retrieval performance by relevance feedback. //
Journal of the American Society for Information Science, 41, 4(1990), 288–297.

- Sanger, Larry. Who says we know?: On the new politics of knowledge. The Edge, 208. Retrieved March 22, 2008. Dostupno na URL: http://www.edge.org/3rd_culture/sanger07/sanger07_index.html (15.06.2015.).
- Sažetak 3. sjednice Vijeća projekta „Hrvatska kulturna baština“, održane 09. Siječnja 2013. godine. / Ministarstvo kulture Republike Hrvatske. Dostupno na URL: <http://www.min-kulture.hr/default.aspx?id=9946> (6.06.2016.).
- Schema.org. Dostupno na URL: <http://schema.org/> (26.10.2015.).
- Schmitt, Ingo; Saake, Gunter: Merging inheritance hierarchies for database integration. // 3rd IFCIS Intl. Conf. on Cooperative Information Systems: proceedings, August 20-22, 1998, New York City, New York, USA.
- Scholarpedia. Dostupno na URL: <http://www.scholarpedia.org/> (15.06.2015.).
- Schreiber, G.; Wielinga, B.; Jansweijer W. The KACTUS View on the ‘O’ World. // In Proceedings of the National Dutch AI Conference. NAIC’95. 1995. Dostupno na URL: <http://www.math.vu.nl/~guus/papers/Schreiber95a.pdf> (8.2.2016.).
- Searching for common sense : populating Cyc from the web. / Matuszek, C. et al. // Proceedings of the twentieth national conference on artificial intelligence 20(2005). Dostupno na: <http://www.cyc.com/doc/white~paperdAA4IO51MatuszekC.pdf> (27.09.2015.).
- Segaran, Toby; Evans, Colin; Taylor, Jamie. Programming the Semantic Web. Beijing ; Sebastopol, CA : O’Reilly, 2009.
- Kiryakov, Atanas...et al. Semantic Annotation, Indexing, and Retrieval. // Journal of Web Semantics, Vol 2, No 1 (2004). Dostupno na URL: <http://www.websemanticsjournal.org/index.php/ps/article/view/53> (10.03.2015.).
- Semantic Web. // In World Wide Web Consortium. Dostupno na URL: <http://www.w3.org/standards/semanticweb/> (8.9.2015.)
- Shadbolt, Nigel; Hall, Wendy; Berners-Lee, Tim. The Semantic Web Revisited. // J. IEEE Intelligent Systems 21, 3(2006), 96–101.
- Simons, Peter M. Parts: a Study in Ontology. Oxford : Clarendon Press ; New York : Oxford

University Press, 1987.

SKOS: Simple Knowledge Organization System. Dostupno na URL:
<http://www.w3.org/2004/02/skos/> (18.09.2015.).

Smith, Barry. Fiat Objects // Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, Proceedings of the 11th European Conference on Artificial Intelligence on Artificial Intelligence, Amsterdam, 8 August 1994, Amsterdam: European Coordinating Committee for Artificial Intelligence / ed. by N. Guarino, L. Vieu and S. Pribbenow. 1784(1994), 15-23. Dostupno na URL:
<http://link.springer.com/article/10.1023%2FA%3A1017948522031> (8.9.2015.).

Smith, Michael K.; Welty, Chris; McGuinness, Deborah L. OWL Web Ontology Language Guide. Dostupno na URL: http://www.w3.org/TR/owl-guide/#owl_Class (16.09.2015.).

Sowa, John F. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA, 2000. Dostupno na URL: <http://www.jfsowa.com/krbook/> (27.8.2015.).

Sowa, J. The challenge of knowledge soup. Dostupno na:
<http://www.jfsowa.com/pubs/challenge.pdf> (27.09.2015.).

Sparck Jones, Karen. A statistical interpretation of term specificity and its application in retrieval. // Journal of Documentation 28, 1(1972), 11-21.

Special Issue on BalkaNet. / ed. by Dan Tufiş. // Romanian Journal on Science and Technology of Information, 7 (3-4): 9-44, 2004.

Standard Upper Ontology Working Group (SUO WG). Home Page. Dostupno na URL:
<http://web.archive.org/web/20071010030010/http://suo.ieee.org/index.html> (3.06.2016.).

Stanford Encyclopedia of Philosophy. Dostupno na URL: <http://plato.stanford.edu/> (27.03.2015.).

Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard. Yago: A Core of Semantic Knowledge Unifying Wordnet and Wikipedia. Dostupno na URL: .
<http://www2007.org/papers/paper391.pdf> (17.06.2015.).

Suchanek, Fabian M.; Kasneci, Gjergji; Weikum, Gerhard. Yago: A Large Ontology from Wikipedia and WordNet. // Web Semantics 6, 3(2008), 203-217. Dostupno na URL: . <http://dx.doi.org/10.1016/j.websem.2008.06.001> (17.06.2015.).

Suggested Upper Merged Ontology (SUMO). Dostupno na URL: www.ontologyportal.org (27.8.2015.).

Svenonius, Elaine. Intelektualne osnove organizacije informacija. Lokve, 2005.

Swartout, Bill...et al. Toward Distributed Use of Large-Scale Ontologies. // AAAI Technical Report SS-97-06. Dostupno na URL: <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-018.pdf> (8.2.2016.).

Šojat, Krešimir. Struktura glagolskog dijela hrvatskog WordNeta // Filologija 59(2012), 154-156.

Švab, Mladen. Mate Ujević: utemeljitelj suvremene hrvatske enciklopedistike / Radovi Leksikografskog zavoda „Miroslav Krleža“, 2(1992), 9-73.

Tang, Bin...et al. A Comparative Study of Dimension Reduction Techniques for Document Clustering: technical report. Dostupno na URL: https://www.cs.dal.ca/sites/default/files/technical_reports/CS-2004-14.pdf (17.02.2016.).

Tata, Sandeep; Patel, Jignesh M. Estimating the selectivity of *tf-idf* based cosine similarity predicates // ACM SIGMOD Record 36, 2007, 75-80.

Terziev, Ivan; Kiryakov, Atanas; Manov; Dimitar. D1.8.1 Base upper-level ontology (BULO) Guidance // SEKT:Semantically Enabled Knowledge Technologies. Dostupno na URL: [http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20\(BULO\)%20Guidance.pdf](http://www.sekt-project.com/rd/deliverables/wp01/sekt-d-1-8-1-Base%20upper-level%20ontology%20(BULO)%20Guidance.pdf) (18.09.2015.).

The KACTUS Booklet version 1.0. Esprit Project 8145. September, 1996. Dostupno na <http://www.citeulike.org/user/eduardocruz/article/3926072> (8.2.2016.).

The MEANING Multilingual Central Respository. / J. Atserias...[et al.] // In Proceedings of the Second International Global WordNet Conference (GWC'04). Dostupno na URL: <http://www.fi.muni.cz/gwc2004/proc/106.pdf> (27.8.2015.).

The Mondial Database. Dostupno na URL: <http://datahub.io/dataset/mondial> (27.8.2015.).

The OBO Foundry. Dostupno na URL: <http://www.obofoundry.org/> (27.8.2015.).

Schlenoff, Craig ...et al. The Process Specification Language (PSL): Overview and Version 1.0 Specification / NISTIR 6459, National Institute of Standards and Technology, Gaithersburg, MD, 2000. Dostupno na URL: <http://www.mel.nist.gov/msidlibrary/doc/nistir6459.pdf> (27.11.2015.).

Touretzky, David S. The Mathematics of Inheritance Systems. London : Pitman ; Los Altos, Calif. : Morgan Kaufmann Publishers, 1986.

Tumačenja značenja i sadržaja definicije // Hrvatska enciklopedija: 1 : A – Bd / < glavni urednik Dalibor Brozović>. Zagreb : Leksikografski zavod Miroslav Krleža, 1999-

UMBEL 1.00. Dostupno na URL: <http://umbel.org/resources/news/finally-umbel-v-100/> (26.10.2015.).

UMBEL Specification. Dostupno na URL: http://techwiki.umbel.org/index.php/UMBEL_Specification (26.10.2015.).

Understanding metadata / by National Information Standards Organization. Bethesda, MD: NISO Press,2004. Dostupno na URL: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf> (10.02.2014.).

Upper Mapping and Binding Exchange Layer (UMBEL). Dostupno na URL: <http://www.umbel.org/> (27.11.2015.).

Uredba o Leksikografskom zavodu „Miroslav Krleža“. Dostupno na URL: http://narodne-novine.nn.hr/clanci/sluzbeni/1991_02_6_153.html (6.06.2016.).

URIs, URLs, and URNs: Clarifications and Recommendations: Report from the joint W3C/IETF URI Planning Interest Group // World Wide Web Consortium, 21 September 2001. Dostupno na URL: <http://www.w3.org/TR/uri-clarification/> (10.9.2015.).

Uschold, Mike. Building Ontologies: Towards a Unified Methodology. Edinburgh : Artificial Intelligence Applications Institute, University of Edinburgh, 1996. Dostupno na URL:

<http://www.aiai.ed.ac.uk/publications/documents/1996/96-es96-unified-method.pdf>
(29.01.2016.).

Uschold, M.; Grüninger, M. Ontologies: Principles, Methods and Applications. Knowledge Sharing and Review 11, 2 (1996). Dostupno na URL:
<http://www.upv.es/sma/teoria/sma/onto/96-ker-intro-ontologies.pdf> (29.01.2016.).

Uschold, Mike; King, Martin. Towards a Methodology for Building Ontologies. Methodology 80, 1995, 275-280.

Uvjeti za funkcionalnost predmetnih autoriziranih podataka : konceptualni model / urednice Marcia Lei Zeng, Maja Žumer i Athena Salaba ; IFLA-ina Radna skupina za Uvjete za funkcionalnost predmetnih preglednih zapisa (FRSAR) ; odobrio Stalni odbor IFLA-ine Sekcije za klasifikaciju i indeksiranje u lipnju 2010. ; [s engleskog preveo Predrag Perožić].

Van Rijsbergen, C. J. A theoretical basis for the use of co-occurrence data in information retrieval. Journal of Documentation 33, 2(1977), 106-119.

Vossen, Piek. EuroWordNet: A Multilingual Database with Lexical Semantic Networks. // Computational Linguistics 25, 4(1998), 427-430. Dostupno na URL:
<https://www.aclweb.org/anthology/J/J99/J99-4008.pdf> (3.12.2015.).

Wache H...et al. Ontology-Based Integration of Information: A Survey of Existing Approaches. // Proc. International Joint Conference on Artificial Intelligence; Workshop: Ontologies and Information Sharing Workshop: Ontologies and Information Sharing, Seattle, WA, 108-117.

Waiyamai, Kitsana; Taouil, Rafik; Lakhal, Lotfi. Towards an object database approach for managing concept lattices. // Conceptual modeling - ER '97 : 16th International Conference on Conceptual Modeling, Los Angeles, California, USA, November 3-5, 1997 : proceedings. Berlin ; New York : Springer, 1997.

Web services Architecture / ed. by David Booth ... et al. // World Wide Web Consortium, 11 February 2004. Dostupno na URL: <http://www.w3.org/TR/ws-arch/#XML-infoset>
(10.9.2015.).

What is Unicode? Dostupno na URL: <http://www.unicode.org/standard/WhatIsUnicode.html>
(10.09.2015.)

What is WordNet? Dostupno na URL: <https://wordnet.princeton.edu/wordnet/> (30.10.2015.).

Wikipedia : The Free Encyclopedia. Dostupno na URL:http://en.wikipedia.org/wiki/Main_Page (17.06.2015.).

WikiProject Geographical coordinates. Dostupno na URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Geographical_coordinates (17.06.2015.).

Wille, Rudolf. Formal Concept Analysis as Applied Lattice Theory // Concept Lattices and Their Applications / Ben Yahia, Sadok; Mephu Nguifo, Engelbert; Belohlavek, Radim (Eds.). Berlin ; New York : Springer, 2008.

Wille, Rudolf. Restructuring lattice theory: an approach based on hierarchies of concepts // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5548 LNAI, 2009, Str. 314-339.

Wolff, Karl Erich. A first course in Formal Concept Analysis // Advances in Statistical Software 4, 1993, 429-438.

Wordnet : A lexical database for English. Dostupno na URL: <http://wordnet.princeton.edu/> (17.06.2015.).

WordNet Domains. Dostupno na URL: <http://wndomains.fbk.eu/> (17.06.2015.).

World Wide Web Consortium (W3C). Dostupno na URL: <http://www.w3.org/> (16.09.2015.).

W3C Geospatial Vocabulary. Dostupno na URL: <http://www.w3.org/2005/Incubator/geo/XGR-geo/> (18.09.2015.).

W3C. OWL 2 Web ontology language document overview (second edition). 12 December 2012. Dostupno na URL: <http://www.w3.org/TR/owl2-overview> (26.10.2015.).

W3C. OWL Web Ontology Language Guide. 10 February 2004. Dostupno na URL: <http://www.w3.org/TR/owl-guide/> (26.10.2015.).

W3C. OWL 2 Web Ontology Language: Quick Reference Guide (Second Edition). 11 December 2012. Dostupno na URL: <http://www.w3.org/TR/owl2-quick-reference/>

(26.10.2015.).

W3C. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. 02 December 2008. Dostupno na URL: <http://www.w3.org/TR/2008/WD-owl2-syntax-20081202/> (26.10.2015.).

W3C. RDF Schema 1.1. 25 February 2014 Dostupno na URL: <http://www.w3.org/TR/rdf-schema> (26.10.2015.).

W3C Semantic Web Activity. Dostupno na URL: <http://www.w3.org/2001/sw/> (9.9.2015)W3C.

SKOS Core Guide. Working Draft 2 November 2005. Dostupno na URL: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/> (26.10.2015.).

W3C. SKOS Mapping Vocabulary Specification. Dostupno na URL: <http://www.w3.org/2004/02/skos/mapping/spec/2004-11-11.html> (26.11.2015.).

W3C. Vocabularies. Dostupno na: [URL:http://www.w3.org/standards/semanticweb/ontology](http://www.w3.org/standards/semanticweb/ontology) (26.10.2015.).

YAGO2. Dostupno na URL: <http://meta-guide.com/data-processing/knowledge-synthesis/yago2> (04.01.2016.).

YAGO: A High-Quality Knowledge Base. Dostupno na URL: <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/> (17.06.2015.).

Hoffart, Johannes ...et al. YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia. Dostupno na URL: <http://resources.mpi-inf.mpg.de/yago-naga/yago/publications/aij.pdf> (27.03.2015.)

Hoffart, Johannes ...et al. YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. Dostupno na URL: <http://suchanek.name/work/publications/www2011demo.pdf> (27.03.2015.)

Yeo, Richard. Encyclopaedic Vision: Scientific Dictionaries and Enlightenment Culture. Cambridge: Cambridge University Press, 2001.

Zakon o Leksikografskom zavodu „Miroslav Kreleža“. Dostupno na URL:
http://narodne-novine.nn.hr/clanci/sluzbeni/2003_06_96_1223.html (31.05.2016.).

Zgusta, L. Priručnik leksikografije. Sarajevo: Svjetlost, 1991.

Žubrnić, Darko. Diskretna matematika. 2. izd. Zagreb: Element, 2002.

12 POPIS TABLICA

Tabela 1: Jednostavne trojke	24
Tabela 2: Klase RDF Scheme	35
Tabela 3: Svojstva i atributi RDFS Sheme.....	39
Tabela 4: Podatkovna svojstva XML Scheme.....	49
Tabela 5: Podržani tipovi podataka u OWL 2 jeziku	54
Tabela 6: Pregled općih enciklopedijskih izdanja LZMK.....	150
Tabela 7: Prikaz strukovnih enciklopedijskih izdanja LZMK.....	153
Tabela 8: Opis vrste i oblika podataka biografskih enciklopedijskih članaka hrvatske književnosti..	169
Tabela 9: Prikaz dijela rječnika termina ontologije književnosti	182
Tabela 10: Hijerarhijski prikaz modula Književne vrste i rodovi	190
Tabela 11: Periodizacija književnosti.....	195
Tabela 12: Hijerarhijski prikaz modula Dokument	197
Tabela 13: Prikaz ontologijskih modula.....	257
Tabela 14: Formalni pojmovi	295
Tabela 15: Formalni pojmovi sadržani u reduciranoj rešetci	297
Tabela 16: Rezultati frekvencije termina za „informativni dnevnik"	315
Tabela 17: Rezultati frekvencije termina za “tako lirski dnevnik”.....	316
Tabela 18: Rezultati inverzne frekvencije dokumenta za termine koji se pojavljuju u izrazima „informativni dnevnik“ i “tako lirski dnevnik”	317
Tabela 19: Izračun TF i IDF koji su potrebni za izračun TF-IDF uzorak upita	319
Tabela 20: Rezultati TF-IDF mjere za pojedini termin	319
Tabela 21: Zbroj TF-IDF vrijednosti za višeterminske uzorke upita	319
Tabela 22: Ekstrahirani termini	334
Tabela 23: Prisutnost određenih riječi grupiranih po pripadnosti određenom članku (istaknuto znakom +) te pripadnost riječi i članaka određenom klasteru.....	339
Tabela 24: Vrijednosti pojedinih indeksnih riječi i dokumenata.....	340
Tabela 25: Razvrstavanje autora tj. natuknica članaka (dokumenata) kao instanci pojmova (klasa) i grupiranje riječi koje opisuju pojedine pojmove	343

13 POPIS SLIKA

Slika 1: Trenutni WWW (lijevo) i semantički web (desno).....	13
Slika 2: Arhitektura semantičkog weba (preuzeto iz: (Antoniou, G.; Van Harmelen, F., 2008.)).....	15
Slika 3: RDF trojka.....	24
Slika 4: Grafički prikaz trojki iz Tablice 1.....	24
Slika 5: SKOS temeljna dokumentacijska svojstva (preuzeto iz: (SKOS Core Guide, 2005.)).....	57
Slika 6: Hijerarhija semantičkih svojstava (preuzeto iz: (SKOS Core Guide, 2005.)).....	58
Slika 7: PROTON (PROTo ONtology) moduli (preuzeto iz: (Terziev, I.; Kiryakov, A.; Manov; D., 2005.)).....	87
Slika 8: Infookvir iz Wikipedijinog članka Split (preuzeto iz: (Sarić, I.; Essert, M., 2011.)).....	93
Slika 9: RDF tripleti ekstrahirani iz Infookvira Split (preuzeto iz: (Sarić, I.; Essert, M., 2011.)).....	94
Slika 10: Glavni prozor Protege-a.....	114
Slika 11: Stvaranje nove klase u Protégé-u.....	115
Slika 12: Prikaz dijela taksonomije klasa iz ontologije književnosti kreiranih u kartici Classes.....	116
Slika 13: Prikaz stvorenih individua u Protégé-u.....	117
Slika 14: Dodavanje individue u odgovarajuću klasu u Protégé-u.....	118
Slika 15: Prikaz podatkovnih odnosa u Protégé-u.....	120
Slika 16: Prikaz stvorenih objektnih odnosa u Protégé-u.....	120
Slika 17: Indirektno svrstavanje resursa u klase.....	121
Slika 18: Definiranje dodatnih karakteristika odnosa u Protégé-u.....	122
Slika 19: Definiranje inverznog odnosa.....	123
Slika 20: Definiranje simetričnog odnosa.....	124
Slika 21: Definiranje tranzitivnog svojstva.....	125
Slika 22: Dodavanje opisnog odnosa naziva (rdfs:label) u Protégé-u.....	126
Slika 23: Dodavanje opisnog odnosa komentar (rdfs:comment) u Protégé-u.....	127
Slika 24: Pripadnost individue višestrukim klasama.....	127
Slika 25: Povezivanje individua u Protégé-u.....	128
Slika 26: Dodavanje doslovnih vrijednosti individui u Protégé-u.....	129
Slika 27: Vizualizacija ontologije pomoću OntoGraf-a.....	129
Slika 28: Agent modul.....	187
Slika 29: Podklase klase Person.....	187
Slika 30: Podklase klase Umjetnik, Zanimanja i Znanstvenik.....	188
Slika 31: Podklase ontologijske klase Književnici.....	189
Slika 32: Prikaz hijerarhije odnosa Svojstva Osoba i OsobaOsobniPodaci.....	200
Slika 33: Prikaz hijerarhije ontologijskih odnosa OsobaDjelatnost.....	202
Slika 34: Prikaz hijerarhije ontologijskih odnosa OsobaOsoba.....	207
Slika 35: Prikaz hijerarhije ontologijskih odnosa OsobaStvaralaštvo.....	212
Slika 36: Hijerarhija SvojstvaAutora.....	217
Slika 37: Podsvojstva DjeloOsoba.....	225
Slika 38: Prikaz podsvojstva Opis djela.....	227
Slika 39: Primjer zaključivanja i klasificiranja entiteta Poezije nakon primjene reasoner-a Hermit 1.3.8. u Protege-u.....	235
Slika 40: primjer klasificiranja knjiž. vrsta Drama nakon primjene reasoner-a Hermit 1.3.8.....	236

Slika 41: Podatkovni odnosi ontologije hrvatskih književnika	237
Slika 42: Opisna svojstva	242
Slika 43: Podhijerarhija svojstava životopis.....	243
Slika 44: Prikaz opisnih svojstava	249
Slika 45: Prikaz objektivnih i podatkovnih relacija individue Antun Gustav Matoš	256
Slika 46: Prikaz opisnih relacija individue Antun Gustav Matoš.....	257
Slika 47: Svojstvo jeAutor povezuje individuu A. G. Matoš sa svim njegovim djelima i prikazuje njegovu pripadnost pojedinim klasama.	267
Slika 48: Odgovor na složeni upit „Koje je pjesme domoljubne tematike napisao Antun Gustav Matoš?“	268
Slika 49: Odgovor na složeni upit „Koje je djelo Antuna Gustava Matoša objavljeno u Vijencu 1892?“	269
Slika 50: Odgovor na složeni upit „Koji je naslov fikcijske proze Antun Gustav Matoš objavio 1899?“	269
Slika 51: Odgovor na složeni upit „Koji je impresionističku novelu objavio Antun Gustav Matoš?“	270
Slika 52: Odgovor na složeni upit „U kojim je sve publikacijama objavljivao A. G. Matoš?“	270
Slika 53: Odgovor na složeni upit „Na koje je djelo travestija djelo Smrt Babe Čengiđkinje?“	271
Slika 54: Odgovor na složeni upit postavljen u Primjeru 7.....	271
Slika 55: Izbor potrebnih kriterija pregledavanja	272
Slika 56: Rezultati pretraživanja	273
Slika 57: Rezultati pretraživanja sa SPARQL.....	273
Slika 58: Rezultati za upit „Koji su dubrovčani pisali latinskim jezikom?“	274
Slika 59: Rezultati za upit „Koji su dubrovčani pisali latinskim jezikom?“ (nastavak).....	274
Slika 60: Rezultati za upit „Koji su dubrovčani pisali grčkim i latinskim jezikom?“	275
Slika 61: Rezultati za upit „Koji su splitski književnici dobili nagradu Vladimir Nazor?“	279
Slika 62: Rezultati za upit „Koji su splitski književnici uređivali Mogućnosti?“	280
Slika 63: Rezultati za upit „Koji su splitski književnici objavljivali u Mogućnostima?“	281
Slika 64: Rezultati za upit „Koje su drame dubrovačkih književnika objavljene u razdoblju 1830-1840?“	281
Slika 65: Rezultati za upit „Koji su dubrovački književnici članovi HAZU-a?“	282
Slika 66: Rezultati za upit „Koji su splitski književnici članovi HAZU-a?“	282
Slika 67: Rezultati za upit „Koji su splitski književnici bili predsjednici Društva hrvatskih književnika?“	283
Slika 68: Ekstrakcija podataka Python programom	292
Slika 69: Formalni kontekst prema zadanim objektima i atributima.....	293
Slika 70: Grafički prikaz formalne rešetke.....	294
Slika 71: Graf pojmovne rešetke (desna strana slike) i podaci za odabrani formalni pojam (tj. objekti i atributi na lijevoj strani slike).....	295
Slika 72: Termin-dokument matrica, njezino premještanje te termin-termin korelacijska matrica za mali primjer zbirke sa samo dva dokumenta i tri termina (preuzeto iz: (Baeza-Yates, R.; Ribeiro-Neto, B., 2011.).).....	314
Slika 73: Odnos dvaju vektora tj. dokumenata (preuzeto iz: (Baeza-Yates, R.; Ribeiro-Neto, B., 2011.).).....	322
Slika 74: Model vektorskog prostora (preuzeto iz(Ivančević, Zvonimir; Jelić, Slobodan, loc cit. (bilj. 361.))).....	323

Slika 75: Model indeksiranja skrivene semantike – LSI model (preuzeto iz(Ivančević, Zvonimir; Jelić, Slobodan, loc cit. (bilj. 361.))).	326
Slika 76: Dekompozicija matrice na singularne vrijednosti (preuzeto iz: (Deerwester...et al., 1990.).)	330
Slika 77: Aproksimacija matrice A matricom A (preuzeto iz: (Deerwester...et al., 1990.).....)	331
Slika 78: Grafički prikaz indeksnih riječi i dokumenata	340

14 ŽIVOTOPIS

Ivana Dević rođena je 1979. godine u Splitu. Diplomirala je Visoku učiteljsku školu u Splitu Sveučilišta u Splitu 2002. godine. U svibnju 2005. godine diplomirala je na Dopunskom dvogodišnjem izvanrednom studiju knjižničarstva u Zadru pri Filozofskom fakultetu u Osijeku te dobila visoku stručnu spremu i stručno zvanje diplomirani knjižničar, a u lipnju 2006. godine je položila stručni ispit za stručno zvanje diplomirani knjižničar pri Povjerenstvu za polaganje stručnih ispita u knjižničarskoj struci Hrvatskog knjižničnog vijeća. Akademске godine 2008/2009 upisuje Poslijediplomski studij informacijskih i komunikacijskih znanosti na Filozofskom fakultetu Sveučilišta u Zagrebu. Nakon položenih ispita u prosincu 2012. brani sinopsis na temu *Razvoj ontologije iz enciklopedički organiziranoga znanja*.

Od listopada 2003. do travnja 2004. u svrhu obrazovanja volonitra u visokoškolskoj knjižnici Ekonomskog fakulteta u Splitu. Od siječnja 2004. do srpnja 2007. je zaposlena kao učitelj razredne nastave u Osnovnoj školi Sućidar u Splitu. Stekla je i radno iskustvo u specijalnoj knjižnici radeći u knjižnici Instituta za javne financije u Zagrebu od kolovoza 2005. do siječnja 2006., a od siječnja 2006. je radila u školskoj knjižnici Osnovne škole Brda u Splitu. Na Odjelu za nabavu i obradu omeđenih publikacija Sveučilišne knjižnice u Splitu je zaposlena od 2011. godine.

Član je Hrvatskog knjižničarskog društva (HKD), Komisije za klasifikaciju i predmetno označivanje te Komisije za informacijske tehnologije pri HKD-u. Sudjeluje u Radnoj skupini za interoperabilnost te Radnoj skupini za opis slikovne, audio i video građe, kontroliranih rječnika i identifikatora repozitorija *Dabar* pri Sveučilišnom računalnom centru Srce. Središnji interes Ivane Dević je semantički web, razvoj ontologija i razvoj kontroliranih rječnika te njihova primjena u enciklopedici i knjižničarstvu.

Autorica je aktivno sudjelovala izlaganjem na skupovima: Seminar Arhivi, knjižnice i muzeji – AKM, Skupštini HKD-a, te stručnim edukativnim predavanjima Splitskog knjižničarskog društva. Objavila je dolje navedene radove.

15 BIBLIOGRAFIJA

Sarić, Ivana; Magdić, Antonio; Essert, Mario. ZOTERO – program otvorenog koda za upravljanje bibliografskim bilješkama. // Vjesnik bibliotekara Hrvatske. 54 (2011) , 1; 158-171.

Sarić, Ivana; Magdić, Antonio; Essert, Mario. Sheme metapodataka značajne za knjižničarstvo s primjerom implemetacije OpenURL standarda. // Vjesnik bibliotekara Hrvatske. 54 (2011) , 1; 134-157.

Sarić, Ivana; Essert, Mario. Uključivanje hrvatskog Web sadržaja u DBpediju // 14. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastructure : zbornik radova / Hasenay, Damir ; Faletar Tanacković, Sanjica (ur.). Osijek, 2011. 165-179.

Sarić, Ivana. Model povezivanja i interoperabilnosti knjižničnih zapisa na semantičkom webu. // Vjesnik bibliotekara Hrvatske, 56 (2013), 1-2; 107-122.