

Scholarly reference trees

Kristina Kocijan, krkocijan@ffzg.hr

Marko Požega, mpozega1@ffzg.hr

Dario Poljak, dpoljak1@ffzg.hr

Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb

Libellarium, IX, 2 (2016): 131 – 144.

UDK: 004.891: 001.891.32=111

DOI: <http://dx.doi.org/10.15291/libellarium.v9i2.281>

Research paper

Abstract

In this paper, we propose, explain and implement bibliometric data analysis and visualization model in a web environment. We use NLP syntactic grammars for pattern recognition of references used in scholarly publications. The extracted information is used for visualizing author egocentric data via tree like structure. The ultimate goal of this work is to use the egocentric trees for comparisons of two authors and to build networks or forests of different trees depending on the forest's attributes. We have stumbled upon many different problems ranging from exceptions in citation style structures to optimization of visualization model in order to achieve an optimal user experience. We will give a summary of our grammars' restrictions and will provide some ideas for possible future work that could improve the overall user experience. The proposed trees can function by themselves, or they can be implemented in digital repositories of libraries and different types of citation databases.

KEYWORDS: science mapping, scholarly data, network visualization, egocentric networks, NLP, information extraction, pattern recognition, reference styles, *ReferenceTree*

Introduction and related work

Lately, bibliometric analysis has become increasingly important as a way to measure and assess research impact of individuals, groups of individuals or institutions (Borić 2008, Tuđman and Pečarić 2009, Aparac and Pehar 2010, Tuđman and Pečarić 2014, Taşkin and Al 2014). If we define the field of bibliometrics as a science of statistical and mathematical methods used to quantitatively analyze citations in academic publications, as proposed in (Hebrang Grgić 2016), than our research falls partially in the area of bibliometrics. On the other side, we are using NLP to prepare the data for the bibliometric analysis. More precisely, we are using grammar driven approach for pattern

recognition and information extraction after which our data is prepared for a visualization whose purpose is to provide a new dimension to, sometimes, plain visualizations of bibliometric data.

Our inspiration comes from *ContactTrees* proposed and designed in (Sallaberry et al. 2012, 2104). Their trees were constructed to show multilevel aspects of social interactions using an egocentric approach and mainly as a help to sociologists. Later on, the idea was reused to show collaborative activities of *InfoVis* researchers (Sallaberry and Kwan-Liu 2012) and for a comparison of majors at the University of California (Fung et al. 2015). These examples showed us the power of such visualizations and we started to explore the idea (Požega et al. 2016) proposing the *ReferenceTrees* as a tool that can be used in the field of bibliometrics. We believe that the proposed egocentric approach to visualization of publication data may open a whole new category of bibliometric research questions (Pehar 2010) i.e. the category of scientific characteristics of each individual scientist.

In this paper, we propose a tree design with a purpose of presenting two types of information. The first one deals with author's publication modus and includes the number of publications a single author has published throughout his/her career, the type of each publication and the possible co-authorships. The second type deals with his/her referencing modus i.e. the number of references per publication, source type of a reference, and the number of reused references throughout his/her publications.

We also intend to show how we applied the third major and final redesign of our citation trees, as well as the two new features to the web application arsenal. In our previous work (Požega et al. 2016), we ran into a problem of presenting trees of authors with big number of citations per paper because of our tree's branches original vertical orientation. We managed to modify our visualization algorithm to reduce the mentioned anomaly and change it to a more readable format. This new structure drastically changes horizontal to vertical ratio and gives a clearer perception of data in the user's tree. After the first part of the redesign, we were not pleased with the concept of a realistic tree because it was still out of 'realistic' scope and we decided to do a completely opposite theme. We kept the old structure of our original concept of data distribution in a tree, but instead of constantly trying to simulate a real tree we agreed to a more futuristic approach. We substituted tree branches with *Printed Circuit Board (PCB)* templates and leaves with small circles, which are blank in case of paper/book/chapter in book citation and colored in case of a web citation. This futuristic design also lead us to change the design and interface of our webpage.

In the following sections, we will explain in more details all three interpretations of our tree design tackling the problems we have encountered and solutions we considered. We will also describe how we build the learning and testing corpora, and build the grammar for recognizing and annotating data needed for building the trees. Before we conclude, we will illustrate the two new upgrades: comparison of two authors and building the forest.

Our interpretations of a *tree design*

During the development phase of our system, our tree visualization diagrams have passed through three stages of design. In this section, we will show how the design changed over time using the same author for demonstrational purposes.

Our first design proposed in Požega et al. (2016) was conceived as a real life tree under the influence of Sallabery et al. (2014) and their *Contact trees*. Each main branch, representing a year of publication, is perpendicular to the trunk while the twigs placed on a branch and representing a single work by the given author, was parallel with the trunk. If the author was the sole author of a publication, the twig was positioned on the upper side of a branch. Publications written in a co-authorship were positioned on the opposite i.e. bottom side of a branch. With this layout, we were able to add another dimension to our visualization (number of papers written with or without co-authors) and still keep the design readable and neat. The references used in each paper (twig) are represented as leaves and were placed along the left and right side of a twig with the single leaf on the top in the cases where the number of references was odd. The leaves were color coded depending on the type of a source the reference came from. Thus, if the reference came from the web source it was colored in green, while the print source of a reference was indicated in orange. Hovering over a certain citation node (leaf) lights up all the other nodes on a tree where the same publication is referenced.

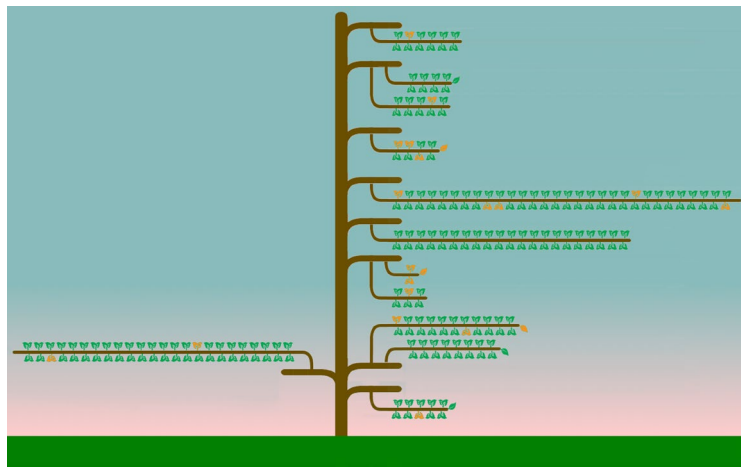
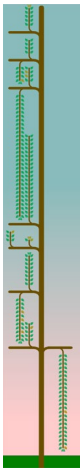


Figure 1. Design of the first version of citation tree (for author: Tatjana Aparac Jelušić)

Figure 2. Design of the second version of citation tree (for author: Tatjana Aparac Jelušić)

Although everything in the tree is modular, including the length of each branch allowing it to 'grow' depending on the number of citations the publication had, the completely straight trunk design and protrusion of branches under right

angles along the main trunk did not prove to be a very satisfactory or viable solution (Fig.1). In theory, the design looked impressive. However, it proved to be quite inconvenient when it came to real world application. The design worked well when there were no big bibliographies to visualize. At the moment when we had to place publications containing more than ten references, the trees would grow to incredulous heights and all readability was lost. When we considered the fact that there will be many authors publishing works with a plethora of references it became clear to us that it will be a problem to accommodate them all on a single branch.

This made us realize the flaws of our design and we started to look for another approach. We quickly came to the conclusion that to deal with massive amounts of references we had to rotate our twigs so they no longer grow parallel to the trunk but parallel to the main branch. This way we could handle even the larger number of references that proved to be everything but uncommon among our training data. Still, the question of arranging the twigs along the branch remained. Aligning them one after the other, with some space between, was out of the questions since our tree branches could then potentially span far out to the left or right side. This would again take us to a position where we would lose every advantage in readability of our visualization since we would not be able to fit enough relevant data on a single screen or stay big enough to be comprehensible.

Therefore, we decided to give each new twig an origin point very close to the start of a branch leaving some space between them, and stacking them one over the other (Fig. 2). The publication with the most references would occupy the twig closest to the main trunk while each following publication will be the one having the number of references smaller or equal to the previous one. Our design proved to be true when we tested it on our first authors who had a publication consisting of astounding 60 references. The sacrifice we made in the aesthetics, proved to be truly minor in regard to the precise display of data.

During the overhaul of the whole system, we were considering the ways our tree could look less fabricated. As the research progressed and we added more and more works to our tree gallery, we saw fewer and fewer ways to solve that conundrum. Then we asked ourselves why we should shy away from a fabricated look. After all, this is an algorithmic, grammar driven, computational approach. So, we decided to digitize our tree design, opting out of a hazel trunk, green and yellow leaves for a printed circuit board (PCB) inspired design (Fig. 3).

According to Katherine (2013), printed circuit board is a non-conductive material with conductive lines on it, onto which electronic components can be mounted and connected by traces to form a working circuit or assembly. Taking inspiration from the conductive lines, we still maintained the tree structure but replaced a hazel colored trunk with a dull gold trunk comprised of three slender lines, instead of the previous burly one line.

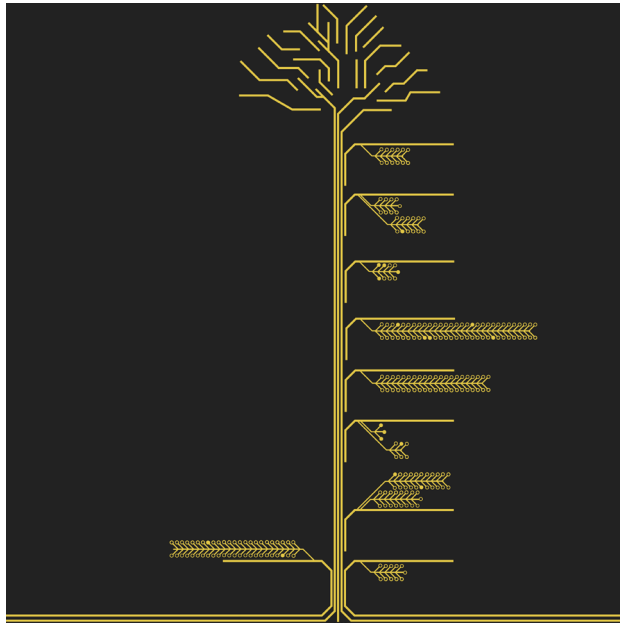


Figure 3. Design of the third version of citation tree (for author: Tatjana Aparac Jelušić)

The tree structure design with the branches is still retained, but now, the branches are more along the lines of connections on a PCB than branches with leaves. We have replaced our green and yellow leaves with simple empty and filled circles, respectively. The monochrome scheme on a simple background left us with an aesthetically pleasing design and a straightforward data representation of our previous tree designs.

Construction of corpora

Our grammar has been trained on various sets of data pertaining APA, MLA and Chicago citing styles. We shaped our grammar from the basic instructions on citing and referencing bibliographies. Data containing basic citation examples was gathered from the University of Pittsburgh Website. It contains various ways of citing works such as books, articles and websites with an example for each one. We have made our grammar able to differentiate between three major citation styles. We have further improved and revised our grammar with data concerning specific cases and exceptions. To do that we have used Purdue Online Writing Lab which was developed exactly for the purpose of eliminating lapses in referencing some rarer sources like video tapes and digital repositories. After we have revised our grammar so it completely recognizes preset examples we have tested it out with real world examples that proved to be quite more challenging due to the human factor.

We have gathered bibliographies for authors pertaining to our field of study since we could access most of their works through digital repository of our library or other public repositories. At the time of writing this paper, our database has authors Tatjana Aparac-Jelušić, Silvio Peroni, Marko Tadić, Božo Bekavac, Nikola Ljubešić and Željko Agić. Between eight and thirteen publications per each author were not only available to us but they also had a structure that we could parse according to the referencing rules (Table 1). Cumulatively, there are 1,120 reference entries in our database that we were able to visualize as circles on a PCB. Their distribution is shown in Table 2 and Table 3.

Table 1. Distribution of articles and books per author, sorted by the highest number of articles

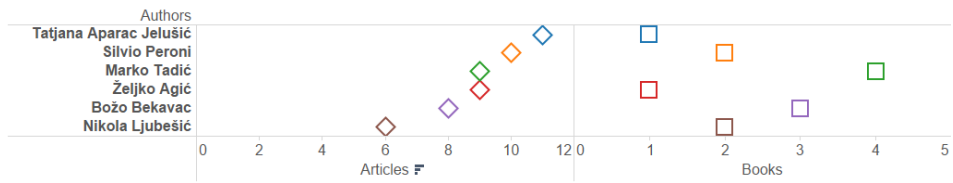


Table 2. Distribution of references per author, sorted by the smallest number of cumulative references

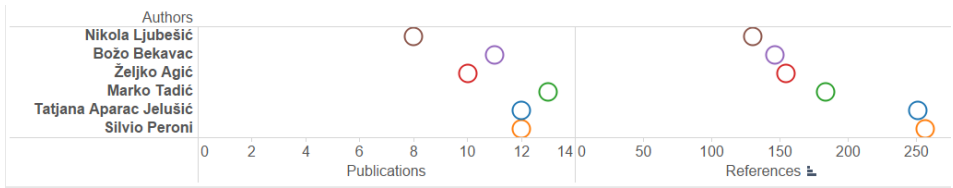


Table 3. Distribution of references per publication per author sorted by the author with the most publications



We started our search for the data at our faculty digital library's repository (www.darhiv.ffzg.unizg.hr). The repository holds the publications written by the scientific staff and students from all the departments of the Faculty of Humanities And Social Sciences at the University of Zagreb. All publications are in a digitized format and most of them are available to students and employees of the faculty and even to wider public. Publications are organized in an orderly fashion and can be browsed by year, department, subject, author, supervisor and document type. We were primarily interested in the department of Information Sciences and Computational Linguistics, where we looked for books and articles to get large enough bibliographies to make a representative visualization. We found enough publications in the repository to give us a proper start (Table 1).

Some data were not available due to publisher policies and some were unfit for use in our work because bibliographies had little to none cited books or articles but rather had cited manuals and standards. It is important to notice here that we are not considering all bibliographical entries for our tree presentation. We are primarily interested in references written by a member of the academic community. In that regard, works like videotapes, manuals, ISO standards and plethora of other cited material is, at this time, skipped in our metrics. Also, we have decided to present both books and book chapters as book entries (left side of the tree), and conference papers and journal publications as articles (right side of the tree).

To close the gap between unavailable and unfit publications, we turned to IEEE *Xplore* Digital Library at <http://ieeexplore.ieee.org/>. The IEEE *Xplore* Digital Library has articles in open access provided by the Faculty of Electrical Engineering and Computing at the University of Zagreb. Every publication has its page with separate tabs for full work, abstract, references, metrics and other plethora of useful data. We were primarily interested in the reference tab where the bibliography is listed in the structure we could easily parse. However, there was not enough articles to completely wrap up our dataset since we got a maximum of 3 bibliographies per author.

The best resource for gathering our bibliographies by far proved to be the Croatian Scientific Bibliography (CROSBI) at <http://bib.irb.hr/>. It is a combined effort by The Ministry of Education, Culture and Sport, Croatian Academic Research Network (CARNet) and the library of Ruđer Bošković Institute. It contains a database of all published work by Croatian authors. The database can be freely browsed and it is structured in a way that every publication type is listed separately. The biggest advantage of CROSBI is that authors themselves input the data relevant to their work and can name their co-authors and associates. They can also provide a link to their full work if it is in open access or to the website of their publisher, where we can usually find all data relevant to us. We have primarily used the CROSBI database for our research since it is neatly structured and we can immediately get to the pages of authors that worked together. It served as a source on data for four authors that have also been co-authors on several papers on natural language processing which allowed us to present some comparisons between similar authors (Fig. 4).

However, some works are published under strict license and cannot be accessed completely online so we were condemned to scurry through specific databases concerning the publisher of said work. Luckily, publishers like Springer have all of their publications neatly laid out and similarly to the IEEE *Xplore* Digital Library have a special tab for references with full bibliographies ready for parsing and analysis.

Thanks to all of these repositories, we have managed to acquire a dataset big enough for parsing and displaying in a neat visualization covered in the next few sections. We have also come across quite a few problems during parsing due

to the poor formatting of published works whether it was due to the technical problems with diacritics in Croatian language or poor knowledge of reference style rules. However, that is a topic that can be a research paper all on its own and we will not go in-depth about it here.

Recognizing different referencing styles

There are several referencing styles (APA, MLA, Chicago, Turabian) that are used widely by many publishers in the scientific domain and that we have decided to include in our research. We have studied the rules for referencing books, book chapters and articles in both paper or web versions and we have used that information to build the syntactic grammars in NooJ. The power of an NLP tool NooJ (Silberztein 2003) allowed us to recognize and annotate the data in the manner that was suitable for our further analysis and visualization. The main purpose of our syntactic grammars is to recognize each reference from the list of references and mark the sections inside each reference appropriately. We were mostly interested in the information about the author(s) of the reference, the title, year of publication, publication type (book or article) and publication source (web or print) that were annotated in the following manner:

```
<LIT+AUTHOR1='Brants, T.'+YEAR='2000'+TITLE='TnT – A Statistical Part-of  
Speech Tagger'+SOURCE='PRINT'+TYPE='ARTICLE'>
```

Such quintuplets are used to populate the ReferenceTrees of our authors.

In spite of the thorough instructions and explanations of these styles, many authors do not follow them and tend to use different combinations of these styles, including some novel approaches to writing references. It is hard to decipher if their (own) citation style is due to the unfamiliarity with the rules, or wrong interpretation of the rules, or with the idea that the referencing style is not that important to the publication and as such should not take too much of the time/effort to write it *'some other way'*. The purpose of this paper is not to find the answers to this dilemma (some can be found in Moed 2005, Taşkin and Al 2014), but rather to deal with the situation in a manner that will allow us to build the *ReferenceTrees* regardless of the reference style(s) used by the author.

We have taken a meticulous approach in writing our grammars from the ground up, taking every variable and rule into account. At the end of our trial period, we could brag with 100% accuracy in detecting every detail about the provided bibliography. Next to the absolutely necessary information, such as author names, year of publication and title, we have also been able to recognize the details such as publisher, place of publishing, page numbers, internet address if there was one, and the citation style used to create the reference list. However, those were only the training references that stuck rigorously to the supplied rules. Real world works turned out to be a whole different case to master.

We came across some seemingly simple problems such as disregarding punctuation rules or capitalization rules. But, we have also stumbled upon reference lists cited by styles completely unknown to us and structurally against all (referencing) logic. At their worst, some reference lists were a blend of different citation styles for different publication type where one reference would be in APA style, the other in Turabian or MLA or Chicago. There were also cases where one single reference would start as one style and finish up as another, as if it had suddenly decided to change its style.

Suddenly, quite unaware of this very liberal approach to populating the reference list, we were facing the situations where our grammars could not find any matches or only some references from the entire reference list were recognized. We considered two ways to deal with the situation: either we rewrite the references so they are in conformance with one of the referencing styles, or we add some new branches in the syntactic grammar that will recognize additional styles. We opted for the second solution and ended up with additional fourteen main branches for recognizing articles and four for recognizing books (Fig. 4).

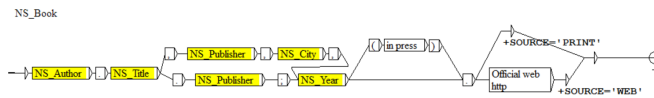


Figure 4. Section of a grammar recognizing no-style book references

However, after the detailed audit of our results, we still ended up correcting some of the references, but only where it was necessary for the results we needed. Such corrections included mostly spelling errors like separating two words written as one or adding/deleting/editing a letter in a word but only for the data needed for our quintuplets. Since our main purpose is to build the ReferenceTrees and visualize the references used throughout the author’s scientific life span, we did not find such ‘tempering with the data’ to be in foul with our research.

Reference lists and citation styles are designed with several things in mind. Their standardization of input is not a purpose in itself, nor is it mainly to maintain readability across publications. It also provides authors with a tool to help them show where they found their sources. At the same time, publishers, readers and reviewers can more easily deduce whether the source is reputable or not. In the extreme cases, references have a power to protect an author against accusations of plagiarism (Hebrang Grgić 2016). Moreover, lately, many citation databases and scientific social networks (Google Scholar, Scopus, WoS, ResearchGate, Academia) are using them with their analytical tools for evaluating authors impact (Brajnović-Milić 2014). However, surely unaware of the ways they are hurting themselves, some of the authors we came across during our research, have provided reference lists that cannot even begin to qualify for any of these functions.

New features

We processed the papers of multiple authors in NooJ and expended our database as explained in previous sections. Then, we explored an idea to add a feature of comparing public trees side by side instead of taking only a preview of a single citation tree. We have further developed this idea that led us to the final step of our project i.e. building forests dependent on the specific user's query. We explain these two upgrades at more length in the following two sections.

Two trees—a comparison

We have included the capability to display two trees at the same time next to each other for different authors. Because of the tree size and limited computer screen size (and resolution), we concluded that the optimal number of parallel trees to compare is two. The clear differentiation between the two trees is the color of the whole design of the second tree and for now, we have opted for a cobalt shade of blue. In that regard, we feel that comparing two authors has never been easier and more straightforward than now.

The comparison feature is only possible through public trees (forest) preview, which we will explain in the next sub-section. The main benefit of this feature is not only for the user to see differences in tree structure, but when hovered over citation of one author user can see if another author has used the same reference in some of his work.

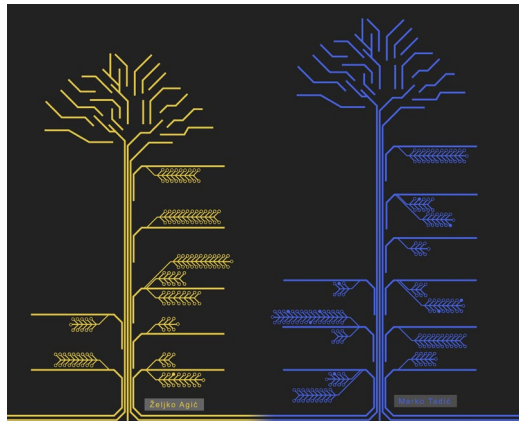


Figure 5. Comparison of two trees for authors Ž. Agić on the left and M. Tadić on the right side

To implement this feature, we use the same function we have already defined for single tree visualization. By clicking the 'compare' button and selecting an author from the database, the canvas in which the first tree is previewed horizontally extends by the number of pixels to the right needed for the second tree canvas. This procedure prepares the area where the second tree is drawn by setting the starting position as a difference between full canvas horizontal size and first tree canvas size. Since they share the same canvas, hovering over leaf (reference) of one tree automatically triggers and lights up all the leaves

that present the same reference on both trees.

A big step: from a tree to a forest

As the final step we added a global display of all the gathered trees. As stated earlier and in (Požega et al. 2016) our final goal was to connect the public trees in one global segment that we named ‘the tree forest’. The main idea of this step is to present all the trees on a single page and then cluster them by specific attributes. Thus, the authors sharing a specified attribute are grouped closer together, whereas authors that do not share this attribute are further away and can be seen as a start of their own cluster. Gradually, as we add more and more authors the trees will get smaller and smaller until they are just dots building the big picture. Still, the clusters will be easily expandable which will make them more appropriate for any further analysis.

For this task we had to drastically change the size of trees so they could fit the screen. We did that by saving small .png images in our database which by clicking on it would open a full interactive tree on a screen with the feature to compare the selected tree with another author within or out of the cluster. So far, we only managed to cluster trees (authors) by the field of study and institutions (Fig. 6) in which they do their scientific work. However, our primary goal and task is to cluster authors by the frequency of citations between the authors. The second possible clusterization would be to group authors who share the most citations of third party author. For sure, there are many possibilities and much space to improve this idea of visualizing interconnections between multiple authors.



Figure 6. Presentation of clustered authors by the institution and research field

Once the raw paper is uploaded and citation data is extracted to our database, the algorithm takes a snapshot of the tree canvas, resizes it and saves it in the database. In the public tree section, where we can find all the trees with public access, trees are shown as saved images and not as live html elements. By clicking on an image, the *drawTree()* function is called and the appropriate tree is displayed in its full size, as a DOM object. The key part in this process is a PHP function that groups images in separate clusters defined by user defined attributes. These attributes are: number of publications, number of books, number of citations, number of web citations, same research field, authors that share at least X given number of citations and most interesting and algorithmically complicated clustering method which groups the same authors which reference to each other more than X given number of times. Some of these methods are still in the development and optimization stage and we plan to add some new ones in the next stages of our project.

Conclusion and future work

In this paper, we have demonstrated the power of data visualization in the field of bibliometry. As a scientific discipline, bibliometry is revolving around analyses and display of data related to authors and their citations through various tables, charts and graphs. We have taken a more organic approach in that endeavor to close the gap between scientists and ordinary users. Our system is designed around clarity to allow everyone some range of comprehension of the data displayed: from members of academia, used to standard visualization methods, to non-academic users who could have been perplexed by the complexity and starkness of bibliometric measures. We deem our more organic approach may help shaping a new viewpoint on data visualization built with *clarity first* methodology in mind.

The benefits of the proposed scholarly reference trees can be described as twofold. The first is more egocentric and gives us novel insight into individual author's publication and referencing modus that include the number and type of publications and possible co-authorships on one side and the number of (re)used references and their source type preferences on the other. The second allows different levels of clusterization of egocentric representations.

There are still many (un)charted possibilities that this project may further take like adding new attributes for clusterization of the forest, or combining several attributes into one query. In our further research, we will be more focused on improving on our data recognition fidelity since we have stumbled upon many problems directly related to the rigidity of the rules and not analyzing context. We are also considering the ways to allow more autonomy to the user that would, for example, allow him/her to adjust the color of trees in order to get the maximum user experience from the visualization. We hope to add these optionalities in the near future.

References

- Aparac, Tatjana, and Franjo Pehar. 2010. "Information sciences in Croatia: A view from the perspective of bibliometric analysis of two leading journals." In *The Janus Faces Scholar: A Festschrift in Honour of Peter Ingwersen*, edited by Birger Larsen, Jesper W. Schneider, and Fredrik Astrom. 325-338. Copenhagen: Royal School of Library and Information Science.
- Baranovskiy, Dmitry. 2015. "Raphaël - JavaScript library." Accessed January 17, 2016. <http://raphaeljs.com>.
- Borić, Vesna. 2008. "Analiza citata radova objavljenih u časopisu *Acta stomatologica Croatica* zabilježeni u bazi podataka Web of Science." *Acta Stomatologica Croatica* 42, 2: 123-139.
- Brajenović-Milić, Bojana. 2014. "Bibliometrijski pokazatelji znanstvenog odjeka autora i časopisa." *Medicina fluminensis* 50, 4: 425-432.
- Carroll, Katherine. 2013. *Key words for electrical engineering*. Glasgow: HarperCollins Publishers.
- Chen, Chen, Rachael Dubin, and Timothy Schultz. 2014. "Science mapping." In *Encyclopedia of Information Science and Technology, Third Edition*, edited by M. Khosrow-Pour. IGI Global. DOI:10.4018/978-1-4666-5888-2.ch410. <https://doi.org/10.4018/978-1-4666-5888-2.ch410>
- Fung, Tsai Ling, and Kwan-Liu Ma. 2015. "Visual characterization of personal bibliographic data using a botanical tree design". In *Electronic Proceedings of IEEE VIS 2015 Workshop on Personal Visualization: Exploring Data in Everyday Life*. Accessed December 15, 2015. <http://www.vis4me.com/personalvis15/papers/fung.pdf>
- Fung, Tsai Ling, Jia-Kai Chou, and Kwan-Liu Ma. 2015. "Comparing characteristics of majors using egocentric botanic-trees." Accessed January 15, 2016. <http://vacommunity.org/ieeevpg/viscontest/2015/entries/6.html>.
- Hebrang Grgić, Ivana. 2016. *Časopisi i znanstvena komunikacija*. Zagreb: Naklada Ljevak.
- Moed, Henk F. 2005. *Citation analysis in research evaluation*. Dordrecht: Springer.
- Pehar, Franjo. 2010. "Od statističke bibliografije do bibliometrije. Povijest razvoja kvantitativnog prisupta istraživanju pisane riječi." *Libellarium* 3, 1: 1-28. <https://doi.org/10.15291/libellarium.v3i1.144>
- Požega, Marko, Dario Poljak, and Kristina Kocijan. 2016. "Building scholarly data forest." In *Lecture Notes in Computer Science: SAVE-SD 2016*, edited by Alejandra Gonzalez-Beltran, Francesco Osborne, Silvio Peroni. Springer International Publishing.

- Sallaberry, Arnaud, and Kwan-Liu Ma. 2012. "Visualizing InfoVis researchers with ContactTrees". Accessed January 15, 2016. <http://web.cse.ohio-state.edu/~raghu/teaching/CSE5544/Visweek2012/infovis/posters/sallaberry.pdf>
- Sallaberry, Arnaud, Yang-Chih Fu, Hwai-Chung Ho, and Kwan-Liu Ma. 2014. "Contact trees: A technique for studying personal network data." CoRR, abs/1411.0052.
- Silberstein, Max. 2003. NooJ Manual. Accessed September 15, 2010. <http://www.nooj4nlp.net> (223 pages).
- Taşkin, Zehra, and Umut Al. 2014. "Standardization problem of author affiliations in citation indexes." *Scientometrics* 98, 1:347-368. DOI: 10.1007/s11192-013-1004-x. <https://doi.org/10.1007/s11192-013-1004-x>
- Tuđman, Miroslav and Đilda Pečarić. 2014. "Ko-autorstvo najcitiranijih autora: analiza slučaja informacijskih znanosti u Hrvatskoj." In *Komunikacijski obrasci i informacijske znanosti* edited by Radovan Vrana and Đilda Pečarić. Zagreb: Zavod za informacijske studije Odsjeka za informacijske i komunikacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu, 5-16.
- Tuđman, Miroslav, and Đilda Pečarić. 2009. "Prilozi dubinskoj analizi komunikacijskih obrazaca." *Informatologija* 42, 2, 87-92.