

Identifikacija online imidža organizacija temeljem analize sentimenata korisnički generiranog sadržaja na hrvatskim portalima

Hrvoje Jakopović*
Nives Mikelić Preradović**

SAŽETAK

Izgradnja pozitivnog imidža organizacije predstavlja središnju aktivnost odnosa s javnošću koja se često ostvaruje kroz odnose s medijima. Evaluacija i mjerenje učinaka u medijima danas predstavljaju važna pitanja za područje odnosa s javnošću te se pozitivan ili negativan imidž može proučavati kao ishod aktivnosti odnosa s javnošću. Budući da internet predstavlja sve značajnije sredstvo komuniciranja, temeljni je cilj ovog istraživanja bio identificirati online imidž organizacija Croatia Airlines, Hrvatske Željeznice i Zagrebački električni tramvaj na internetskim portalima analizom korisnički generiranog sadržaja, tj. analizom komentara korisnika portala koji se manifestiraju kao pozitivan, negativan ili neutralan online imidž tih organizacija. U radu se ispituje metoda računalne analize sentimenta na komentarima korisnika dva news portala - 24sata.hr i Index.hr. Usporedno s referentnom ljudskom analizom sentimenta se ispitala i preciznost računalne analize sentimenta za hrvatski jezik. Računalna analiza sentimenta se u istraživanjima za engleski jezik već pokazala kao koristan alat za profesije kao što su odnosi s javnošću. U skladu s time su ispitani sentimenti korisnika hrvatskih news portala koji su komentirali medijske objave o organizacijama iz domene putničkog prijevoza - Croatia Airlines, ZET i HŽ.

* Hrvoje Jakopović, poslijedoktorand, Odsjek za novinarstvo i odnose s javnošću, Fakultet političkih znanosti Sveučilišta u Zagrebu, Lepušićeva 6, 10000 Zagreb, +385917392225, hrvoje.jakopovic@fpzg.hr

** Nives Mikelić Preradović, izvanredna profesorica, Odsjek za informacijske i komunikacijske znanosti, Filozofski fakultet Sveučilišta u Zagrebu, I. Lučića 3, 10000 Zagreb, +38514092323, nmikelic@ffzg.hr

Temeljem analize sentimenta i frekvencije riječi u korisničkim komentarima izrađen je računalni leksikon riječi obilježenih sentimentom za domenu prijevoza. Također, u radu su identificirani ključni izazovi s kojima se suočava analiza sentimenta za hrvatski jezik, ali i za druge morfološki bogate slavenske jezike, temeljem usporedbe, korelacijske analize i mjerenjem relativne pogreške računalne analize sentimenta u odnosu na referentnu ljudsku analizu.

Rezultati istraživanja ukazuju na potencijal korištenja analize sentimenta za identificiranje pozitivnih, negativnih i neutralnih emocija prema određenom objektu, u ovome slučaju organizacijama koje se bave transportom putnika. Zaključak rada je da su transportne organizacije među korisnicima hrvatskih news portala obilježene dominantno negativnim sentimentom, tj. imaju negativan online imidž.

Ključne riječi: analiza sentimenta, korisnički komentari, korisnički generiran sadržaj, online vijesti, prijevoz, putnici, odnosi s javnošću

Uvod

Krajnji dosezi medija očituju se u promjenama mišljenja, stavova i, konačno, ponašanja medijskih publika. S obzirom na preobrazbu medija u formu interaktivnih platformi za dvosmjernu komunikaciju, korisnički generiran sadržaj (eng. *user-generated content* - UGC) otvorio je pristup promišljanjima brojnih korisnika te time postao predmet istraživanja i proučavanja različitih društvenih fenomena. Komentari korisnika na novinskim (eng. *news*) portalima, osim recenzija, objava na društvenim mrežama i blogovima, čine posebnu kategoriju jer su direktno povezani s popularnim medijima, odnosno onima koji predstavljaju „glavnu struju“ (eng. *mainstream*). U tom pogledu, istraživanja komentara na novinskim portalima podrazumijevaju različite navike korisnika. Primjerice, Coe i dr. (2014: 673) u svome radu analiziraju neuljudnost korisnika na novinskim portalima. Njihovo istraživanje pokazuje da je u prosjeku jedan od pet komentara neuljudan, a 55.5 % komentara uključuje neku vrstu neuljudnosti, dok se među vrstama neuljudnosti najčešće pronalazi verbalno zlostavljanje (eng. *name-calling*). Huston (2011: 89) zaključuje da bi medijske kuće koje produciraju internetske / online vijesti trebale imati na umu kako komentari korisnika također mogu pridonijeti kreiranju mišljenja o određenoj temi i u tom smislu ostaviti dojam pristranosti jer vijesti koju uređuju urednici i novinari ne funkcioniraju izolirano u online okruženju. Doprinos i značaj komentara, u smislu povećanja raznolikosti diskursa na novinskim portalima, naglašavaju Baden i Springer (2014: 545). Utjecaj komentara na kreiranje mišljenja je i u fokusu studije Hsueh i dr. (2015: 9) koja otkriva kako u online okruženju predrasude iščitane u komentarima drugih korisnika potiču pristranost osobnih komentara.

U ovome se radu ispituju mogućnosti implementacije računalne analize sentimenta na hrvatskom jeziku za procjenu mišljenja korisnika prema organizacijama iz domene putničkog prijevoza kao što su Croatia Airlines, ZET i HŽ. Domena putničkog prijevoza izabrana je zbog činjenice da ispitivanje sentimenta korisnika u toj domeni nailazi na osobito plodno tlo zbog čestih oscilacija i kriznih situacija u prometu. Analiza sentimenta pri računalnoj obradi nailazi na određene prepreke, a u ovom radu je poseban naglasak stavljen na hrvatski jezik i sentiment. Pritom se ispitivanje računalne analize sentimenta smatra korisnim indikatorom i za druge slavenske jezike. Prepreke računalne analize sentimenta javljaju se kod otkrivanja značenja pojedinih izraza i rečeničnih konstrukcija jer računalo ne prepoznaje višeznačnost i metaforična značenja koja su specifična za ljudsku komunikaciju. Glucksberg ističe: "Metafore dopuštaju interpretacije koje se sudaraju s doslovnim značenjima, stoga riječi moraju biti doslovno shvaćene kako ne bi izgubile značenje" (2011: 1). Računala stoga ne mogu biti sasvim precizna u određivanju značenja, no razrada i obogaćivanje leksikona koji se implementiraju u programe za analizu sentimenta smanjuju mogućnost pogrešne interpretacije poruka. Prema Schwartz i dr. (2013), analiza jezika daje podatke koji su kvantificirani i koji dovode do zaključaka o ponašanju korisnika. Objave korisnika na društvenim medijima stoga su pogodne za takve vrste analiza koje obuhvaćaju veliki broj komentara. Osim toga, može se pretpostaviti kako su korisnici slobodniji u izražavanju svojih misli i osjećaja putem društvenih medija nego što bi to bili prilikom anketiranja ili intervjuiranja.

Korisnička aktivnost i online vijesti

Navike korisnika društvenih medija vrlo su promjenjive, a ponajviše ovise o tehničkim postavkama koje uvode ili predstavljaju korporacije poput Facebooka (koji su nedavno uveli novi način izražavanja sentimenta pomoću emojija). Promjene su vidljive i u domeni *news* portala pa je sve većem broju korisnika privlačno uključiti se u komentiranje medijskih objava. Iz tog razloga Kaplan i Haenlein (2010) definiraju društvene medije kao grupu internetskih aplikacija sagrađenih na ideološkim i tehnološkim temeljima Weba 2.0, koji dopuštaju kreiranje i razmjenu korisnički generiranog sadržaja. Može se reći i da su društveni mediji splet tehnologije i ljudske interakcije kroz koji se stvaraju određene vrijednosti. Kietzmann i dr. (2011: 243) objašnjavaju društvene medije kroz sedam funkcionalnih blokova:

1. Identitet – mjera u kojoj se korisnik otkriva
2. Prisutnost – mjera u kojoj korisnici znaju jesu li drugi korisnici dostupni
3. Odnos – mjera u kojoj se korisnici povezuju s drugima
4. Dijeljenje – mjera u kojoj korisnici razmjenjuju, distribuiraju i primaju sadržaj

5. Konverzacija – mjera u kojoj korisnici komuniciraju jedni s drugima
6. Skupine – mjera u kojoj korisnici formiraju online zajednice
7. Reputacija – mjera u kojoj korisnici znaju društvenu poziciju drugih i dostupnog sadržaja

Od spomenutih funkcionalnih blokova, profesija odnosi s javnošću (PR) dotiče se svih, a osobito posljednjeg - „reputacije“. Reputacija se tako često promatra kao ishod aktivnosti odnosa s javnošću (Grunig, 1993, Coombs, 2005), a unapređuje se uz pomoć oblikovanja prepoznatljivog identiteta, prisutnosti, razvijanja odnosa i komuniciranja. Informacijsko komunikacijska tehnologija (IKT) je postavila tehničke uvjete za aktivno sudjelovanje korisnika, što je prema Edwards (2009: 548) dosad bilo najteže postići u kampanjama odnosa s javnošću. U takvom okruženju PR više ne traži posrednika u medijima. Kako tvrdi Meerman Scott: „Internet je učinio odnose s javnošću opet javnima, nakon godina gotovo isključivog fokusa na medije. Blogovi, *on-line* priopćenja i drugi oblici Web sadržaja dopuštaju organizacijama izravno komuniciranje s kupcima“ (2007: 11). S tim se slaže i Broom (2010: 249), prema kojemu je jedna od najvećih prednosti interneta kao medija za odnose s javnošću što “osigurava izravnu i trenutnu dostupnost konkretnih ciljnih javnosti, čime se zaobilazi novinare i tradicionalne medije”. Korisničko uključivanje u prvom redu podrazumijeva njihovo izlaganje objavljenom sadržaju. Ksiazek i dr. (2014) napominju kako se primarni razlozi za uključivanje i interaktivnost mogu objasniti kroz teoriju koristi i zadovoljstva pa tako motive čine: 1. Potraga za informacijama, 2. Socijalizacija ili socijalna interakcija i 3. Zabava. Na tragu toga moguće je razlikovati interaktivnost na relaciji korisnik-sadržaj i korisnik-korisnik, gdje je prvo potaknuto potrebom za informiranjem, a drugo potrebom za socijalnom interakcijom. Longitudinalna studija Lee i dr. (2014: 519) ukazuje kako klikovi (eng. *clicks*) korisnika utječu na pozicioniranje vijesti na internetskim portalima, zaključujući da se snaga učinaka klikova pojačava tijekom dana, dok nema značajnog ukupnog učinka naknadnog pozicioniranja vijesti na broj klikova, stoga je učinak klikova korisnika pri početnom pozicioniranju vijesti značajniji nego što ta kasnije pozicionirana vijest ima učinka na povećanje broja klikova korisnika. Bobkowski (2015: 15) naglašava kako su konzumenti online vijesti danas postali digitalni dostavljači koji koriste društvene medije kako bi ih distribuirali među svojim online kontaktima. Istraživanje je pokazalo kako u prosjeku korisnici dijele vijesti koje prema njihovom viđenju sadržavaju određenu informacijsku korisnost, dok vođe javnog mišljenja (eng. *opinion leaders*) dijele vijesti bez obzira na informacijsku korisnost, a drugi korisnici ih najčešće (zbog toga što dolaze od vođa javnog mišljenja) opažaju kao informacijski korisne (Bobkowski, 2015: 15). Tenenboim i Cohen (2015: 2012-2013) ispitali su u kojoj su mjeri povezani klikovi i komentari korisnika sa sadržajem online vijesti s obzirom na visok broj pogleda/klikova i visok broj ko-

mentara kod određenih vijesti. Njihovi rezultati potvrđuju da se među najkomentiranim vrstama nalaze vijesti koje uključuju politička i društvena pitanja, ali su istovremeno među najposjećenijima one online vijesti koje se bave kriminalnim radnjama. Zhou i Moy (2007: 92-93) u svojoj su studiji demonstrirali potencijal online javnog mišljenja u smislu doprinosa javnom diskursu, budući da korisnici uvode raznolike „interpretacijske okvire“ (eng. *frames*) dodajući značenje i vrijednost određenoj temi te na taj način transformiraju događaj u temu koja privlači pažnju izvan okvira virtualnog prostora. Kako se objavljuje više informacija o određenoj temi, tako se intenzivira i određeni dominantni okvir oko predmeta koji je izložen komentarima. Istovremeno, što je više korisnika uključeno u komunikaciju oko određene teme, suma njihovih mišljenja producirat će bolju evaluaciju proizvoda, usluge, organizacije ili osobe.

Materijal i metode

U ovom radu ispitana je mogućnost prilagodbe programa za analizu sentimenta hrvatskom jeziku te način na koji se suvremeni alati, poput računalne analize sentimenta, mogu primijeniti za ispitivanje korisnički generiranog sadržaja. Na temelju analize korisnički generiranog sadržaja sastavljen je leksikon riječi obilježenih sentimentom za specifičnu domenu istraživanja. Postupak vrednovanja komentara korisnika na internetskim portalima se sastoji od referentne ručne i računalne analize i polariteta pozitivno/negativno te neutralno. Uz to se ispituje korelacija na primjeru organizacije Croatia Airlines i relativna pogreška u slučajevima Zagrebačkog električnog tramvaja (ZET-a), Hrvatskih željeznica (HŽ-a) i Croatia Airline-sa (CA).

Istraživanjem se pokušava odgovoriti na dva istraživačka pitanja:

- 1:** Kakav je sentiment korisnika news portala prema organizacijama Croatia Airlines, Hrvatske željeznice i Zagrebački električni tramvaj?
- 2:** U kojoj mjeri je računalna analiza sentimenta precizna u odnosu na ručnu analizu sentimenta?

Na početku istraživanja postavljene su dvije polazne hipoteze:

H1 – Sentiment prema organizacijama Croatia Airlines, Hrvatske željeznice i Zagrebački električni tramvaj među korisnicima internetskih portala 24sata.hr i Index.hr je dominantno negativan

H2 – Računalna analiza sentimenta ostvaruje preciznost veću od 50 % u odnosu na referentnu ljudsku analizu sentimenta za korisnički generiran sadržaj na temu transporta

Uzorak

Ukupno je analizirano 3068 (N=3068) korisničkih komentara. Najviše je komentara analizirano u medijskim objavama o Zagrebačkom električnom tramvaju (N=1266), zatim o Croatia Airlinesu (N=1135) i Hrvatskim željeznicama (N=667). Jedinicu analize sentimenta predstavlja jedan komentar korisnika koji pripada medijskoj objavi (članku) o Croatia Airlinesu, Hrvatskim željeznicama i Zagrebačkom električnom tramvaju na internetskim portalima *24sata.hr* i *Index.hr*. Razdoblje analize medijskih objava (članaka) na internetskim portalima *24sata.hr* i *Index.hr* i pripadajućih komentara korisnika obuhvaća period od dvije kalendarske godine (od 1. siječnja 2013. do 31. prosinca 2014.). Prikupljanje medijskih objava i pripadajućih korisničkih komentara o organizacijama ZET, HŽ i CA, za razdoblje analize, provedeno je putem pretraživanja arhive internetskih portala *24sata.hr* i *Index.hr* prema ključnim riječima – nazivima triju hrvatskih organizacija iz domene transporta. Izbor jednostavnog slučajnog uzorka medijskih objava napravljen je uz pomoć računalnog generiranja slučajnih brojeva putem besplatnog programa *Research Randomizer* (<http://www.randomizer.org/>, 2015). Zastupljenost korisničkih komentara po medijskoj objavi kretala se u rasponu od najmanje 1 do najviše 89 komentara. Analizom su obuhvaćeni samo oni komentari koji se odnose na spomenute organizacije, a pritom su se iz analize izuzimali drugi komentari istih korisnika u istoj medijskoj objavi.

Internetski je portal *24sata.hr* odabran na temelju mjesečnog dosega od 1.165 709 jedinstvenih korisnika (gemiusAudience, veljača/2015). Prema istraživanju Gemiusa (gemiusAudience, veljača/2015), *24sata.hr* ima doseg 48,17 %, a to ga čini najposjećenijim internetskim portalom u Hrvatskoj, što se odražava u broju komentara korisnika. *24sata* su ujedno i najčitanije dnevne novine u Hrvatskoj (De facto, listopad – prosinac, 2014) koje prema dnevnom prosjeku imaju 458.838 čitatelja. S druge strane, *Index.hr* jedan je od popularnijih internetskih portala s dosegom od 39,79 % korisnika (gemiusAudience, siječanj/2014). *Index.hr* je prema posljednjem mjerenju Gemiusa imao mjesečno 992.336 jedinstvenih korisnika (gemiusAudience, siječanj/2014). Za razliku od najpopularnijih novinskih portala, *Index.hr* djeluje kao samostalni *news* portal bez dodatnog medijskog formata i karakterističan je po brojnosti korisničkih komentara. *Index.hr* je tijekom 2014. godine istupio iz istraživanja Gemius (<http://www.index.hr>, 2015).

Metoda

Analiza sentimenta (eng. *sentiment analysis*) još se naziva i dubinsko pretraživanje mišljenja (eng. *opinion mining*). Podrazumijeva računalnu obradu prirodnog jezika

kojom se, analitičkim pristupom, može dobiti uvid u stavove i mišljenja pojedinaca. Naime, sentiment se u ovom kontekstu interpretira kao osjećaj, mišljenje, stav o pojedinom proizvodu, usluzi, osobi ili nekom drugom predmetu analize. Pang i Lee (2008) ističu kako je u procesu donošenja odluke oduvijek bilo važno imati informaciju što drugi ljudi misle o predmetu odluke. Wilson i dr. (2005) naglašavaju da je zadatak analize sentimenta identificiranje pozitivnih i negativnih mišljenja, emocija i procjena. Pang i Lee (2004) definiraju analizu sentimenta kao računalnu metodu koja služi identificiranju određenih gledišta unutar pojedinog teksta. Sentiment može predstavljati subjektivno viđenje koje je samim time potpomognuto osjećajima, ali i objektivno čime se isto tako izražava stav. Prema tome, sentiment se najčešće iskazuje kroz suprotstavljene pojmove kao što su pozitivan/negativan te mogućnost neizjašnjavanja kroz sentiment koji ujedno može biti identificiran kao neutralan. No, ovisno o odabiru instrumenta koji se koristi za analizu, raspon mogućih odgovora može biti i veći. Primjerice, to mogu biti ljestvice u kojima se intenzitet pozitivnog i negativnog sentimenta mjeri rasponom ocjena od 1 do 5. Jurafsky i dr. (2014) su uz pomoć analize sentimenta za engleski jezik analizirali recenzije korisnika o restoranima na internetskoj stranici *Yelp.com*. Analiza sentimenta korisnika je pokazala kako su korisnici ukusnu hranu u skupim restoranima opisivali senzualnim metaforama, dok se dobra hrana u jeftinim restoranima povezivala s ovisnošću (Jurafsky i dr., 2014). Time dolazimo do zaključka da mišljenja, odnosno komentari korisnika, mnogo govore i o njima samima.

Budući da hrvatski jezik pripada jezicima s malim brojem računalnih jezičnih resursa (eng. *low-resourced language*), a istovremeno je i morfološki složen jezik, još uvijek ne postoji javno dostupan računalni program za analizu sentimenta teksta na hrvatskom jeziku. Glavaš i dr. (2012) su izradili sentiment leksikon *CroSentiLex* za hrvatski jezik (<http://takelab.fer.hr/sentilex>, 2015) koji se sastoji od 1200 lema (s pridruženim pozitivnim i negativnim polaritetom, ali bez intenziteta) koje su ručno označene od strane 12 analitičara, a s obzirom na stupanj njihovog slaganja razlikuje se zlatni skup¹ riječi s niskim slaganjem većine analitičara (eng. *low agreement set*), zlatni skup sa srednjim slaganjem barem osam analitičara (eng. *medium agreement set*) i zlatni skup s visokim slaganjem (eng. *high agreement set*) barem deset analitičara (Glavaš i dr., 2012). Doprinos razradi i upotrebi analize sentimenta za hrvatski jezik također su dali Merkler i Agić (2013) implementacijom programa *Sentiscope* za detekciju sentimenta u dnevnom horoskopu na *news* portalima.

S obzirom na gore navedenu prepreku automatskoj analizi sentimenta, za računalnu analizu sentimenta komentara objavljenih na hrvatskom jeziku odabran je program *SentiStrength* (<http://sentistrength.wlv.ac.uk/>, 2015). *SentiStrength* je besplatan program s otvorenim pristupom za akademska istraživanja, dostupan na engleskom jeziku. Riječ je o programu koji omogućuje analizu različitog tekstualnog sadržaja

pisanog engleskim jezikom, no posjeduje mogućnost prilagodbe za analizu teksta na različitim prirodnim jezicima, uključujući i hrvatski. Sentiment se u programu iskazuje kroz polaritete pozitivno / negativno, odnosno polariteti su opisani kao pozitivne i negativne emocije. Ocjena tekstualnog sadržaja se prikazuje kroz pozitivnu ljestvicu od 1 do 5 (1 – neutralno, 5 – izrazito pozitivna emocija) i negativnu ljestvicu od -1 do -5 (-1 – neutralno, -5 – izrazito negativna emocija). Program se sastoji od tablice emotikona (*EmoticonLookupTable*), tablice slenga (*SlangLookupTable*), popisa engleskih riječi (*EnglishWordList*), popisa negacijskih riječi (*NegatingWordList*), tablice idioma (*IdiomLookupTable*), popisa riječi koje pojačavaju značenje (*BoosterWordList*) i upitnih riječi (*QuestionWords*). *SentiStrength* provodi analizu nakon što se unese tekst u prozor programa te određuje sentiment za svaki uneseni tekst posebno.

Budući da je hrvatski jezik izrazito flektivan, tj. morfološki bogat jezik, prilagodba programa hrvatskom jeziku napravljena je prevodenjem više od 2.500 engleskih riječi programa *SentiStrength* i ispitivanjem njihove višeznačnosti putem Hrvatskog leksikona (<http://www.hrleksikon.info/>, 2015) koji je nastao na Leksikografskom zavodu “Miroslav Krleža”. Kako bi se leksikon dodatno obogatio riječima obilježenim sentimentom, pridodan je i leksikon s gotovo 6.000 hrvatskih riječi koje je obradilo pet studenata/analičara Filozofskog fakulteta Sveučilišta u Zagrebu. Zbog usklađivanja intenziteta pozitivnih, negativnih i neutralnih riječi, pet analitičara koji su ih obilježavali kroz raspon od -2 do 2 (-2 – izrazito negativno; -1 – negativno; 0-neutralno; 1-pozitivno, 2-izrazito pozitivno) s ljestvicom *SentiStrength* (od -5 do 5), tražila se aritmetička sredina, kao mjera centralne tendencije, odnosno prosjek za svaku pojedinu riječ koju su analitičari obilježili sentimentom. S obzirom na aritmetičku sredinu koja se kretala za svaku riječ od -2 do 2 kod procjene analitičara Filozofskog fakulteta Sveučilišta u Zagrebu, izrađena je ljestvica u rasponu od -5 do 5 (koja odgovovara ljestvici *SentiStrength*) na temelju sljedećih zastupljenih vrijednosti:

1 → 0.2 i 0.4;	-1 → -0.2 i -0.4
2 → 0.6 i 0.8;	-2 → -0.6 i -0.8
3 → 1 i 1.2;	-3 → -1 i -1.2
4 → 1.4 i 1.6;	-4 → -1.4 i -1.6
5 → 1.8 i 2;	-5 → -1.8 i -2

Nakon spajanja leksikona, izbačene su riječi koje se ponavljaju temeljem slaganja dvaju analitičara oko sentimenta kojim su obilježene. Na taj način dobiven je leksikon od 7.615 hrvatski riječi obilježenih sentimentom koji je ubačen u program *SentiStrength*.

Nadalje, komentari korisnika pripadajućih medijskih objava su analizirani ručno i uzeti kao referentni u odnosu na računalnu analizu sentimenta zbog nedostataka koji

su uočeni pri računalnoj analizi sentimenta kod mnogih autora (Glucksberg, 2011; Liu, 2012; Pang i Lee, 2008). Ručna analiza je provedena uz pomoć dva obučena nezavisna analitičara. Holstijev test pouzdanosti (1969) je proveden na 100 nasumično odabranih komentara iz odabrane populacije te je postignut prosječan rezultat 0,97. Neusklađenosti među analitičarima su se riješile putem rasprave i dogovora među analitičarima. Ručna i računalna analiza sentimenta se temeljila na određivanju pozitivnih, negativnih i neutralnih komentara.

Relativno odstupanje računalne procjene od ljudske analize sentimenta izračunato je uzimajući ljudsku analizu sentimenta kao pravu ili referentnu vrijednost. Broj ljudskih istoznačnih procjena sentimenta je prava ili referentna vrijednost k_{lj} , a razlika između broja ljudskih i računalnih istoznačnih sentimenta je apsolutno odstupanje $\Delta k = k_{lj} - k_r$ od referentne vrijednosti. Relativno je odstupanje računalnih istoznačnih procjena komentara:

$$r = \frac{\Delta k}{k_{lj}} 100\%.$$

S obzirom da je skup procjena komentara zbroj pozitivnih, negativnih i neutralnih procjena komentara, taj skup je označen s $k_u = k_p + k_n + k_{ne}$. Njihovo je zajedničko apsolutno odstupanje računalnih procjena od ljudskih procjena komentara kao referentne vrijednosti: $\Delta k_u = \Delta k_p + \Delta k_n + \Delta k_{ne}$. Stoga je zajedničko relativno odstupanje računalnih procjena od referentnih procjena komentara:

$$r_z = \frac{\Delta k_p + \Delta k_n + \Delta k_{ne}}{k_p + k_n + k_{ne}} 100\%.$$

Rezultati

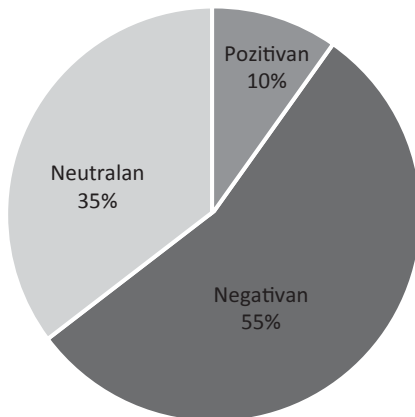
Analizom oba hrvatska portala *24sata.hr* i *Index.hr* utvrđeno je da ne postoje značajnije razlike u iskazanom sentimentu korisnika prema tri hrvatske organizacije iz domene transporta (ZET, CA i HŽ) pa su rezultati u skladu s time prikazani skupno. Iz komentara korisnika (Tablica 1) proizlazi da je imidž svih triju organizacija po

Tablica 1. Online imidž organizacija po medijskoj objavi

Table 1. Online image of organizations according to users' comments

Online imidž organizacije po medijskoj objavi (N=310)	Pozitivan	Negativan	Neutralan
Croatia Airlines, N=102	9 %	55 %	36 %
Zagrebački električni tramvaj, N=102	9 %	54 %	37 %
Hrvatske željeznice, N=106	3 %	58 %	39 %

**Ukupna zastupljenost komentara u ljudskoj
analizi sentimenta (N=3068)**



Slika 1. Ukupna zastupljenost komentara u ljudskoj analizi sentimenta
Figure 1. The total representation of comments in human sentiment analysis

Tablica 2. Online imidž organizacija prema ljudskoj analizi sentimenta
Table 2. Online image of organizations according to human sentiment analysis

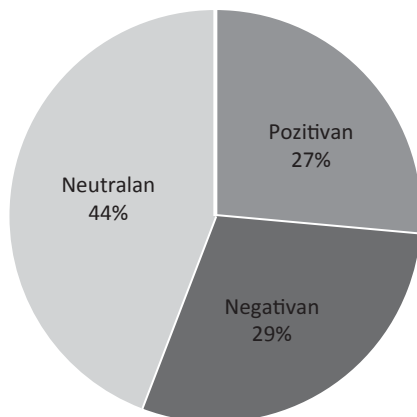
Online imidž organizacija prema komentarima (N=3068)	Pozitivan	Negativan	Neutralan
Croatia Airlines, N=1135	13 %	55 %	32 %
Hrvatske željeznice, N=667	3 %	70 %	27 %
Zagrebački električni tramvaj, N=1266	12 %	47 %	41 %

medijskoj objavi dominantno negativan. Najnegativniji je imidž HŽ-a (58 %), potom Croatia Airlinesa (55 %) i zatim ZET-a (54 %). Online imidž po medijskoj objavi jest jedinica analize koja je određena polaritetima pozitivno-negativno i neutralno, a čini ju dominantni sentiment iz sume korisničkih komentara za svaku pojedinačnu medijsku objavu (N=310).

Komentari korisnika u medijskim objavama o organizacijama CA, HŽ i ZET na internetskim portalima *24sata.hr* i *Index.hr* istraženi su ljudskom analizom sentimenta kako bi se utvrdio odgovarajući online imidž navedenih organizacija.

Slika 1 ukazuje na prevladavajuću zastupljenost negativnih komentara u medijskim objavama o CA, HŽ-u i ZET-u (55 %), zatim slijede neutralni komentari (35 %), a najmanja je zastupljenost pozitivnih komentara (10 %). Jedinica analize u ovome je

**Ukupna zastupljenost komentara u računalnoj
analizi sentimenta (N=3068)**



Slika 2. Ukupna zastupljenost komentara u računalnoj analizi sentimenta
Figure 1. The total representation of comments in computer sentiment analysis

slučaju jedan korisnički komentar koji izražava sentiment prema hrvatskim organizacijama iz domene transporta. Stoga je ukupan broj korisničkih komentara iskazan kao N=3068.

S obzirom na dominantan broj negativnih komentara, Tablica 2 prikazuje kako sve tri organizacije (CA, HŽ i ZET) s obzirom na komentare imaju negativan imidž među korisnicima na internetskim portalima *24sata.hr* i *Index.hr*. U tom smislu je najnegativniji online imidž *Hrvatskih željeznica* (70 %), potom slijedi *Croatia Airlines* (55 %) i *Zagrebački električni tramvaj* (47 %). Najveći broj pozitivnih komentara ima CA (13 %), zatim ZET (12 %), a samo 3 % pozitivnih komentara je usmjereno prema HŽ-u. Neutralni komentari su u najvećoj mjeri zabilježeni kod ZET-a (41 %), CA (32 %) i HŽ-a (27 %).

Odnos između negativnih, pozitivnih i neutralnih komentara u računalnoj analizi sentimenta (*SentiStrength*) pokazuje značajnija odstupanja u odnosu na ljudsku analizu sentimenta (Slika 2). Naime, program *SentiStrength* zabilježio je najviše neutralnih komentara (44 %), potom negativnih (29 %) i pozitivnih komentara (27 %). Različiti autori (Glucksberg, 2011; Liu, 2012; Pang i Lee, 2008; Pang i Lee 2004) uočavaju kako je računalna analiza sentimenta još uvijek suočena s nedostacima koji se očituju u neprepoznavanju metafora, ironije, sarkazma i općenito šireg konteksta u koji je smješten sadržaj koji se vrednuje kroz polaritete pozitivno, negativno ili neutralno. Pokazuju to i primjeri analize sentimenta komentara korisnika u

slučaju izvještavanja o CA, HŽ-u i ZET-u, gdje postoji neslaganje između ljudske analize i računalne analize sentimenta:

- „Pa ti leti sa hrvatskim avionom“ (ljudska procjena – negativno; *SentiStrength* – neutralno)
- „Čudo da se dosad nije srušio nijedan“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „Trebalo biti hrabar letjeti s CA.“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „Šta ti je hrvatski inženjering :-)“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „HŽ savršen za odlazak na izlete. Izadete iz vlaka, prošetate, pa opet uđete u vlak u vožnji.“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „Bez obzira na vašu točnost sa nama kasnite 100% HŽ!!!!“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „Dragi ZET! Smanjite cijene karata i ovakvi i slični slučajevi će otić u prošlost. Hvala!“ (ljudska procjena – negativno; *SentiStrength* – pozitivno)
- „Zet je uvijek spreman za sve : („ (ljudska procjena – negativno; *SentiStrength* – pozitivno)

Slaganja u procjenama sentimenta ljudske i računalne analize sentimenta iskazana su u sljedećim primjerima:

- „Neka, neka...govno jedno šofersko!“ (ljudska procjena – negativno; *SentiStrength* – negativno)
- „Zatvorite Firmu ! Nemojte nas stalno sramotiti po svijetu.“ (ljudska procjena – negativno; *SentiStrength* – negativno)
- „Njima je očito fetish gubiti prtljagu. Moju su izgubili dva puta.“ (ljudska procjena – negativno; *SentiStrength* – negativno)
- „ZET je kao Splitski Promet, magični miks najgoreg socijalizma i najgoreg kapitalizma, socijalizam kad treba parazitirati na grbači naroda cuclajući porez i prirez, kapitalizam kad tom istom narodu naplate najskuplje cijene karata javnog prijevoza na svijetu“ (ljudska procjena – negativno; *SentiStrength* – negativno)
- „Odavno ne putujem s HŽ, redovno kašnjenje, vlak u snijegu, pad napona, pretprani smrdljivi, hladni vagoni a masu puta crkne vlak. Stoje pol sata da prođe teretni vlak. To ima samo kod nas. Zaista imam razumijevanja pa putujem busom.“ (ljudska procjena – negativno; *SentiStrength* – negativno)
- „Hvala dragom Bogu da nikome nije ništa, ipak je bio vozač jako smiren i sposoban. Bravo čovječe“ (ljudska procjena – pozitivno; *SentiStrength* – pozitivno)

- „Živio HŽ!!!!“ (ljudska procjena – pozitivno; *SentiStrength* – pozitivno)
- „Bravo za pilote, takve si uvijek poželim“ (ljudska procjena – pozitivno; *SentiStrength* – pozitivno)
- „Ma svaka čast cure , ja znam dvije isto koje ne voze u zet-u, voze manje autobuse kod svojih očeva !!“ (ljudska procjena – pozitivno; *SentiStrength* – pozitivno)
- „Baš su ostavili dojam simpatičnih i skromnih ljudi, a uz to još i sposobnih, smirenih, koncentriranih. CA može s pravom biti ponosna na ovakve zaposlenike.“ (ljudska procjena – pozitivno; *SentiStrength* – pozitivno)

Na temelju frekvencije riječi u komentarima korisnika medijskih objava o Croatia Airlinesu, Hrvatskim željeznicama i Zagrebačkom električnom tramvaju, definirao se leksikon riječi za domenu prijevoza (Tablica 3). S obzirom na specifičan kontekst izvještavanja hrvatskih internetskih portala, koji je prepoznat kroz teme o kojima

Tablica 3. Leksikon najzastupljenijih riječi obilježenih sentimentom za domenu prijevoza

Table 3. *Lexicon of the most frequent sentiment words in the transport domain*

Pozitivno	Negativno	Neutralno
dobro	problem	karte
sigurno	krivi	tramvaj
bolje	otkaz	vrijeme
bravo	sramota	plaće
super	protiv	putnik
besplatno	gubitke	avion
čast	previše	vozač
red	gore	vlakovi
dragi	loše	prijevoz
čudo	kvar	piloti
čestitke	pada	bus
sreća	štrajk	cijene
jeftinije	nažalost	uprava
uredno	bahati	firma
prijatelj	strašno	kontrolor
stručnjaci	kasni	pruge
podrška	nesreća	država
sposobni	gluposti	novac
pametni	teško	kompanija
znanje	uhljebi	promet

se izvještava, izvore informacija, tipove naslova te sveobuhvatne medijske okvire, zatim kontekst u kojem navedene hrvatske organizacije posluju i obilježja korisničkih komentara (pozitivnih, negativnih i neutralnih), može se uočiti kako se, osim riječi koje su posve jasno povezane s domenom prijevoza, pojavljuju i one koje upućuju na kontekst u kojemu se nalaze organizacije. Tako se među najzastupljenijim neutralnim riječima često spominje „država“, a među negativnima „uhljebi“, što upućuje na širi društveni kontekst koji proizlazi iz korisničkih komentara.

Korelacijskom analizom se vrednovala preciznost računalne analize sentimenta u odnosu na referentnu ljudsku analizu sentimenta na primjeru Croatia Airlinesa. Pearsonov koeficijent korelacije je u području srednje pozitivne povezanosti (Jackson, 2009), $r=0,312$, uz signifikantnost $p<0,01$ (Tablica 4). Pearsonov koeficijent korelacije se prema Jackson (2009: 67) dijeli na: slabu povezanost – od ± 0.00 do 0.29 ; srednju povezanost – od ± 0.30 do 0.69 te snažnu povezanost od ± 0.70 do 1.00 .

Tablica 4. Korelacija ljudske analize sentimenta i računalne analize sentimenta (CA)
 Table 4. Correlation of human sentiment analysis and computer analysis of sentiment (CA)

Pearsonova korelacija (N=1135)	Ljudska analiza sentimenta za CA	
	koeficijent	signifikantnost
Računalna analiza sentimenta za CA	0,312	p<0,001

Relativno odstupanje računalne analize od ljudske analize sentimenta računalo se u skladu s formulom iznesenom u planu obrade podataka te je, prema tome, relativno odstupanje računalne procjene od referentnih procjena komentara za Croatia Airlines: $r_z = 43\%$.

Rasprava

Ljudska analiza sentimenta je pokazala da su korisnički komentari dominantno negativno usmjereni prema organizacijama CA, HŽ i ZET, što je u skladu s početnom hipotezom H1. S obzirom na komentare, Hrvatske željeznice imaju najnegativniji imidž među korisnicima na internetskim portalima, dok je imidž Croatia Airlinesa najpozitivniji, a imidž ZET-a najneutralniji. U usporedbi s računalnom analizom sentimenta, provedenom uz pomoć programa *SentiStrength* koji je modificiran i prilagođen hrvatskom jeziku, vidljiva su značajnija odstupanja u odnosima pozitivnih, negativnih i neutralnih komentara. Računalna analiza sentimenta je prepoznala najviše neutralnih komentara o organizacijama CA, HŽ i ZET pa tek onda negativne i, naposljetku, pozitivne komentare. Može se reći kako je veći broj pozitivnih komen-

tara u računalnoj analizi, u odnosu na ljudsku analizu sentimenta, posljedica nemoćnosti prepoznavanja ironije i sarkazma u komentarima korisnika. No, usprkos tome, korelacijska analiza referentne ljudske analize i računalne analize sentimenta na primjeru Croatia Airlinesa je pokazala korelaciju u području srednje pozitivne povezanosti $r=0,312$ uz signifikantnost na razini 1 %. Iz toga se može zaključiti kako računalna analiza sentimenta programa *SentiStrength* prilagođenog hrvatskom jeziku pokazuje srednju razinu preciznosti u odnosu na referentnu analizu, ali uz perspektivu da se i dalje radi na unapređenju programa. Primjerice, Thelwall (2012) pokazuje kako preciznost *SentiStrengtha* u odnosu na referentnu analizu sentimenta varira ovisno o formi teksta koji se analizira (*BBC forum, My Space, Twitter, YouTube*) pa se tako najviše korelacije kreću do $r=0,647$, a najniže do $r=0,286$ za engleski jezik. Rezultati *SentiStrengtha* prilagođenog hrvatskom jeziku osobito su značajni kad se uzme u obzir morfološka složenost hrvatskog jezika (tj. sedam padeža hrvatskog jezika za jedninu i za množinu imenica). Potom je prisutan i problem višeznačnosti vezan uz dijakritičke znakove koji mnogi korisnici ne upotrebljavaju prilikom pisanja online komentara, kao u primjeru:

„Postoje diskretniji nacini da se covjek ubije. Ovako traumatiziraju vozace vlaka koji nisu ni krivi ni duzni.“ (*Index.hr*, 28.5.2014.)

Posljedično, postoji mogućnost da program ne raspoznaje riječi obilježene sentimentom, iako *SentiStrength* nudi i mogućnost uključivanja nepravilnog pisanja (eng. *include incorrect spellings*). Posljednja prepreka je i specifičan kontekst komentara korisnika na internetskim portalima. Naime, njihov komentar je svojevrsna reakcija na izvještavanje u medijskoj objavi koje može biti usmjereno na više priča kao dio medijskog događaja pa izraženo mišljenje korisnika o organizaciji te stoga može obuhvatiti puno širi kontekst nego što je potrebno. Za razliku od komentiranja vijesti, komentari kao što su recenzije proizvoda i usluga usmjereniji su i direktniji te je komentiranje uglavnom fokusirano samo na predmet recenziranja.

Relativno odstupanje računalne analize od ljudske analize sentimenta prikazano je na primjeru Croatia Airlinesa te iznosi $r_z=43\%$. Thelwall (2010) je u svom radu utvrdio kako *SentiStrength* može predvidjeti pozitivan sentiment s preciznošću od 60,6 %, a negativan sentiment s preciznošću od 72,8 %. Uzme li se u obzir relativno odstupanje na primjeru Croatia Airlinesa, može se reći da je preciznost programa *SentiStrength* prilagođenog hrvatskom jeziku 57 %, čime je potvrđena hipoteza (H2) - računalna analiza sentimenta ostvaruje preciznost veću od 50 % u odnosu na referentnu ljudsku analizu sentimenta za korisnički generiran sadržaj.

Određivanjem frekvencije riječi u korisničkim komentarima se utvrdilo da se uz kontekst hrvatskog prijevoza usko veže „država“, ali i riječi obilježene negativnim sentimentom kao što je riječ „uhljebi“, na što ukazuju sljedeći primjeri:

- „Prvo treba potjerat debele menadžere i direktore, ako su se negdje uhljebili onda je to CA! Na svakih par radnika dođe jedan debeli uhljeb, ko i u HŽu... Rastjerajte te hodajuće kante masti, nek guraju malo avione ak hoće zadržat poso!!!“ (*Index.hr*, 17.5.2013.)
- „Debele veze za poslove, debeli uhljebi...što reći...“ (*Index.hr*, 13.5.2013.)
- „Jedino koncesija može spasiti naš željeznički promet. Tu su sve uhljebi do uhljeba i nikad ništa nije uloženo u modernizaciju, neg’ je sve odlazilo uhljebljenicima, fimi-medijama, rođijacima, kumovima, političarima, ... ma muzle su se milijarde kuna godišnje pod izlikom “očuvanje” radnih mjesta. Mo’š mislit!!!!!!!“ (*24sata.hr*, 19.7.2013.)
- „Potpis. I mi, zetovski uhljebi, ne snimamo putnike kada oni pričaju na mobitel. Molio bih Upravu da kraj vozača postavi upozorenje “Ne snimati vozača dok priča na mobitel.”“ (*Index.hr*, 6.11.2014.)

Zaključak

Ispitivanje računalne analize sentimenta korisničkih komentara prema organizacijama Croatia Airlines, ZET i HŽ daje uvid u prepreke automatske analize prirodnih jezika, koje se odnose na hrvatski, ali i ostale morfološki složene slavenske jezike koji nemaju velik broj računalnih jezičnih resursa. Istraživanje daje jedinstven pregled različitih sentimenata, kao i vrijednosno obojenih riječi koje su identificirane među korisničkim komentarima za specifičnu domenu putničkog prijevoza. Upravo iz toga proizlazi i leksikon riječi koji se može koristiti u budućim programima za računalnu obradu teksta i određivanje sentimenta. Analiza sentimenta se pokazala kao perspektivan i primjenjiv pristup za vrednovanje korisničkih komentara na velikom uzorku. Usporedna analiza referentne ljudske analize sentimenta i računalne analize sentimenta ukazala je na određena odstupanja računalne analize. Računalna je analiza prepoznala najviše neutralnih komentara u odnosu na najzastupljenije negativne komentare u ljudskoj analizi, ali je također ukazala i na više pozitivnih komentara nego referentna analiza. To se može protumačiti kroz već utvrđene nedostatke računalne analize sentimenta koja ne prepoznaje negativan sentiment kroz ironiju, sarkazam, metafore, a nema ni uvid u širi kontekst koji imaju ljudi. No, korelacijska analiza potvrđuje značajan potencijal računalne analize sentimenta u identifikaciji pozitivnog, negativnog i neutralnog sentimenta među korisnicima. Iz toga razloga bi i dalje svakako trebalo proučavati i unaprjeđivati ovakav pristup analizi korisnički generiranog sadržaja. U ovom istraživanju je program *Senti-Strength* odabran kao sredstvo računalne analize sentimenta i prilagođen hrvatskom jeziku kako bi se identificirali ključni izazovi hrvatskog sentimenta među korisnicima *news* portala, a to su: 1. morfološka složenost prirodnog jezika (sedam hrvat-

skih padeža jednine i množine), 2. dijakritički znakovi (odnosno odsustvo istih u korisnički generiranom sadržaju) te 3. komentari korisnika koji se kao reakcija na medijsku objavu, osim organizacijom, bave i drugim kontekstima zastupljenima u medijskoj priči. Važnost implementacije ovakvog pristupa u analizi tekstualnog sadržaja posebno je usredotočena na profesije kao što su odnosi s javnošću i marketing kojima je u opisu aktivnosti osluškivati potrebe njihovih ciljanih javnosti. Posljedično, računalna analiza sentimenta se prepoznaje kao metoda koja ima tržišnu vrijednost. U slučaju hrvatskih organizacija koje se bave prijevozom, sentiment korisnika je indikator njihove nepovoljne reputacije među korisnicima istraženih hrvatskih portala.

ENDNOTES

- ¹ zlatni standard – eng. *gold standard* upućuje na zadovoljavajuću razinu slaganja među analitičarima
-

LITERATURA

- Baden, C. & Springer, N. (2014) “Com(ple)menting the news on the financial crisis: The contribution of news users’ commentary to the diversity of viewpoints in the public debate”, *European Journal of Communication*, 29 (5), 529-548.
- Bobkowski, P. S. (2015) “Sharing the News Effects of Informational Utility and Opinion Leadership on Online News Sharing”, *Journalism & Mass Communication Quarterly*, 92 (2), 320-345.
- Broom, G. M. (2010) *Cutlip & Center ‘s Učinkoviti odnosi s javnošću*. Zagreb: MATE.
- Coe, K., Kenski, K. & Rains, A. S. (2014) “Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments”, *Journal of Communication*, 64 (4), 658-679.
- Edwards, L. (2009) “Public relations origins: definitions and history”, 3-18. U Tench, R. & Yeomans, L.: *Exploring Public Relations*. Harlow. Pearson Education.
- Glavaš, G., Šnajder, J. & Dalbelo Bašić, B. (2012) “Semi-Supervised Acquisition of Croatian Sentiment Lexicon”, 166-173. U: *Proceedings of 15th International Conference on Text, Speech and Dialogue*. Brno. TSD 2012.
- Glucksberg, S. (2011). “Understanding Metaphors: The Paradox of Unlike Things Compared”. *Affective Computing and Sentiment Analysis*, 45, 1-12.
- Grunig, J. E. (1993) “Image and Substance: From Symbolic to Behavioral Relationships”, *Public Relations Review*, 19 (2), 121-139.

- Heath, Robert (ur.) (2005) *Encyclopedia of Public Relations*. Thousand Oaks: Sage.
- Houston, J. B., Hansen, G. J. & Nisbett, G. S. (2011) "Influence of User Comments on Perceptions of Media Bias and Third-Person Effect in Online News", *Electronic News*, 5 (2), 79-92.
- Hsueh, M., Yogeewaran, K. & Malinen, S. (2015) "Leave Your Comment Below": Can Biased Online Comments Influence Our Own Prejudicial Attitudes and Behaviors?, *Human Communication Research*, 41 (4), 557-576. doi: 10.1111/hcre.12059.
- Index.hr (2015) "*Index opet najčitaniji hrvatski portal*". <http://www.index.hr/mobile/clanak.aspx?category=vijesti&id=840507> (20.08.2016).
- Jackson, S. L. (2009) *Research Methods and Statistics: A Critical Thinking Approach*. Belmont, CA: Wadsworth.
- Jurafsky, D., Chahuneau, V., Routledge, B. R. & Smith, N (2014) "Narrative framing of consumer sentiment in online restaurant reviews", *First Monday*, 19 (4), <http://journals.uic.edu/ojs/index.php/fm/article/view/4944/3863> (12.12.2015.).
- Kaplan, A. M. & Haenlein, M. (2010) "Users of the world, unite! The challenges and opportunities of social media", *Business Horizons*, 53 (1), 59-68.
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P. & Silvestre, B. S. (2011) "Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media", *Business Horizons*, 54, 241-251.
- Ksiązek, T. B., Peer, L. & Lessard, K. (2014) "User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments", *New Media & Society*, 18 (3), 502-520.
- Lee, A. M., Lewis, S. C. & Powers, M. (2014) "Audience Clicks and News Placement A Study of Time-Lagged Influence in Online Journalism", *Communication Research*, 41 (4), 505-530.
- Agić, Ž. & Merkler, D. (2012) "Rule-Based Sentiment Analysis in Narrow Domain: Detecting Sentiment in Daily Horoscopes Using Sentscope", 115-124. S. Bandyopadhyay & M. Okumura (ur.): *Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology*. Mumbai. COLING.
- Mumbai, Indija : The COLING 2012 Organizing Committee, 2012. 115-124
- Pang, B. & Lee, L. (2004). "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics". <http://aclweb.org/anthology/P/P04/P04-1035.pdf?CFID=247293914&CFTOKEN=70536811>(12.02.2016.).
- Pang, B. & Lee, L. (2008). "Opinion Mining and Sentiment Analysis", *Information Retrieval*, 2 (1-2), 1-135.

- Scott, M. (2007) *The New Rules of Marketing and PR*. Wiley: New Jersey.
- Shwartz, A. H. et al. (2013). „Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach“, *PLoS ONE*, 8 (9), 1-16.
- Tenenboim, O. & Cohen, A. A (2015) “What prompts users to click and comment: A longitudinal study of online news”, *Journalism*, 16 (2), 198-217.
- Thelwall, M., Buckley, K. & Paltoglou, G. (2012) “Sentiment strength detection for the social Web”, *Journal of the American Society for Information Science and Technology*, 63 (1), 163-173.
- Thelwall, M., Buckley, K., Paltoglou, G. & Cai, D. (2010) “Sentiment strength detection in short informal text”, *Journal of the American Society for Information Science and Technology*, 61 (12), 2544–2558.
- Wilson, T., Wiebe, J. i Hoffmann, P. (2005) “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis”, 347-354. U: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. <http://acl.ldc.upenn.edu/H/H05/H05-1044.pdf> (15.02. 2016.).
- Zhou, Y. & Moy, P. (2007) “Parsing Framing Processes: The Interplay Between Online Public Opinion and Media Coverage”, *Journal of Communication*, 57, 79–98.

Identification of Organizations' Online Image through Sentiment Analysis of User-generated Content on Croatian Web Portals

Hrvoje Jakopović
Nives Mikelić Preradović

SUMMARY

Building a positive image of an organization represents the central activity of public relations that is often achieved through media relations. Today, the evaluation and measurement of the effects of the media represent important issues in the field of public relations (PR), while a positive or a negative image can be studied as a result of the PR activities. Since the Internet represents an increasingly important means of communication, the fundamental objective of this study was to identify the online image of the passenger transport organizations (Croatia Airlines, Croatian Railways and the Zagreb Electric Tram) on the internet portals through the analysis of user-generated content. This paper studies the method of computer sentiment analysis of user comments in two news portals - 24sata.hr and Index.hr. In parallel with the human sentiment analysis, the paper examines the accuracy of sentiment analysis of Croatian language, testing the sentiment of the users of Croatian news portals (more precisely, users' comments on media reports about the organizations in the domain of passenger transport - Croatia Airlines, Zagreb Electric Tram and Croatian Railways). Based on the sentiment analysis and word frequencies in the users' comments, a computer lexicon of sentiment words for the domain of transportation was generated. Also, this paper identifies the key challenges of the sentiment analysis for the Croatian language, but also for other morphologically rich Slavic languages, based on the comparison, correlation analysis and measurement of the relative error of computer sentiment analyzes in relation to the human sentiment analysis.

The results indicate the potential use of sentiment analysis to identify the positive, negative and neutral emotions towards organizations dealing with the passenger transport. The paper concludes that the users of Croatian news portal have the predominantly negative sentiment towards the passenger transport organizations, i.e. they have a negative online image.

Key words: sentiment analysis, user comments, user-generated content, online news, transport, passengers, public relations