

Sveučilište u Zagrebu

Filozofski fakultet

Odsjek za lingvistiku

Računalnolingvistička analiza korisničkih komentara na internetskim portalima

Diplomski rad

Studentica: Antonia Matković

Mentor: dr. sc. Božo Bekavac

Sumentor: dr. sc. Željko Agić

Zagreb, rujan 2014. g.

SADRŽAJ

1. UVOD	Error! Bookmark not defined.
1. 1. Obilježja računalno posredovane komunikacije.....	3
2. CILJEVI I SVRHA RADA	4
3. TEORIJSKI KONCEPT RADA I ISTRAŽIVAČKA PITANJA	5
3. 1. Analiza sentimenata.....	5
3. 2. Analiza polarnosti.....	7
3. 3. Analiza jezika u računalno posredovanoj komunikaciji.....	8
3. 4. Odabir teme i portala za analizu	9
3. 6. Istraživačka pitanja	11
4. METODOLOGIJA.....	13
4. 1. Sortiranje i označavanje komentara.....	13
4. 2. Analitička strategija.....	14
5. REZULTATI ANALIZA	15
5. 1. Rezultati kvantitativne analize podataka	15
5. 1. 1. Usporedba označavanja dvaju označivača	15
5. 1. 2. Usporedba portala prema duljini komentara, frekvenciji sentimenata, ukupnom broju polarnih frazi i korištenom jeziku	16
5. 1. 3. Usporedba komentatora na temelju iskazanih sentimenata i korištenog jezika.....	21
5. 2. Rezultati kvalitativne analize podataka	27
5. 2. 1. Analiza sentimenata.....	27
5. 2. 2. Analiza polarnosti.....	30
5. 2. 3. Analiza korištenog jezika	31
6. RASPRAVA I ZAKLJUČCI	34
7. LITERATURA.....	36
8. SAŽETAK.....	42

1. UVOD

U proteklih se dvadesetak godina paralelno s razvojem interneta, razvijalo i umreženo društvo – novi oblik društva koje karakterizira sve šira upotreba interneta te informacijskih i komunikacijskih tehnologija općenito. Glavno obilježje umreženog društva je transformacija domene komunikacije, koja uključuje i medije. Novi komunikacijski sustav opisuju tri trenda: organiziranost komunikacije oko velikih poslovnih medijskih konglomerata koji su globalni i lokalni u isto vrijeme (ovo uključuje televiziju, radio, tiskovine, audiovizualnu produkciju, izdavaštvo, snimanje i produkciju glazbe te komercijalne *online* tvrtke); sve veća digitalizacija sustava komunikacije koji postaje i sve interaktivniji te ubrzani razvoj horizontalnih mreža komunikacija neovisnih o medijima i vladama – tzv. samoupravljana (inicirana od strane pojedinca ili grupe) masovna komunikacija (Castells, 2005).

Razvoj ovih komunikacijskih interaktivnih tehnologija koje korisnicima omogućavaju da proizvode i objavljuju vlastite sadržaje potaknuo je stvaranje nove „sudjelujuće“ kulture. Granice između proizvođača medijskog sadržaja i konzumenta istog polako nestaju (fenomen „*prodsusage*“), a dugu tradiciju pisama čitatelja uredništvu novina i časopisa zamjenjuju korisnički komentari na internetskim portalima (Millioni & Vadaratsikas, 2012).

Potencijalni utjecaj tih korisničkih komentara na razvoj javnog mnijenja, demokratskog društva te na same medije postaje predmetom proučavanja brojnih znanstvenika iz humanističkih i društvenih područja. Dok neki smatraju da korisnički komentari doprinose jačanju i širenju demokracije povećavajući dotok informacija i omogućujući pojedincima da stvore stavove poduprte činjenicama te istovremeno stvarajući diskurzivne prostore u kojima javnost može razmijeniti mišljenja i uključiti se u javnu diskusiju, drugi su skeptični prema ovakvom demokratizirajućem potencijalu interneta. Smatraju da ovo dovodi do fragmentiranja javnog prostora gdje individualna sloboda izražavanja rezultira polariziranim i ekstremnim stavovima, a ne javnom mnijenju formiranom prema Habermasovom kriteriju za kritičko-racionalni govor (Millioni & Vadaratsikas, 2012).

1. 1. Obilježja računalno posredovane komunikacije (CMC)¹

Bubaš (2001) računalno posredovanu komunikaciju definira kao razmjenu informacija u tekstualnim, audio i/ili video formatima koji se prenose i kontrolirani su uz pomoć računala i telekomunikacijskih tehnologija. CMC predstavlja bazu za interpersonalnu interakciju putem *groupware* sustava koji su dizajnirani tako da omoguće ljudima ostvarivanje zajedničkog cilja (ovdje spadaju softverski proizvodi poput e-pošte, čavrljaonica i kalendara, a općenitiji pojam društvenog softvera obuhvaća i *online dating* servise i društvene mreže poput Facebooka i Twittera).

Osim očitih prednosti CMC-a nad komunikacijom licem-u-lice (npr. prevladavanje geografskih i socioekonomskih granica i podjela, lakoća prijenosa i širenja informacija i slično), kad se te dvije vrste komunikacije uspoređuju, mogu se zamijetiti i određeni nedostaci računalno posredovane komunikacije (pogotovo one putem pisanog teksta) u odnosu na komunikaciju licem-u-lice. Primjerice, komunikacija licem-u-lice je neposredna, brža, sadrži informacije nedostupne u CMC-u (npr. govor tijela, intonacija, različiti neverbalni signali kojim se govornici svakodnevno služe i slično), dok je u CMC-u povratna informacija često reducirana i u kvantiteti i u kvaliteti (što povećava mogućnost nesporazuma), osobe uključene u komunikaciju mogu biti posve anonimne, a „društvena prisutnost“ primatelja poruke je reducirana, što olakšava problematična ponašanja poput dezinhibicije, burnih reakcija, agresije i slično (Bubaš, 2001).

Kako navodi Schroeder (2011), ljudi se ponašaju „društvenije“ u komunikaciji licem-u-lice; pristojniji su i suzdržaniji. Jedna od posljedica reducirane „društvene prisutnosti“ i relativne anonimnosti CMC-a je tendencija ka zloupotrebi interneta kao medija. Zbog nedostatka negativnih povratnih informacija, korisnici često znaju kršiti društvene norme i iskazivati niži stupanj samokontrole ne obazirući se pritom na efekt vlastitih postupaka i riječi (jer taj efekt im nije neposredno vidljiv). Neki od najproblematičnijih ponašanja vezanih za interpersonalne aspekte korištenja interneta su nedostatak inhibicije u ponašanju, verbalna agresija, vrijeđanje i zlonamjerno komentiranje (*flaming*) te neprimjereno samootkrivanje (Bubaš, 2001).

¹ CMC skraćena je engleskog termina *computer-mediated communication*, u hrvatskom prijevodu „računalno posredovana komunikacija“. U tekstu ćemo dalje upotrebljavati isključivo englesku skraćenicu spomenutog termina (a ne njenu potencijalnu hrvatsku inačicu), zbog njene uvriježenosti i prepoznatljivosti u široj akademskoj zajednici.

2. CILJEVI I SVRHA RADA

U ovom radu bavimo se analizom korisničkih komentara na hrvatskim internetskim portalima, točnije trima razinama analize – analizom sentimenata (na više razina, tzv. *aspect-based* pristup – Feldman, 2013), analizom polarnih izraza (unigrama, bigrama te ostalih n-grama) te analizom korištenog jezika.

Sve su navedene analize obavljane ručno kako bi se dobio uvid u specifičnosti komentara i komentatora na hrvatskim internetskim portalima te kako bi se uočili potencijalni problemi pri izradi računalnih alata za hrvatski jezik koji bi služili za automatsku detekciju sentimenata, stavova i polarnosti u tekstu. Većina takvih dosad postojećih alata izrađena je za engleski jezik, i to uglavnom za njegovu standardnu varijantu (npr. Balahur i sur., 2011; Duric i Song, 2011 te Sharma i Bhattacharya, 2013).

Dosad je za hrvatski jezik izrađen jedan računalni alat za traženje i prikupljanje tekstova (dnevnih horoskopa) na *web* portalima te automatsku detekciju polarnih frazi i analizu ukupnih sentimenata u dnevnim horoskopima - *Sentiscope* (Agić i Merkle, 2012) i jedan prototipni modul za automatsku detekciju i klasifikaciju polarnih frazi u neviđenom tekstu iz područja financija (Agić i sur., 2010). Oba navedena alata analiziraju tekstove napisane standardnim hrvatskim jezikom. Također, Glavaš i suradnici (2013) su testirali poluautomatsku metodu prikupljanja stavova korisnika (metoda je testirana na recenzijama restorana) i identifikacije tih stavova uz pomoć posebnih indikatora.

Korist ovakvih alata je višestruka – mogu se tako koristiti pri brandiranju proizvoda te u marketingu i oglašavanju jer analize sentimenata i polarnih frazi daju u vid u javno mnijenje, tj. osjećaje građana prema određenoj temi (Kennedy, 2012). Albert i sur. u svom radu iz 2011. g. navode da je analiza stavova bitna za proizvođače i pružatelje određenih usluga jer im može dati uvid u argumente kupaca ili klijenata o njihovom proizvodu, hotelu, restoranu i slično, tj. uvid u njihovu *e-reputaciju*. Još jedna zanimljiva i korisna upotreba analize sentimenata predstavljena je u gorenavedenom radu Agića i sur. (2010). Oni su pronašli statistički značajnu povezanost između ukupnog tržišnog trenda na Zagrebačkoj

burzi i broja pozitivno i negativno intoniranih članaka napisanih za vrijeme „vladavine“ tih trendova.

3. TEORIJSKI KONCEPT RADA I ISTRAŽIVAČKA PITANJA

Temeljita pretraga baza domaćih znanstvenih radova uz pomoć *online* repozitorija hrvatskih znanstvenih radova² pokazuje da su se analizom sentimenata i analizom polarnih frazi u Hrvatskoj bavili samo Agić zajedno sa suradnicima Ljubešićem, Merkler i Tadićem s Filozofskog fakulteta u Zagrebu (Agić i Merkler, 2012 te Agić i sur., 2010) te Glavaš sa suradnicima Korenčićem, Šnajderom i Dalbelo Bašić sa zagrebačkog FER-a (Glavaš i sur., 2010 te Glavaš i sur., 2012). S druge strane, u svijetu je na ovu temu napisano preko 7 000 članaka (Feldman, 2013).

Analizom hrvatskog jezika na internetu bavilo se više znanstvenika, iz različitih područja (npr. Filipan-Žignić³, Mihaljević⁴, Nemeth-Jajić i Milinović⁵, Omrčen⁶, itd.).

U ovom poglavlju ćemo uz definicije metoda koje ćemo koristiti u analizi komentara na portalima, predstaviti i radove (mahom inozemne) koji se bave tim područjima. U nastavku poglavlja, obrazložiti ćemo odabir teme i portala za analizu, a na samom kraju poglavlja navest ćemo istraživačka pitanja.

3.1. Analiza sentimenata

Analizu sentimenata (ili „rudarenje stavova“, *opinion mining*) možemo definirati kao zadatak traženja stavova autora o specifičnim temama. Naime, na proces donošenja odluka utječu i stavovi autoriteta, ali i običnih ljudi (npr. prije kupnje nekog proizvoda, ljudi uglavnom traže recenzije tog proizvoda na internetu). Zahvaljujući razvoju društvenih mreža poput Twittera i Facebooka, raznih foruma i blogova, danas smo svjedoci eksplozije sentimenata na internetu. Isječci teksta koji sadrže sentimente predstavljaju „rudnik zlata“ za tvrtke i pojedincima koji žele

² Hrvatska znanstvena bibliografija ili CROSBİ (Croatian Scientific Bibliography). URL: <https://bib.irb.hr/>

³ <https://bib.irb.hr/prikazi-rad?&rad=589329>

⁴ Mihaljević, 2005

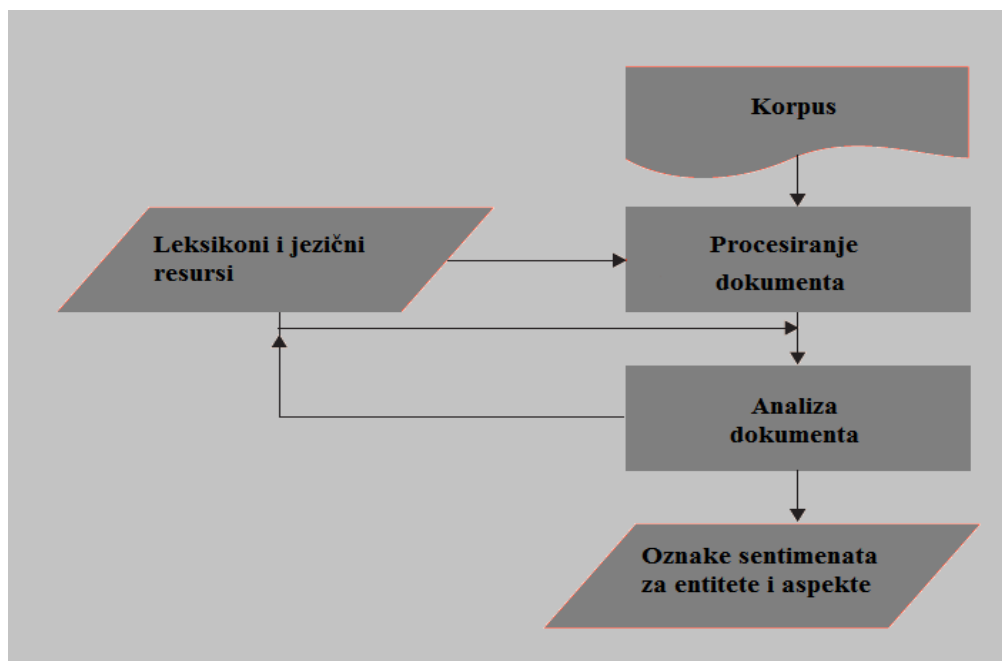
⁵ Nemeth-Jajić i Milinović, 2012

⁶ <https://bib.irb.hr/prikazi-rad?&rad=52052>

pratiti mijenjanje vlastite reputacije na internetu i dobiti pravovremene povratne informacije o svojim proizvodima i/ili radnjama. Od analize sentimenata tako mogu imati dobit različitih marketinški menadžeri, tvrtke koje se bave odnosima s javnošću, voditelji različitih kampanja, političari, različiti ulagači, ali i obični ljudi koji koriste kupnju putem interneta (Feldman, 2013).

Kad je riječ o pristupima analizi sentimenata, Feldman (2013) ih navodi pet – analiza sentimenata na razini dokumenta (najjednostavniji oblik; pretpostavlja se da se u dokumentu nalazi jedinstveni stav autora); analiza sentimenata na razini rečenice (pretpostavlja se da svaka rečenica sadrži stav); analiza sentimenata na razini aspekata (pretpostavlja se da kad ljudi govore o određenoj temi ili proizvodu, zapravo imaju stav o više aspekata ili svojstava te teme ili proizvoda); komparativna analiza sentimenata (koristi se kad korisnici daju svoj stav uspoređujući više proizvoda) te analiza sentimenata uz pomoć usvojenog leksikona (leksikoni sentimenata koriste u algoritmima za analizu sentimenata; bili oni uneseni ručno, uz pomoć rječnika ili uz pomoć alata poput WordNeta).

Na slici 1 prikazano je kako arhitektura generičkih sustava za analizu sentimenata izgleda u praksi (kod automatskog označavanja sentimenata; različiti alati i algoritmi).



Slika 1. Arhitektura generičkih sustava za analizu sentimenata (Feldman, 2013)

Pri analizi sentimenata, analitičari se susreću s nekoliko problema. Ključni problemi odnose se na preciznost analize, količinu te „čistoću“ podataka. Naime, analiza sentimenata u medijima nije nimalo lagan zadatak, jer se često ni ljudski označivači ne mogu složiti oko označavanja sentimenata zbog različitih načina izražavanja korisnika (korištenje ironije, sarkazma, humora i skraćenica). Zato se 70% ispravno detektiranih sentimenata u sustavu za analizu sentimenata (70% slaganja s ljudskim označivačima) smatra dobrim postotkom. Neki istraživači tvrde da je mnogo realniji postotak 30% ispravno detektiranih sentimenata. Zbog toga neke tvrtke koriste isključivo ljudske označivače (Kennedy, 2012). Duric i Song u svom radu iz 2011. g. navode da je postotak slaganja između ljudskih označivača najčešće između 58 i 76%, a Bonev i suradnici u svom radu iz 2012. godine analizu sentimenata smatraju tek predviđanjem kako bi većina čitatelja reagirala na tekst, jer interpretacija sentimenata ovisi o lingvističkim vještinama te emocionalnoj inteligenciji čitatelja, ali i autora teksta.

Osim ispravnosti označavanja, dodatni problem pri analizi sentimenata predstavlja „čistoća“ podataka. Kad je riječ o pisanju recenzija na proizvode, javlja se problem pretjeranog pozitivizma, tj. korisnici su skloniji ostavljanju pozitivnih komentara nego negativnih. K tome, neke tvrtke vjerojatno i same pišu recenzije za vlastite proizvode te tako potiskuju negativne recenzije stvarnih korisnika (Kennedy, 2012).

3.2. Analiza polarnosti

Klasifikacija polarnosti ima za cilj identificirati pozitivne i negativne polarnosti u tekstu, na različitim razinama (razine dokumenta, rečenice, fraze ili riječi). Pritom se možemo voditi principom slaganja, koji glasi: „Značenje kompleksnog izraza se određuje uz pomoć njegove strukture i značenja njegovih sastavnica.“ Ako slijedimo ovaj princip, polarnost djelića teksta može se odrediti iz polarnosti njegovih sastavnica. Za primjer takvog slaganja može poslužiti rečenica: „On je dobar lažljivac.“ koja se klasificira kao negativna jer pozitivni pridjev i negativna imenica zajedno tvore negativnu imensku frazu (Klenner i sur., 2009).

Sharma i Bhattacharyya u svom radu iz 2013. godine objašnjavaju važnost izrade leksikona polarnih frazi za specifične domene i njihove prednosti u odnosu

na univerzalne leksikone. Takvi leksikoni bi uvelike olakšali analizu sentimenata jer neka polarna fraza ili riječ u jednoj domeni može bit pozitivna, dok je u drugoj veoma negativna – primjerice, polarna riječ „nepredvidljiv“ je u domeni filma pozitivna, dok je u recenziji nekog kućanskog aparata negativna.

Unatoč potencijalnoj upotrebi, i analiza polarnosti u tekstu susreće se s nizom problema, baš kao i analiza sentimenata. Neki od tih problema su: ovisnost principa slaganja o upotrebi (npr. je li fraza „savršeni špijun“ pozitivna ili negativna?), neodređenost principa slaganja (npr. ako je fraza „savršeni špijun“ pozitivna, zašto je fraza „savršena gnjavaža“ negativna?), neutralne riječi zajedno mogu tvoriti polarnu frazu (npr. „hladni odgovor“), implicitni stavovi i figurativni govor (npr. u rečenici „Bio sam sretan kad je moj očuh nestao.“ negativan stav prema očuhu je impliciran), a prije određivanja polarnosti trebalo bi razriješiti višeznačnost frazi (npr. „jeftina terapija“ može biti pozitivna fraza ako se pridjev „jeftin“ odnosi na nisku cijenu, a može biti i negativna fraza ako se pridjev „jeftin“ odnosi na nisku kvalitetu) (Klenner i sur., 2009).

3.3. Analiza jezika u računalno posredovanoj komunikaciji

Szécsi (2013) iznosi teoriju da u svijetu računalno posredovane komunikacije norme usmene i pismene komunikacije polako gube na važnosti i prestižu. Ekspanzija nestandardnih jezičnih varijeteta koje preferiraju umreženi pojedinci ubrzana je upotrebom elektroničkih medija (računalo i mobilni uređaji), a osobito korištenjem multimedijalnih poruka (prijenos govora, teksta i slika u realnom vremenu).

Tekstovi većine e-mailova i SMS poruka zapravo pripadaju žanru govora, a ne pisanih tekstova. Gramatičke i stilske karakteristike tih poruka mogu se smatrati oznakama posebne vrste usmene komunikacije. Ovi gramatički i stilski elementi su posredstvom novih komunikacijskih tehnologija integrirani u tekstove mnogo slobodnije nego u usmene iskaze. Upotreba ovih posebnih lingvističkih formi dovodi do približavanja i miješanja usmenosti i pismenosti. Zahvaljujući pojavi posebnog slikovnog jezika (pojava emotikona ili „smajlića“), proces miješanja i sinteze lingvističkih obilježja usmenih i pismenih oblika komunikacija uvelike je

ubrzan. Pojava ovakvog slikovnog jezika može doprinijeti usvajanju novog, dodatnog, globalnog jezika (Szécsi, 2013).

Kada je riječ o upotrebi hrvatskog jezika u pisanoj komunikaciji posredovanoj računalom, Mihaljević (2005) uočava oživljavanje arhaičnog glagolskog oblika - aorista (zbog kratkoće), široku upotrebu emotikona, povećanje pravopisnih i jezičnih pogrešaka (dijelom zbog brzine), upotrebu poštapalica, dijalektalizama i žargonizama, anglizama (pa i engleskih riječi). Nemeth-Jajić i Milinović (2012) navode još i pojavu slovopisnih pomodnica, uporabu višeslovnih kombinacija grafema umjesto dijakritika (npr. „sh“ umjesto „š“), oživljavanje glagoljice i morfonološkoga pravopisa, redukciju samoglasnika, suglasnika pa i slogova, uporabu brojaka u funkciji tvorbenih elemenata, ustaljene kratice kao dio internetskoga žargona (npr. „lol“ – *laughing out loud* ili „smijem se naglas“) te odražavanje funkcionalne sintakse razgovornoga diskursa s mogućnošću umnogostručenja replika.

3.4. Odabir teme i portala za analizu

U ovom radu analiziram korisničke komentare na članke napisane na temu prikupljanja potpisa za referendum o braku. Udruga U ime obitelji je 12. svibnja 2013. g. započela prikupljati potpise za referendum o braku (predloženo referendumsko pitanje: “Jeste li za to da se u Ustav RH unese odredba po kojoj je brak životna zajednica žene i muškarca?”)⁷ i taj proces završen je dva tjedna kasnije, 26. svibnja 2013. g. To je ujedno bio i vremenski okvir unutar kojeg su komentari koje analiziram i nastali. Navedena tema istraživački je bila posebno zanimljiva jer je u hrvatskom društvu izazvala brojne kontroverze te su čitatelji ispod članaka koji su se nje doticali producirali više komentara no obično.

Portali za analizu odabrani su prema čitanosti⁸ te uz uvjet da su komentari na njima pisani preko Facebook računa, jer bi komentari forumskog tipa (uz višestruka citiranja unutar jednog unosa te citiranja jednog komentara nakon nekoliko ispisanih stranica ostalih komentara) predstavljali praktični problem za ručnu analizu. Uz

⁷ Detaljnije: <http://uimeobitelji.net/>

⁸ Prema portalu <http://www.alexa.com/>, najčitaniji news internetski portali u RH su, redom: index.hr, jutarnji.hr, 24sata.hr i net.hr.

jutarnji.hr i net.hr koji se nalaze unutar 4 najposjećenija portala u Hrvatskoj, analiziran je i jedan katolički portal – bitno.net kako bi se dobili što raznovrsniji komentari. Naime, portali jutarnji.hr i net.hr su od strane su komentatora često proglašavani pristranima (tj. protivnicima inicijative i referendumu). Primjeri takvih komentara:

„To je tipična retorika zlonamjernih medija sa ideološkim predznakom. Ako promatrač (čitatelj) ne zna koristiti zdrav razum, bilo tko će mu retorikom uspjeti inputirati polu-istine. Osim toga, jasno jest da su argumenti "psihologa, terapeuta i pravnika", čije je izjave Yutarnji plasirao, neozbiljna, neutemeljena, neobjektivna i pristrana tumačenja pojava u društvu, a posebno apstrakta "ljudska prava" na koji se svi danas površno pozivaju.“ – komentar korisnika I. S. na članak <http://www.jutarnji.hr/-pitanje-braka-nije-nesto-sto-se-definira-ustavom-/1105243/>

„čestitam autorici! napisati ovakav tekst, potpisati ga i objaviti u javnom mediju, pa to je za rubliku nemo'sh vjerovat! ovaj tekst treba svakako pohraniti i uvrstiti u program obrazovanja mladih novinara kao primjer subjektivne manipulacije informacijama. ima se što analizirati u svakoj rečenici... sjajan materijal... ma ne, zapravo ne, jer ovdje je ta manipulacija tako očita da je to u stvari jedno... a baš kad sam pomislila da u našim medijima gore ne može...“ – komentar korisnice D. B. na članak <http://www.jutarnji.hr/pankerice-razbile-stand-crkvenih-aktivista-koji-su-protiv-gay-brakova/1103259/>

„Podsjetimo, mediji su već pisali o sumnjivim metodama kojima se koristi homofobna inicijativa U ime obitelji" Zar se ova inicijativa sama prozvala homofobnom? ili je ovo etiketiranje tekstopisca. No a gdje navodite da je ovo plaćeni tekst jednog homofila, nema nigdje..Jadno i vulgarno.“ – komentar korisnika I. L. na članak

<http://danas.net.hr/hrvatska/i-policija-intervenirala-volonteri-iz-u-ime-obitelji-ometaju-izbore>

„kako to da ljudima nije očito da velika većina portala jednostavno potkopava ovu akciju? šire se samo laži i dezinformiraju se ljudi... kakva je to demokracija i objektivnost? nikakva!“ – komentar korisnika D. T. na članak <http://danas.net.hr/hrvatska/i-policija-intervenirala-volonteri-iz-u-ime-obitelji-ometaju-izbore>

Uoči referenduma (koji je održan 1. 12. 2013. godine) uredništvo Jutarnjeg lista se izjasnilo protiv referenduma, odbili su objektivno izvještavati o kampanji te su oglasni prostor na portalu i u tiskanom izdanju lista darovali osobama koje se osjećaju ugroženima referendumom⁹, čime se odluka o odabiru portala bitno.net kao „protuteže“ pokazala ispravnom.

3.5. Istraživačka pitanja

Kako je već ranije najavljeno, ovaj rad ima nekoliko ciljeva – otkriti potencijalne probleme za izradu računalnih alata za analizu sentimenata i analizu polarnosti u hrvatskim tekstovima te opisati obilježja komentara tj. komentatora na hrvatskim portalima. Kako je literatura u ovom području prilično oskudna, a u području hrvatske lingvistike gotovo nikakva (izuzevši radove Agića i suradnika), odlučili smo se na postavljanje istraživanja uz pomoć istraživačkih pitanja (dalje u tekstu: IP), radije nego da postavljamo hipoteze i predviđamo smjer povezanosti i razlika između obilježja koje mjerimo.

Odredili smo 6 skupina istraživačkih pitanja. U prve tri skupine nalaze se pitanja vezana za specifičnosti analize sentimenata, analize polarnih frazi i analize jezika na hrvatskim portalima. Odgovori na ova pitanja mogu služiti kao smjernice prilikom izrade računalnih alata za navedene analize. U skupinama 4, 5 i 6 nalaze se pitanja vezana za specifičnosti komentara na ovu temu i općenito za specifičnosti komentara na portalima.

⁹ <http://www.jutarnji.hr/odluka-urednistva-jutarnji-donira-oglasni-prostor-gradanima-ugrozenima-referendumom/1138602/>

Prva skupina istraživačkih pitanja (IP1): Je li analiza sentimenta prikladna analiza za ovako kompleksnu temu? Komentiraju li uopće temu komentatori? Koliko je komentara u kojima komentatori ne komentiraju direktno članak ili inicijativu?

Druga skupina istraživačkih pitanja (IP2): Kakav je odnos polarnih frazi na portalima? Koliko je komentara u kojima nema polarnih frazi, a nose jasno određen stav (sentiment) prema temi?

Treća skupina istraživačkih pitanja (IP3): Kakvim se jezikom generalno služe komentatori na portalima? Postoje li razlike između među portalima?

Četvrta skupina istraživačkih pitanja (IP4): Kakva je povezanost između sentimenta prema članku i sentimenta prema inicijativi? Postoje li razlike u statističkoj značajnosti i snazi te veze između tri analizirana portala?

Peta skupina istraživačkih pitanja (IP5): Kakva je povezanost između sentimenta prema inicijativi i ukupnog sentimenta komentara? Postoje li razlike u statističkoj značajnosti i snazi te veze između tri analizirana portala?

Šesta skupina istraživačkih pitanja (IP6): Kakva je povezanost između ukupnog sentimenta komentara i korištenog jezika? Postoje li razlike u statističkoj značajnosti i snazi te veze između tri analizirana portala? Postoji li povezanost između stava prema inicijativa i korištenog jezika? Mijenja li se korišteni jezik komentatora s obzirom na sentiment komentara (tj. postoje li *outlieri*)?

4. METODOLOGIJA

4.2. Sortiranje i označavanje komentara

Nakon što su kopirani u Excel tablicu, svi komentari su ručno označeni od strane dvoje označivača kako bi se uočili potencijalni problemi ovakvog označavanja sentimenata, polarnih frazi i korištenog jezika, kao i potencijalni problemi koji bi se mogli pojaviti pri automatskom označavanju.

Pri kopiranju komentara u Excel tablicu te pri njihovom označavanju, korišten je sljedeći redoslijed, prema stupcima:

- stupac A: link na originalni članak
- stupac B: prvi (početni) komentar
- stupac C: odgovori na početni komentar
- stupac D: sentiment komentatora prema inicijativi
- stupac E: sentiment komentatora prema članku
- stupac F: sentiment komentatora prema ishodišnom komentaru
- stupac G: ukupni sentiment
- stupac H: korišteni jezik
- stupac I: broj pozitivnih frazi u komentaru
- stupac J: broj negativnih frazi u komentaru

Sentimenti su tako označavani po sljedećem principu:

- / - sentiment nije izražen (uz napomenu da ova ocjena ne vrijedi za stupac G; ukupni sentiment mora biti označen)
- POZ – pozitivan sentiment (npr.: „*Svaka čast na inicijativi na prikupljenim potpisima!*“)
- NEU – neutralan sentiment (npr.: „*Pitat ću te što i mnoge druge na ovom portalu... Bi li ti dao svoju kćer u brak s homoseksualcem da imaju djecu i poboljšavaju natalitet?*“)
- NEG – negativan sentiment (primjer komentara s NEG sentimentom: „*Prava kršćanska ljubav - prava koja za sebe tražiš, oduzeti svomu bližnjemu. Sramite se.*“)

Korišteni jezik označen je po sljedećem principu:

- 1 – standardni jezik (npr. „*Brak bi trebao biti zajednica isključivo ljudi koji se vole, razumiju i poštuju. Mrzim netoleranciju, licemjerstvo, sebičnost i diskriminaciju svake vrste!*“)
- 2 – nestandardni jezik, malo odstupanje od standarda (npr. korištenje stranih riječi, slenga i sl.: „*Sve te koji zagovaraju *pederluk* neka svoju djecu daju na čuvanje (babysitting) tim i takozvanim LGBT osobama*“)
- 3 – nestandardni jezik, veliko odstupanje od standarda (npr. korištenje dijalekata: „*To Bozy, bez podrške vas na vrju nećemo uspit.*“)

	A	B	C
1	http://danas	T F : Imperija uzvraća udarac. :-)	
2			L E : a tko je imperija

U nastavku reda:

	D	E	F	G	H	I	J
	/	/	/	POZ	1	0	0
	/	/	NEU	NEU	1	0	0

Slika 2. Primjer sortiranja i označavanja komentara u Microsoft Excel tablici

4.3. Analitička strategija

Nakon što su komentari označeni, daljnja obrada uključivala je neke jednostavne kvantitativne analize poput izračuna prosječnog broja riječi po komentaru, frekvencija sentimenata i slično te izračun povezanosti između sentimenata komentara na više razina i korištenog jezika uz pomoć parametrijskih procedura - Pearsonovog koeficijenta korelacije (gdje su komentari bili uglavnom polarni – pozitivni ili negativni) te jednosmjernog ANOVA testa (uz Laveneov test homogenosti varijanci i post hoc test; za usporedbu između sentimenata i korištenog jezika). Za usporedbu označavanja dvaju označivača korišten je Cohenov Kappa koeficijent. Sve su ove analize izvršene uz pomoć programskog paketa SPSS v19. Polarne fraze su iz praktičnih razloga označivane u NE Simple Annotatoru te potom kopirane natrag u Excel tablicu.

Daljnje kvalitativne analize uključivale su pregledavanje komentara uz uočavanje potencijalnih problema u automatskom označavanju, a uz usporedbu s rezultatima dobivenima u (mahom inozemnoj) literaturi.

5. REZULTATI ANALIZA

Rezultati analiza obavljenih nakon označavanja sentimenata, polarnosti i korištenog jezika od strane autorice rada (pri ovim analizama iz praktičnih razloga zanemareno je označavanje drugog, pomoćnog označivača) podijeljeni su u dva potpoglavlja na rezultate kvantitativne analize podataka i na rezultate kvalitativne analize podataka.

U potpoglavlju „Rezultati kvantitativne analize podataka“ prikazani su rezultati usporedbe dvaju označivača (uz pomoć frekvencije razlike i Cohenovog Kappa koeficijenta), analiza frekvencija te izračuni povezanosti ili razlike između dvije varijable. Svi navedeni rezultati koristit će pri analizi komentara (i komentatora općenito) na hrvatskim internetskim portalima.

U potpoglavlju naslovljenom „Rezultati kvalitativne analize podataka“ prikazani su rezultati kvalitativne analize sentimenata, polarnih frazi i korištenog jezika u komentarima. Navedeni rezultati koristit će se za analizu specifičnosti teme obrađene u člancima, kao i za analizu učinkovitosti korištenja ovih metoda na korisničkim komentarima na hrvatskim internetskim portalima.

5.1. Rezultati kvantitativne analize podataka

5.1.1. Usporedba označavanja dvaju označivača (frekvencija razlike i Kappa koeficijent)

Kad je riječ o slaganju dvaju označivača u označavanju sentimenata, frekvencija razlike 0 (što znači da se oba označivača slažu u ocjeni) iznosi 76%, a Cohenov Kappa koeficijent, konzervativna mjera koja u obzir uzima i slučajno slaganje, iznosi 0,63, što se smatra značajnim slaganjem.

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement Kappa	,629	,011	56,324	,000
N of Valid Cases	3183			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Slika 3. Prikaz izračuna Cohenovog Kappa koeficijenta (slaganje označivača u označavanju sentimenata) u SPSS-u. U stupcu „Value“ su najbitniji podaci: iznos Kappa koeficijenta te broj komentara koji su uzeti u obzir pri izračunu.

S druge strane, pri označavanju korištenog jezika, označivači se slažu u 72% slučajeva (prema frekvenciji razlika), a Cohenov Kappa koeficijent iznosi 0,53, što se smatra umjerenim slaganjem.

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement Kappa	,533	,026	18,992	,000
N of Valid Cases	795			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Slika 4. Prikaz izračuna Cohenovog Kappa koeficijenta (slaganje označivača u označavanju korištenog jezika) u SPSS-u.

5.1.2. Usporedba portala prema duljini komentara, frekvenciji sentimenata, ukupnom broju polarnih frazi i korištenom jeziku

	Ukupan broj komentara	Ukupan broj riječi	Prosječan broj riječi po komentaru
Portal jutarnji.hr	282	12 383	43,91
Portal net.hr	308	11 313	36,73
Portal bitno.net	206	15 781	55,96

Kako je vidljivo u gornjoj tablici, najdulje komentare pisali su čitatelji portala bitno.net (prosječna duljina komentara: 55,96 riječi), zatim slijede komentari čitatelja portala jutarnji.hr (prosječna duljina komentara: 43,91 riječi) te komentari čitatelja portala net.hr (prosječna duljina komentara: 36,73 riječi).

Očekivano, najpozitivniji odnos prema inicijativi U ime obitelji imaju čitatelji katoličkog portala bitno.net (31,2% pozitivnih komentara, 6,3% negativnih te 62,5% komentara u kojima sentiment prema inicijativi nije direktno izražen ili je neutralan). Na druga dva portala može se primijetiti značajna razlika u postotku negativnih komentara o samoj inicijativi u odnosu na portal bitno.net. Tako na portalu jutarnji.hr ima 16,7% negativnih i 12,5% pozitivnih komentara (70,8% komentara nema izražen ili ima neutralan sentiment prema inicijativi), a na portalu net.hr 22,1% negativnih i 18,6% pozitivnih (59,3% komentara nema izražen ili ima neutralan sentiment prema inicijativi).

Najnegativniji odnos prema članku (tj. češće – autoru članka) imaju čitatelji portala jutarnji.hr - 16,4% negativnih, 7,5% pozitivnih te 76,1% komentara s neizraženim stavom i neutralnih komentara. Gotovo obrnuta situacija je u slučaju portala bitno.net gdje je 18% pozitivnih, 5,9% negativnih te 76,1% s neizraženim stavom i neutralnih komentara.

Gledajući ukupan sentiment komentara, portal bitno.net ponovno se znatno razlikuje od druga dva portala. Naime, na tom su portalu, ukupno gledajući, komentari najpozitivniji (67,3% pozitivnih i neutralnih te 32,7% negativnih komentara). Najnegativniji portal je ponovno jutarnji.hr sa 65,8% negativnih te 34,2% pozitivnih i neutralnih komentara. Nakon njega slijedi portal net.hr sa sličnim postocima – 61,2% negativnih te 38,8% pozitivnih i neutralnih komentara.

Frequencies

[DataSet2]

Statistics

bitno_odn_ukupno

N	Valid	205
	Missing	0

bitno_odn_ukupno

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NEG	67	32,7	32,7	32,7
	NEU	96	46,8	46,8	79,5
	POZ	42	20,5	20,5	100,0
	Total	205	100,0	100,0	

Slika 5. Izračun frekvencija ukupnog sentimenta komentara pomoću SPSS-a; portal bitno.net. U prvoj tablici prikazan je broj komentara koje smo analizirali, u prvom stupcu druge tablice prikazane su oznake za sentimente, u drugom su frekvencije, u trećem postoci, u četvrtom valjani postoci, a u petom kumulativni iznos postotaka

Frequencies

[DataSet1]

Statistics

JL_odn_ukupno

N	Valid	281
	Missing	0

JL_odn_ukupno

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NEG	185	65,8	65,8	65,8
	NEU	71	25,3	25,3	91,1
	POZ	25	8,9	8,9	100,0
	Total	281	100,0	100,0	

Slika 6. Izračun frekvencija ukupnog sentimenta komentara pomoću SPSS-a; portal jutarnji.hr

Frequencies

[DataSet1]

Statistics

net_odn_ukupno

N	Valid	307
	Missing	0

net_odn_ukupno

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NEG	188	61,2	61,2	61,2
	NEU	105	34,2	34,2	95,4
	POZ	14	4,6	4,6	100,0
	Total	307	100,0	100,0	

Slika 7. Izračun frekvencija ukupnog sentimenta komentara pomoću SPSS-a; portal net.hr

Odnos pozitivnih i negativnih frazi na portalu bitno.net je gotovo pola-pola (45 pozitivnih i 57 negativnih frazi u 71 komentaru; dakle 135 komentara je bez polarnih frazi), dok je na Jutarnjem oko 85% negativnih (38 pozitivnih i 215 negativnih frazi u 129 komentara; dakle 153 komentara nemaju polarne fraze), a na net.hr-u oko 95% negativnih (8 pozitivnih i 138 negativnih frazi u 115 komentara; dakle 193 komentara nemaju polarne fraze).

Jezik koji su čitatelji koristili pri komentiranju se ne razlikuje mnogo na navedenim portalima – približno polovica čitatelja koristi standardni, oko 40% čitatelja koristi nestandardni jezik s malim odstupanjem od standarda (npr. korištenje stranih riječi, slenga i sl.), a tek oko 10% komentatora koristi nestandardni jezik s velikim odstupanjem od standarda (npr. korištenje dijalekata).

Frequencies

[DataSet2]

Statistics

bitno_jezik

N	Valid	205
	Missing	0

bitno_jezik

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	109	53,2	53,2	53,2
2	73	35,6	35,6	88,8
3	23	11,2	11,2	100,0
Total	205	100,0	100,0	

Slika 8. Izračun frekvencija korištenog jezika u komentaru pomoću SPSS-a; portal bitno.net

Frequencies

[DataSet1]

Statistics

JL_jezik

N	Valid	281
	Missing	0

JL_jezik

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	150	53,4	53,4	53,4
2	96	34,2	34,2	87,5
3	35	12,5	12,5	100,0
Total	281	100,0	100,0	

Slika 9. Izračun frekvencija korištenog jezika u komentaru pomoću SPSS-a; portal jutarnji.hr

Frequencies

[DataSet1]

Statistics

net_jezik

N	Valid	307
	Missing	0

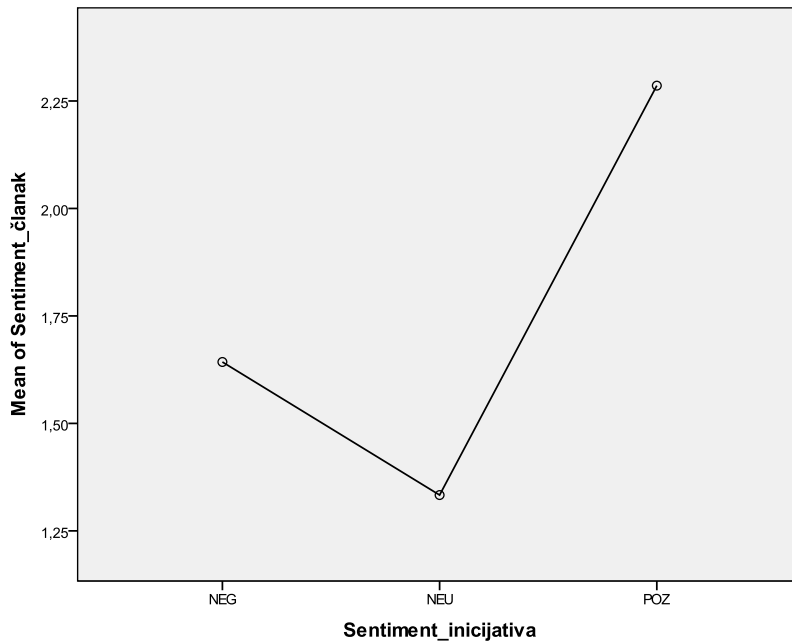
net_jezik

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	139	45,3	45,3	45,3
	2	140	45,6	45,6	90,9
	3	28	9,1	9,1	100,0
	Total	307	100,0	100,0	

Slika 10. Izračun frekvencija korištenog jezika u komentaru pomoću SPSS-a; portal net.hr

5.1.3. Usporedba komentatora na temelju iskazanih sentimenata i korištenog jezika

ANOVA testom analizirali smo statističku značajnost razlike između sentimenata korisnika prema inicijativi i prema analiziranim člancima na portalu Jutarnji.hr. Analiza je rađena na 32 komentara u kojima je bio iskazan i sentiment prema inicijativi i sentiment prema članku. Razlika između navedenih sentimenata nije statistički značajna ($F=2,68$; $df=2$; $p>0,05$)



Slika 11. Grafički prikaz odnosa aritmetičke sredine (NEG=1, NEU=2, POZ=3) sentimenta prema člancima na portalu jutarnji.hr (ordinatna os) i sentimenta komentatora prema inicijativi (apscisna os)

Na portalima net.hr i bitno.net sentimenti prema inicijativi i člancima bili su isključivo pozitivni ili negativni pa smo koristili Pearsonov koeficijent korelacije. Za portal net.hr iznos tog koeficijenta nije statistički značajan. Drugim riječima, nema statistički značajne povezanosti između sentimenta prema inicijativi i prema člancima na tom portalu. Analiza je rađena na 17 komentara u kojima je sadržan i sentiment prema inicijativi i sentiment prema članku.

Correlations - net.hr			
		Odn_članak	Odn_inicijativa
Odn_članak	Pearson Correlation	1	-,310
	Sig. (2-tailed)		,226
	N	17	17
Odn_inicijativa	Pearson Correlation	-,310	1
	Sig. (2-tailed)	,226	
	N	17	17

Slika 12. Iznos Pearsonovog koeficijenta korelacije između sentimenta komentatora prema članku i prema inicijativi na portalu net.hr prikazani je u prvom redu u trećem stupcu, u prvoj liniji. U drugoj liniji prikazana je značajnost korelacije na dvosmjernom testu, a u trećoj liniji prikazan je broj komentara u kojima se mogao odrediti sentiment i prema članku i prema inicijativi

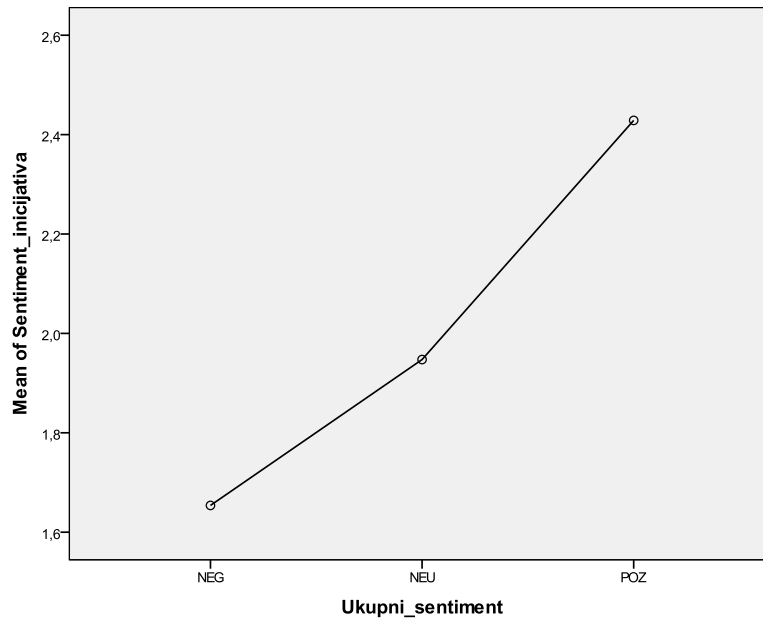
Očekivano, na portalu bitno.net, Pearsonov koeficijent korelacije iznosi $r=1$, što znači da na tom portalu postoji potpuna povezanost između sentimenta prema inicijativi i sentimenta prema članku. Analizirano je ukupno 37 komentara koji su imali izražene sentimente i prema članku i prema inicijativi.

Correlations - bitno.net			
		Odn_članak	Odn_inicijativa
Odn_članak	Pearson Correlation	1	1,000**
	Sig. (2-tailed)		,000
	N	37	37
Odn_inicijativa	Pearson Correlation	1,000**	1
	Sig. (2-tailed)	,000	
	N	37	37

** . Correlation is significant at the 0.01 level (2-tailed).

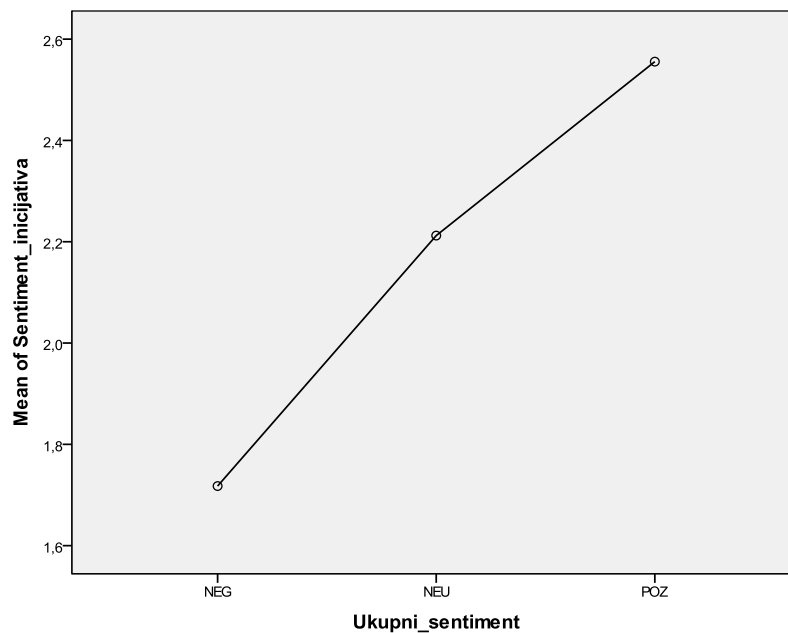
Slika 13. Iznos Pearsonovog koeficijenta korelacije između sentimenta komentatora prema inicijativi i prema člancima na portalu bitno.net

Sljedeća analiza je analiza odnosa sentimenta prema inicijativi i ukupnog sentimenta komentara. Na portalu jutarnji.hr ta je razlika statistički značajna ($F=3,9$; $df=2$; $p<0,05$), što se može vidjeti i u *means plotu* (slika 14). Što je sentiment prema inicijativi pozitivniji, pozitivniji je i ukupni sentiment komentara. Analizirano je ukupno 86 komentara u kojima je bio izražen sentiment prema inicijativi.



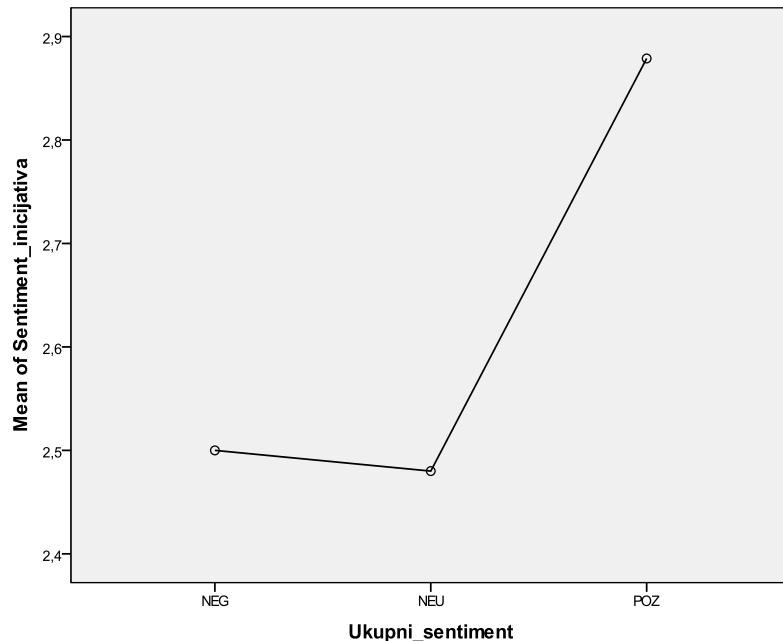
Slika 14. Grafički prikaz (*means plot*) odnosa aritmetičke sredine (NEG=1, NEU=2, POZ=3) sentimenta prema inicijativi u odnosu na ukupni sentiment (Jutarnji.hr)

Slična je situacija i s portalom net.hr. Razlika je i ovdje značajna ($F=5,44$; $df=2$; $p<0,01$). Analizirano je ukupno 128 komentara s iskazanim sentimentom prema inicijativi.



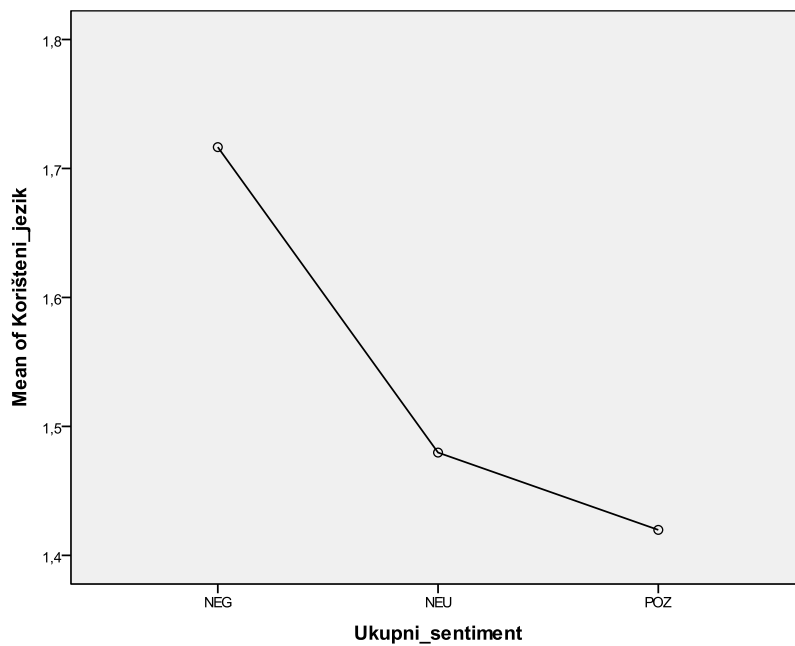
Slika 15. Grafički prikaz (*means plot*) odnosa aritmetičke sredine (NEG=1, NEU=2, POZ=3) sentimenta prema inicijativi u odnosu na ukupni sentiment (net.hr)

Jedino na portalu bitno.net razlika između ukupnog sentimenta i sentimenta prema inicijativi nije statistički značajna ($F=2,68$; $df=2$; $p>0,05$). Analizirano je ukupno 78 komentara s izraženim sentimentom prema inicijativi. Kao što se može vidjeti u *means plotu* na slici 13, sentimenti prema inicijativi su uglavnom pozitivni, dok se ukupni sentiment kreću od negativnih ka pozitivnima.



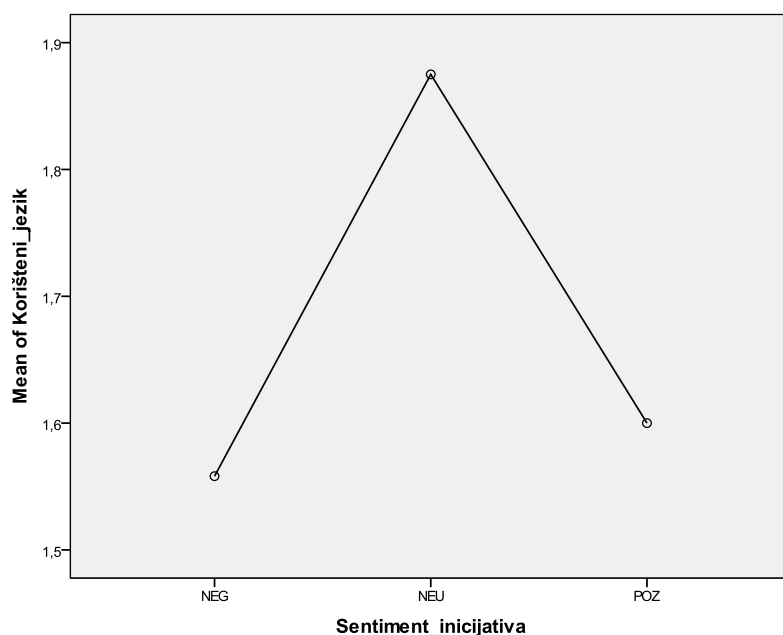
Slika 16. Grafički prikaz (*means plot*) odnosa aritmetičke sredine (NEG=1, NEU=2, POZ=3) sentimenta prema inicijativi u odnosu na ukupni sentiment (bitno.net)

Sljedeće smo analizirali postoji li razlika između ukupnog sentimenta i korištenog jezika. Prema rezultatima dobivenima ANOVA testom - ta razlika je značajna ($F=14,17$; $df=2$; $p<0,005$) – što je ukupni sentiment komentara negativniji, to je korišteni jezik u komentaru bliži dijalektu (tj. udaljeniji od standarda). Ipak, kako se može vidjeti i u *means plotu* na slici 17, aritmetičke sredine korištenog jezika u odnosu na ukupne sentimente razlikuju se tek za 0,30 (aritmetička sredina za negativni ukupni sentiment je oko 1,7, tj. ne doseže 2 ili nestandardni jezik, dok je za pozitivni 1,4).



Slika 17. Grafički prikaz (*means plot*) odnosa aritmetičke sredine korištenog jezika (1=standardni jezik; 2=nestandardni jezik, malo odstupanje od standarda; 3=nestandardni jezik, veliko odstupanje od standarda) s obzirom na ukupni sentiment komentara (svi portali)

Posljednji se kvantitativni izračun odnosi na razliku između korištenog jezika i sentimenta prema inicijativi. Ta razlika nije statistički značajna ($F=0,94$; $df=2$; $p>0,05$). Razlika između aritmetičke sredine korištenog jezika za negativni sentiment prema inicijativi i one za pozitivni sentiment prema inicijativi iznosi tek 0,04.



Slika 18. Slika 14. Grafički prikaz (*means plot*) odnosa aritmetičke sredine korištenog jezika (1=standardni jezik; 2=nestandardni jezik, malo odstupanje od standarda; 3=nestandardni jezik, veliko odstupanje od standarda) s obzirom na sentiment komentatora prema inicijativi (svi portali)

Za analizu moguće promjene u korištenom jeziku komentatora s obzirom na sentiment komentara, izabrali smo po 10-ak najčešćih komentatora sa svakog od portala, no nijedan komentator ne mijenja svoj jezik značajno. Eventualno postoje varijacije između 1 (standardni jezik) i 2 (nestandardni jezik, malo odstupanje od standarda; npr. korištenje stranih riječi, slenga i sl.), no dosta često to ovisi i o duljini komentara (npr. standardni jezik ako je komentar duljine jedne proste rečenice i nestandardni jezik ako je komentar nešto duži).

5.2. Rezultati kvalitativne analize podataka

5.2.1. Analiza sentimenata

Pristup koji smo koristili pri označavanju sentimenata je *aspect-based*, jer smo analizirali više različitih razina sentimenata vezanih za temu. Jednostavnije rečeno, u tekstu komentara smo prepoznali i označili sve izražene sentimente te ih svrstali u neku od tri grupe (sentiment prema inicijativi, sentiment prema članku, sentiment prema ishodišnom komentaru). Na kraju smo dali ocjenu ukupnom sentimentu komentara. Jasno, u svakom komentaru nisu bili izraženi svi aspekti (npr. neki

komentatori nisu uopće izražavali sentiment prema inicijativi, neki nisu izražavali prema članku i slično). Sustavi za automatsku detekciju sentimenta u tekstu imali bi problema s detekcijom više razina sentimenta koje gore navodimo jer je teško odrediti kriterije prema kojima bi se sentimenti razdvojili (npr. koji dio komentara se odnosi na sentiment prema članku, a koji na sentiment prema komentatoru), stoga je upitno koliko je automatska detekcija sentimenta na ovako kompleksnim temama uopće primjerena procedura.

Kada je riječ o komentiranju same teme, u gotovo 50% komentara nije direktno izražen sentiment komentatora niti prema inicijativi, niti prema članku, iz čega bi se dalo zaključiti da uopće ne komentiraju temu članka. U nekim slučajevima to uistinu jest tako – komentatori su se odlučili uključiti u raspravu kako bi diskutirali s drugim komentatorima (češće: kako bi ih vrijeđali) ili pak dali svoj stav o hrvatskom društvu koji nema direktne veze s temom. Primjeri za to:

„Hajde novinari JL pišite malo i o krypto komunističkoj buržuaziji koja je preuzela vlast u RH.“ – komentar korisnika B. B. na članak: <http://www.jutarnji.hr/tko-su-ljudi-na-celu-neokonzervativne-revolucije-u-hrvatskoj--društvene-vrijednosti-i-duse-mladih/1107371/>

„A da pišu o krumpiru od jagode? Imalo bi više smisla od ovog što predlažeš, u svakom slučaju.“ – odgovor korisnika O. M. na gornji komentar (članak: *ibid.*)

„A, kao četnici nisu bili prava opasnost?..Daj se umij, operi zube i nemoj kontaminirati medijski prostor "gosp'oyo" ili ti ispravnije drugarice..Tvoje je da mješaš kuhačom, da ne zagori ručak, ili da mješaš dok te branitelji Vukovara karaju...“ – komentar korisnika A. R., kao odgovor korisnici S. B. koja ima negativan stav prema inicijativi (članak: *ibid.*)

Ipak, dosta tih komentara ipak nosi sa sobom implicitni stav o temi. Primjeri takvih komentara:

*„Sve te koji zagovaraju *pederluk* neka svoju djecu daju na čuvanje (babysitting) tim i takozvanim LGBT osobama (neka izrazito traže takvu osobu za čuvanje djece)...“* – komentar

korisnika f.m. na članak <http://danas.net.hr/hrvatska/djecu-bez-roditeljske-skrbi-koriste-kao-argument-za-borbu-protiv-istospolnih-zajednica>

„Tko sa kime i zašto, baš me briga, to je pravo svakoga od nas da bira i mislim da se nitko nema pravo u to mješati!!!Mene osobno se to ne tiče i imam važnijih problema za rješavati u životu!!!“ – komentar korisnika A. M. na članak:

<http://danas.net.hr/hrvatska/jadranka-kosor-objavila-zanimljivu-poruku-kako-bi-bilo-da-u-ustav-ugradimo>

Zahvaljujući općem znanju, možemo zaključiti da je prvi komentar usmjeren ka prihvaćanju stavova inicijative (implicirano je da homoseksualne osobe ne bi trebale odgajati djecu), a drugi komentar protiv stavova inicijative (implicirano je da je pitanje koje su postavili zagovornici inicijative nevažno i da bi se svi trebali baviti drugim stvarima, a ne definicijom braka). Automatskom detekcijom sentimenata ne možemo doći do ovakvih rezultata. Sustavi za automatsku detekciju sentimenata, mogli bi jedino detektirati ukupni sentiment komentara.

Kako je naglašavano i u drugim radovima (npr. Kennedy, 2012), sustavi za automatsku detekciju sentimenata imaju problema i s detekcijom jezičnih i stilskih figura kao što su ironija, sarkazam, humor i slično. To zasad mogu detektirati samo ljudski označivači koji imaju dovoljno jezičnog (ali i općeg) znanja. Primjeri takvih komentara:

„a daj ti ljepotice stavi svoju sliku...sigurno si prelijepa...zsa zsa gabor...“ - komentar korisnika D. O. kao odgovor na komentar korisnice N. J. na članak: <http://danas.net.hr/hrvatska/je-li-sprajc-u-dnevniku-uvrijedio-gradjane>

„D. O. pravi si frajer.“ – komentar korisnice D. M. kao odgovor na gornji komentar korisnika D. O. (članak: *ibid.*)

Ako nam je poznata fraza „ružan kao gabor“, u prvom je komentaru očigledno da korisnik ne misli na mađarsku glumicu, već isključivo na značenje njenog prezimena u hrvatskom jeziku. Jednako tako, uz pomoć jezičnog znanja, možemo zaključiti da je i drugi komentar također sarkastičan.

5.2.2. Analiza polarnosti

Kada je riječ o analizi polarnosti, osim na probleme zabilježene u literaturi (npr. Klenner i sur., 2009), nailazimo na problem komentara koji ne sadrže polarne riječi, fraze ili rečenice, no nose jasan stav (sentiment) prema temi. U našem uzorku gotovo 50% komentara nema jasne polarne odsječke teksta (riječi, fraze ili rečenice). Primjeri takvih komentara:

„Jučer navečer sam napisao da će zbog straha od gubitaka sponzorskih ugovora uslijediti demantiji, isprike... I evo ga: "Ne dijelim ljude prema boji kože, socijalnom statusu, ne osuđujem one različitih seksualnih orijentacija, vjera..." Pravi primjer kako lova i strah od gubitka iste može čovjeka učiniti tolerantnijim hahahaha“ - komentar korisnika Y. na članak: <http://danas.net.hr/hrvatska/nakon-svega-oglasila-se-i-blanka-vlasic>

„bog je stvorio adama i evu ne adama i stevu“ – komentar korisnika Z. J. na članak: <http://www.bitno.net/vijesti/hrvatska/agresivni-napadi-na-volontere-inicijative-u-ime-obitelji-jedan-stand-zapaljen-volonterka-ozlijedena/>

Iz prvog je komentara očito da se korisnik protivi inicijativi i izjavama Blanke Vlašić. U drugom je pak komentaru implicirano kako se dva muškarca ne bi trebala vjenčati (ili biti zajedno?), jer je to neprirodno. Drugi komentator vjerojatno podupire ideje inicijative.

Nadalje, iz nekih je komentara jasno da polarnost često nadilazi granicu jedne riječi ili fraze, zbog čega ju je teže detektirati sustavima za detekciju polarnosti. Primjeri takvih komentara:

„Nemojte mi samo vi talibani kršćanski moralizirati, veliki vjernici sa 3 ljubavnice i 6 vanbračne djece koji usput malo šamaraju doma ženu a ponekad i djete.. Od njih treba zaštititi ljude Ustavom..“ – komentar korisnika v. na članak: <http://www.jutarnji.hr/inicijativa--u-ime-obitelji--danas-u-povorci-dolazi-u-sabor-predati-potpise/1108564/>

„Ljudi se rađaju gay i čistoća njihove duše nije upitna. Upitna je čistoća duše, a i inteligencija osoba koje ovakve stvari izjavljuju. Sram vas može biti.“ – komentar korisnika C. R. na članak:

<http://www.bitno.net/vijesti/hrvatska/zagreb-fizicki-napadnuta-volonterka-gradanske-inicijative-u-ime-obitelji/>

U prvom je komentaru jasno da korisnik proziva vjernike na licemjerje. Moglo bi se reći da je čitav odsječak: „veliki vjernici sa 3 ljubavnice i 6 vanbračne djece koji usput malo šamaraju doma ženu a ponekad i djete..“ nositelj polarnosti. U drugom primjeru jasno je da je čitava rečenica „Upitna je čistoća duše, a i inteligencija osoba koje ovakve stvari izjavljuju.“ nositelj polarnosti jer korisnik ljude koji „čistoću duše“ homoseksualnih osoba smatraju upitnom naziva ljudima „upitne čistoće duše“ te „upitne inteligencije“. Ovdje je dodatni problem to što se pridjev „upitan“ ne nalazi kraj imenice „inteligencija“, zbog čega je tu frazu sustavima za detekciju polarnosti veoma teško detektirati kao negativnu.

5.2.3. Analiza korištenog jezika

Kako je već ranije naglašeno, gotovo polovica korisnika na portalima piše komentare nekom varijantom nestandardnog jezika, što je također problem u automatskoj detekciji sentimenata jer sustav jednostavno ne prepoznaje nestandardne riječi. Na primjer, za hrvatski jezik imamo nacionalni korpus HNK koji je znatnim dijelom morfosintaktički označen i koji bi mogao služiti pri procesiranju tekstova, no taj korpus se sastoji od tekstova pisanih standardnim jezikom (razne novine i časopisi, knjige i slično). Primjeri za to:

„Ma da pa to nema veze da je razveden, da je varao ženu, ponapravio vanbračnu djecu a potpisuje kao ovo kao nekakva obiteljska nekakva lol moralna vertikala LOL jbt“ – komentar korisnika L. R. na članak: <http://www.jutarnji.hr/-podupirem-prirodni-brak-između-zene-i-muskarca--igor-stimac-protiv-gay-brakova-/1102645/>

„Christopher B. Stuchly netko koji iza svog komentara ne stoji svojim imenom i prezimenom je obična suga a ne čovik“ – komentar korisnika J. G. na članak:

<http://www.jutarnji.hr/pankerice-razbile-stand-crkvenih-aktivista-koji-su-protiv-gay-brakova/1103259/>

„i sad Vi meni recite što će se po pitanju braka u našim životima promijeniti, da li će brakovi postati zbog ustavne promjene bolji, lipši s vječnim poštovanjem i ljubavi, hoćemo li zbog toga biti bolji prema našoj duci, odgovorniji prema obitelji itd...kozmetički ćemo dobiti zadovoljstvo u suštini sve će biti kao i do sada možda će mo samo produbiti mržnju???“ – komentar korisnika E. B. na članak: <http://danas.net.hr/hrvatska/iz-inicijative-u-ime-obitelji-prikupili-vise-od-700-tisuca-potpisa>

„danas cjeli dan razmišljam baš o ovome o čemu vi sad pišete.....i to mi je jako drago.....PA KOGA ĆE ONI GEY roditi....kome će vjeru prenositi....koje će kršćanske vrednote promovirati.....čije će krizme i pričesti slaviti....pa o čemu mi pričamo....NEK IDU BRE DA SE LEĆE ,BOG IM NA POMOĆ.....I NEKA SVOJ STATUS NAZOVU KOJIM GOD HOĆE IMENOM A NEKA SE U BRAK NE DIRAJU....I TO U IME ISUSOVO....BBL“ - komentar korisnice S. B. na članak: <http://www.bitno.net/vjera/apsurdni-referendumski-zahtjev-u-ime-obitelji/>

„Svaki dan si posvjestim koliko mrzim pedere lezbače i ostale transvestite. Zamislite se u kinderbetu, onako bespomoćni i mali, a iznad vas slinavo bradati roditelji Štef i Zdenko. Ma FUUUUUJ. Zagušio bih sam sebe blenkijem.“ – komentar korisnika M. D. na članak: <http://danas.net.hr/hrvatska/nakon-svega-oglasila-se-i-blanka-vlasic>

Kako se može vidjeti u prethodno navedenim komentarima, više je elemenata korištenog jezika koji se mogu pokazati problematičnima pri automatskoj detekciji sentimenata i polarnosti: korištenje skraćenica (hrv. „jbt“ umjesto „jebote“ i „BBL“ umjesto „Božji blagoslov“ te eng. LOL umjesto „*laughing out loud*“ ili „smijem se naglas“), izbjegavanje dijakritika („obicna suga, a ne covik“ umjesto „obična šuga, a ne čovik“), miješanje č/ć i ije/je („hoćemo“ umjesto „hoćemo“, „kršćanske“ umjesto „kršćanske“, „cjeli“ umjesto „cijeli“ itd.), korištenje anglizama i germanizama (npr. kinderbet, blenki), korištenje dijalektalnih obilježja poput ikavice i ekavice i slično. Također, u komentarima su učestali argumenti *ad hominem* poput: „Ti si govno!“ ili „Kozo jedna...“, kao i razna etiketiranja poput: homofob, katoliban, četnik, klerofašist i slično.

6. RASPRAVA I ZAKLJUČCI

Kako smo mogli vidjeti u rezultatima, ova je prilično kompleksna tema na hrvatskim portalima iznjedrila mnogo komentara. Prema iskazanim sentimentima i upotrijebljenim polarnim frazama, ti su komentari bili mahom negativni – korisnici se ili nisu slagali s inicijativom ili su se s inicijativom slagali pa su se prepirali s neistomišljenicima. Pritom su, očekivano, na katoličkom portalu bitno.net komentatori bili uglavnom pozitivniji i blaži nego na informativnim portalima net.hr i jutarnji.hr.

Na portalima net.hr i jutarnji.hr uočen je i zanimljiv trend kretanja sentimentata – što je sentiment prema inicijativi pozitivniji, pozitivniji je i ukupni sentiment komentara. Na portalu bitno.net ovakav trend nije zamijećen. Premda bi se na prvi pogled na temelju ovih podataka moglo zaključiti da su podupiratelji inicijative generalno pristojniji i pozitivniji, razlika bi se dala objasniti činjenicom da su prikupljeni komentari pisani u rasponu od početka prikupljanja potpisa za održavanje referenduma pa sve do trenutka kad je dovoljan broj potpisa prikupljen. Budući da je cilj inicijative time bio ispunjen, logično je da su podupiratelji inicijative izražavali pozitivne sentimente, dok su istovremeno protivnici inicijative izražavali negativne sentimente.

Također, pri analiziranju iskazanih sentimentata i korištenog jezika uočili smo da je korišteni jezik komentara standardniji što je iskazani ukupni sentiment komentara pozitivniji. Iako je razlika prilično mala, statistički je značajna. Ovaj se trend može objasniti činjenicom da većina vulgarizama pripada uličnom govoru, a ne standardnom jeziku pa čak i ako je komentator u ostatku komentara koristio standardni jezik, nekoliko iskorištenih vulgarizama sprječavaju označivače u tome da navedeni komentar označe oznakom 1 (standardni jezik).

Kada je riječ o poteškoćama pri analizi sentimentata i polarnih frazi naveli već u kvalitativnoj analizi. Zato ćemo se ovdje osvrnuti na jedan problem koji se nije direktno odnosio na obrađena potpoglavlja, a koji se javlja u računalnom, ali i ljudskom označavanju sentimentata. Taj problem predstavljaju tzv. „lajkovi“ (ili oznake „svida mi se“) na komentarima. Na „lajkove“ se također može gledati kao na vrstu sentimenta jer broj „lajkova“ na nekom komentaru u praksi znači broj ljudi

koji se slažu s tim komentaram. Tako je primjerice generalni sentiment čitatelja prema temi članka ispod kojeg je većina komentara negativna, no s malo lajkova, dok s druge strane mali broj pozitivnih komentara ima mnogo lajkova, zapravo generalno pozitivan, što se ne da zaključiti promatranjem isključivo komentara.

Nadalje, kad je riječ o korištenom jeziku, vidjeli smo da je jezik komentatora tek u 50% komentara standardan, dok se preostalih 50% odnosi na korištenje nestandardnih jezičnih varijanti. Pritom smo istaknuli problem brojnih vulgarizama i etiketiranja komentatora. Bubaš (2001) i Schroeder (2011) su u svojim radovima naveli upravo ovakav problem vrijeđanja i verbalne agresije zbog zanemarivanja društvenih normi kao jedan od glavnih problema računalno posredovane komunikacije. Analiziranjem komentara na portalima, uočili smo da korisnici vrlo često umjesto za argumentacijom pri izražavanju nekog svog stava, posežu za vrijeđanjem. Štoviše, za vrijeđanjem posežu čak i kad ne iskazuju svoj stav o temi, već kad druge komentatore „žele staviti na njihovo mjesto“. Navedeni problem ponašanja komentatora na portalima u fokus javnosti dolazi nakon smrti jedne hrvatske glumice¹⁰ jer čak ni moderatori nisu uspijevali brisati sve uvrede zbog brzine kojom su se oni javljali. Zbog toga se čak i jedan analizirani portal (net.hr) odlučuje na drastičan potez – potpuno ukidaju komentare¹¹. Kao što možemo vidjeti danas, navedeni portal ponovno je vratio mogućnost komentiranja članaka, no ispod članaka uglavnom nema komentara ili su oni prilično općeniti, što (osim na dobar rad moderatora) ukazuje i na to da su komentatori očito „naučili lekciju“. Možda sljedeći korak bude argumentirana rasprava umjesto vrijeđanja, no to nam tek preostaje vidjeti i analizirati u nekim budućim istraživanjima komentara na internetskim portalima.

¹⁰ Detaljnije: <http://www.netokracija.com/komentari-portali-zabrana-58646>

¹¹ <http://danas.net.hr/hrvatska/dosta-je-gadosti-ukinuli-smo-komentiranje-tekstova>

7. LITERATURA

Agić, Ž. i Merkler, D. (2012) Rule-Based Sentiment Analysis in Narrow Domain: Detecting Sentiment in Daily Horoscopes Using Sentiscope. U S. Bandyopadhyay i M. Okumura (ur.), *Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012)*. Rad predstavljen na 24th International Conference on Computational Linguistics, Mumbai, prosinac 2012. (str. 115–124). Mumbai: The COLING 2012 Organizing Committee

Agić, Ž. i sur. (2010) Towards Sentiment Analysis of Financial Texts in Croatian. U N. Calzolari i sur. (ur.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Rad predstavljen na Seventh International Conference on Language Resources and Evaluation (LREC'10), Valetta, svibanj 2010. (str.1164-1167). Valetta: European Language Resources Association (ELRA)

Adopta (21. 5. 2014.). Svaka obitelj je bolja od doma! URL:

<http://danas.net.hr/hrvatska/djecu-bez-roditeljske-skrbi-koriste-kaio-argument-za-borbu-protiv-istospolnih-zajednica> (23. 9. 2014.)

Ahmad i sur. (2011) What is new? News media, General Elections, Sentiment, and named entities. U S. Bandyopadhyay i sur. (ur.), *Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*. Rad predstavljen na Fourth International Joint Conference on Natural Language Processing, Chiang Mai, studeni 2011. (str. 80-88). Chiang Mai: Asian Federation of Natural Language Processing

Albert, C. i sur. (2011) Introducing Argumentation in Opinion Analysis: Language and Reasoning Challenges. U S. Bandyopadhyay i sur. (ur.), *Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*. Rad predstavljen na Fourth International Joint Conference on Natural Language Processing, Chiang Mai, studeni 2011. (str. 28-34). Chiang Mai: Asian Federation of Natural Language Processing

Balahur, A. i sur. (2011) Detecting Implicit Expressions of Sentiment in Text Based on Commonsense Knowledge. U *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Rad

predstavljen na The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, lipanj 2011. (str. 53-60). Portland: Association for Computational Linguistics.

Biberović, M. (23. 10. 2013.). Vrijeme da zabranimo komentare na portalima. Znae zašto. URL: <http://www.netokracija.com/komentari-portali-zabrana-58646> (23. 9. 2014.)

Bitno.net (15. 5. 2013.). Agresivni napadi na volontere inicijative ‘U ime obitelji’ – jedan štand zapaljen, volonterka ozlijeđena. URL: <http://www.bitno.net/vijesti/hrvatska/agresivni-napadi-na-volontere-inicijative-u-ime-obitelji-jedan-stand-zapaljen-volonterka-ozlijedena/> (23. 9. 2014.)

Bitno.net (20. 5. 2013.). Zagreb: Fizički napadnuta volonterka Građanske inicijative ‘U ime obitelji’. URL: <http://www.bitno.net/vijesti/hrvatska/zagreb-fizicki-napadnuta-volonterka-gradanske-inicijative-u-ime-obitelji/> (23. 9. 2014.)

Bodrožić, I. (17. 5. 2013.) Apsurdni referendumski zahtjev u ime obitelji. URL: <http://www.bitno.net/vjera/apsurdni-referendumski-zahtjev-u-ime-obitelji/> (23. 9. 2014.)

Bonev i sur. (2012) Opinum: statistical sentiment analysis for opinion classification. U *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Rad predstavljen na The 50th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, lipanj 2012. (str. 29-27). Portland: Association for Computational Linguistics.

Bubaš, G. (2001). Computer mediated communication theories and phenomena: Factors that influence collaboration over the Internet. Preuzeto s: <http://cuc.carnet.hr/cuc2001/papers/b1.pdf> (10. 8. 2014.)

Castells, M. (2005) The Network Society: from Knowledge to Policy. U Castells, M. i Cardoso, G. (ur.) *The Network Society: from Knowledge to Policy*. Washington, DC: Johns Hopkins Center for Transatlantic Relations.

Devčić, K. (28. 5. 2013.). Hrvatski psiholozi, psihoterapeuti i pravnici: 'Pitanje braka nije nešto što se definira Ustavom'. URL: <http://www.jutarnji.hr/-pitanje-braka-nije-nesto-sto-se-definira-ustavom-/1105243/> (23. 9. 2013.)

Duric, A. i Song, F. (2011) Feature Selection for Sentiment Analysis Based on Content and Syntax Models. U *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Rad predstavljen na The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, lipanj 2011. (str. 96-103). Portland: Association for Computational Linguistics.

Feldman, R. (2013) Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 56 (4): 82-89.

Glavaš i sur. (2012). Experiments on Hybrid Corpus-Based Sentiment Lexicon Acquisition. U *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data (Hybrid 2012)*. Rad predstavljen na 12th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, travanj 2012, (str. 1-9). Avignon: Association for Computational Linguistics.

Glavaš i sur. (2013). Aspect-Oriented Opinion Mining from User Reviews in Croatian. U *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing*. Rad predstavljen na pages 4th Biennial International Workshop on Balto-Slavic Natural Language Processing , Sofija, kolovoz 2013 (str. 18-23). Sofija: Association for Computational Linguistics.

Hina (14. 5. 2013.). 'PODUPIREM PRIRODNI BRAK IZMEĐU ŽENE I MUŠKARCA!' Igor Štimac potpisao peticiju protiv homoseksualnih brakova. URL: <http://www.jutarnji.hr/-podupirem-prirodni-brak-izmedu-zene-i-muskarca--igor-stimac-protiv-gay-brakova-/1102645/> (23. 9. 2014.)

Kennedy, H. (2012) Perspectives on Sentiment Analysis. *Journal of Broadcasting & Electronic Media*, 56 (4): 435-450.

Klenner, M. i sur. (2009). Robust Compositional Polarity Classification. U G. Angelova i sur. (ur.), *International Conference RANLP 2009 Proceedings*. Rad predstavljen na International Conference RANLP 2009, Borovets, rujan 2009. (str. 180-184). Shoumen: INCOMA Ltd.

Korljan, Z. (17. 5. 2013.) Pankerice razbile štand crkvenih aktivista udruge 'U ime obitelji' koji su protiv gay brakova! URL: <http://www.jutarnji.hr/pankerice-razbile-stand-crkvenih-aktivista-koji-su-protiv-gay-brakova/1103259/> (23. 9. 2014.)

Lukić, S. (8. 6. 2013.). MLADI I SOLIDNO OBRAZOVANI, POSVEĆENI VJERNICI I USPJEŠNI BIZNISMENI Tko su ljudi na čelu neokonzervativne revolucije u Hrvatskoj. URL: <http://www.jutarnji.hr/tko-su-ljudi-na-celu-neokonzervativne-revolucije-u-hrvatskoj--drustvene-vrijednosti--i-duse-mladih/1107371/> (23. 9. 2014.)

Mihaljević, M. (2005). Hrvatskim jezikom o internetu – hrvatski jezik na internetu. U Stolac, D. i sur. (ur.) *Jezik u društvenoj interakciji. Zbornik radova sa savjetovanja održanoga 16. i 17. svibnja u Opatiji*. Zagreb – Rijeka: Hrvatsko društvo za primijenjenu lingvistiku.

Milioni, D. L. i Vadratsikas, K. (2011). 'Their two cents worth': A content analysis of online readers' comments in mainstream news outlets. Preuzeto s: http://wapor.org/wp-content/uploads/2011/09/Milioni_Vadratsikas_Papa.pdf (5. 9. 2014.)

Nemeth-Jajić, J. i Milinović, A. (2012) Hrvatski jezik na internetskim forumima. *Jezik: časopis za kulturu hrvatskoga književnoga jezika*, 59 (2): 41-53.

Net. hr (19. 5. 2013). Volonteri inicijative 'U ime obitelji' ometali birače, morala ih rastjerati policija! URL: <http://danas.net.hr/hrvatska/i-policija-intervenirala-volonteri-iz-u-ime-obitelji-ometaju-izbore> (23. 9. 2014.)

Net.hr (27. 5. 2013.). Je li Šprajc u Dnevniku uvrijedio građane? URL: <http://danas.net.hr/hrvatska/je-li-sprajc-u-dnevniku-uvrijedio-gradjane> (23. 9. 2014.)

Net.hr (28. 5. 2013.). Jadranka Kosor objavila zanimljivu poruku: 'Kako bi bilo da u Ustav ugradimo...' URL: <http://danas.net.hr/hrvatska/jadranka-kosor-objavila-zanimljivu-poruku-kako-bi-bilo-da-u-ustav-ugradimo> (23. 9. 2014.)

Net.hr (29. 5. 2013.). Inicijativa 'U ime obitelji' prikupila više od 710 tisuća potpisa. <http://danas.net.hr/hrvatska/iz-inicijative-u-ime-obitelji-prikupili-vise-od-700-tisuca-potpisa> (23. 9. 2014.)

Net. hr (4. 11. 2013.). Ukinili smo komentiranje tekstova: Dosta je gadosti. URL: <http://danas.net.hr/hrvatska/dosta-je-gadosti-ukinuli-smo-komentiranje-tekstova> (23. 9. 2014.)

Pang, B. i sur. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. U *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Rad predstavljen na Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, srpanj 2002 (str. 79-86). Philadelphia: Association for Computational Linguistics.

Radić, V. (9. 11. 2013.) ODLUKA UREDNIŠTVA Jutarnji donira oglasni prostor građanima ugroženima referendumom. URL: <http://www.jutarnji.hr/odluka-urednistva-jutarnji-donira-oglasni-prostor-gradanima-ugrozenima-referendumom/1138602/> (23. 9. 2014.)

Schroeder, R. (2011) *Being There Together. Social Interaction in Shared Virtual Environments*. New York: Oxford University Press, Inc.

Sharma, R. i Bhattacharyya, P. (2013) Detecting Domain Dedicated Polar Words. *ACL Anthology*. U *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Rad predstavljen na Sixth International Joint Conference on Natural Language Processing, Nagoya, listopad 2013 (str. 661-666).

Szécsi, G. (2013). Language, Media and Community in the Information Age. *Santalka: Filozofija, Komunikacija*, 21 (2): 119-127

Vlašić, B. (14. 6. 2013.). UZ BALONE U SABOR STIGLO 749.316 POTPISA Reporteri Jutarnjeg ispratili veselu povorku 'U ime obitelji'. URL:

<http://www.jutarnji.hr/inicijativa--u-ime-obitelji--danas-u-povorci-dolazi-u-sabor-predati-potpise/1108564/> (23. 9. 2014.)

VLM agencija (22. 5. 2013.). Nakon svega oglasila se i Blanka Vlašić. URL: <http://danas.net.hr/hrvatska/nakon-svega-oglasila-se-i-blanka-vlasic> (23. 9. 2014.)

8. SAŽETAK

U rujnu 2013. godine prikupili smo korisničke komentare s tri hrvatska internetska portala (jutarnji.hr, net.hr i bitno.net) na temu prikupljanja potpisa za referendum o braku. Komentari su nastali u periodu od 12. do 26. svibnja 2013. godine. Na tim komentarima obavili smo nekoliko računalnolingvističkih analiza – analizu sentimenata, analizu polarnih frazi i analizu jezika korištenog u komentarima. Analize su pokazale da su komentari na portalima mahom negativni i prema sentimentu i prema polarnim frazama (najpozitivniji su na katoličkom portalu bitno.net), a na portalima jutarnji.hr i net.hr statistički značajnima su se pokazale razlike između ukupnog sentimenta komentara i sentimenta prema inicijativi (što je sentiment prema inicijativi pozitivniji, pozitivniji je i ukupni sentiment komentara). Također, statistički značajnom se pokazala i razlika između ukupnog sentimenta i jezika korištenog u komentarima (što je sentiment prema inicijativi pozitivniji, jezik komentatora je standardniji). Kada je riječ o korištenom jeziku, oko 50% svih komentara pisano je nestandardnim jezikom, s mnogo vulgarizama.

Ključne riječi: korisnički komentari na internetskim portalima, analiza sentimenata, analiza polarnih frazi, analiza korištenog jezika, računalno posredovana komunikacija

In September 2013 we collected users' comments on three Croatian news websites (jutarnji.hr, net.hr i bitno.net), made on articles about collecting signatures for a marriage referendum in Croatia. Comments were made between 12 May 2013 and 26 May 2013. We then conducted several computational linguistics analyses (sentiment analysis, polarity analysis and language analysis) on those comments. Results showed that comments on websites are mostly negative, with mostly negative sentiments and many negative polar words, phrases and sentences (the most positive website is Catholic website bitno.net). When it comes to quantitative analysis of sentiments and language, we found that there are statistically significant differences between general sentiment in comments and sentiment towards the initiative (the more positive sentiment towards the initiative, the more positive general sentiment) and between general sentiment and language used in comments (the more positive general sentiment, the more standard the language). When it comes to language analysis, we found that around 50% of all comments was written in non-standard language variety, with many vulgarisms.

Keywords: users' comments on web portals, sentiment analysis, polarity analysis, language analysis, computer-mediated communication