

Analiza sentimenta recenzija hotela

Dalibor Bužić¹, Anita Hučić²

¹Visoka škola za informacijske tehnologije, Klaićeva 7, 10000 Zagreb (dalibor.buzic@vsite.hr)

²Visoka škola za informacijske tehnologije, Klaićeva 7, 10000 Zagreb (anita.hucic@vsite.hr)

UDK 004.4 '4

Na internetu se svakodnevno generira velik broj recenzija proizvoda odnosno usluga. Te recenzije mogu poslužiti kao vrijedan izvor informacija ponuđačima usluga o (ne)zadovoljstvu korisnika s ciljem poboljšanja ponude. Glavni problem je mukotrpnost analize recenzija. Zato se ovaj proces nastoji automatizirati za što služi analiza sentimenta – metoda primjene računalne obrade prirodnog jezika koja identificira, izdvaja i analizira subjektivne informacije u tekstu. Dva su temeljna pristupa analizi sentimenta: strojno učenje i leksikonski pristup. U ovom radu korišten je leksikonski pristup korištenjem triju rječnika: Bing, AFFIN i NRC. NRC rječnik otkriva ne samo pozitivni i negativni aspekt rečenica, već i konkretne osjećaje kao što su ljutnja, povjerenje, gađenje i iznenađenje. Za odabrani hotel u Dubrovniku napravljena je analiza sentimenta na temelju više od tisuću i petsto recenzija objavljenih na sustavu TripAdvisor. Analiza otkriva najčešće riječi i fraze te riječi koje doprinose pozitivnom i negativnom sentimentu. Razmatrane su i negacije – riječi koje mijenjaju polarnost rečenica. Rezultati istraživanja pokazuju i promjenu sentimenta tijekom vremena.

Ključne riječi: *rudarenje mišljenja, analiza sentimenta, strojno učenje, recenzije hotela*

1. Uvod

S obzirom na to da svake godine raste broj korisnika interneta, sukladno s time raste i broj korisnika koji izražavaju mišljenje na raznim platformama. Prema podacima o broju korisnika na svjetskoj razini iz travnja 2019. godine, čak 56,1% populacije svijeta ima pristup internetu i informacijama koje su ponuđene na pretraživačima (Internet World Stats, 2019), a velik dio aktivnih korisnika interneta pretražuje recenzije i ocjene korisnika. Sa stajališta kupca, provjera online recenzija smatra se neophodnima prilikom kupnje putem interneta (Schuckert, Xianwei i Rob, 2015).

Upravo iz navedenih razloga ovaj rad stavlja poseban naglasak na pretraživanje recenzija i analizu koja će pojednostaviti i ubrzati tumačenje velikog broja recenzija o određenoj temi.

Da bi recenzije i ocjene bile relevantne, potrebno se uvjeriti da je ocjenjivač iskren, da je recenzija objektivna, odnosno da nije plaćena recenzija. Uz to, bitno je dobiti pravovaljanu informaciju kojoj će korisnik koji pretražuje određen pojam – vjerovati.

Web stranice poput Booking.com i Airbnb.com tvrde da nude ocjene stvarnih korisnika. Stranice za organizaciju smještaja traže povratnu informaciju svojih posjetitelja, a to mogu i potvrditi njihovi korisnici.

Rade to iz više razloga, a jedan od njih je da navedu nove posjetitelje da posjete upravo hotele i ostale smještaje koje oni preporučuju. S druge strane, takve platforme žele dobiti povratnu informaciju da bi mogle poboljšati svoju uslugu ili maknuti one hotele koji ne zadovoljavaju kriterije prema ocjenama posjetitelja.

Analiza sentimenta je automatizirani proces kojemu je cilj sabrati relevantna mišljena korisnika te na temelju analiziranja njihovih tekstova dati povratnu informaciju odnosno ocjenu onome tko traži ta mišljenja. Konkretnije, hoteli žele brzu povratnu informaciju koji njihovi posjetitelji dijele u virtualnom svijetu, a sve u svrhu poboljšanja svojih usluga i privlačenja više posjetitelja.

2. Analiza sentimenta

Analiza sentimenta područje je koje analizira osjećaje, mišljenja, stavove i procjene ljudi prema entitetima i njihovim atributima u tekstualnim podacima (Liu, 2015). Analiza daje smisao velikoj količini podataka kojima se koriste poduzeća, u ovom slučaju – hoteli. Analiza sentimenta se još naziva i Opinion Mining, odnosno u prijevodu - rudarenje mišljenja. Ona služi za izgradnju sustava koji pokušava identificirati i izdvojiti mišljenja unutar samog teksta. Uz to, sustavi rade na principu izlučivanja atributa nekog izraza. Primjerice, izraz je moguće razdvojiti na nekoliko komponenti: polarnost, temu i nositelja mišljenja.

Polarnost pomaže odrediti je li mišljenje pozitivno ili negativno, tema je ono o čemu se govori u datoj recenziji, a nositelj mišljenja je osoba koja izražava svoje mišljenje na temu.

Upravo zbog velike količine informacija koja se može pronaći privatno i javno na blogovima, forumima i društvenim mrežama, moguće je naći pregršt tekstova koji su ustvari mišljenja zadovoljnih i nezadovoljnih korisnika. Uz pomoć analize sentimenta ove informacije mogu postati strukturirani podaci javnog mišljenja o bilo kojoj temi.

Što se tiče hotela i usluga koje pružaju svojim posjetiteljima, ova analiza omogućava brzo i jednostavno prikupljanje podataka, njihovo strukturiranje i prikaz mišljenja posjetitelja na jednom mjestu. Ovi podaci postaju bitni za povratnu informaciju o uslugama koje hotel pruža.

Primjerice, posjetitelji hotela će rado objaviti svoje mišljenje nakon nekog putovanja bez obzira bilo ono pozitivno ili negativno. Ako je mišljenje pozitivno, nositelj mišljenja će to rado podijeliti u virtualnom svijetu, ponekad fotografijom na društvenim mrežama, a ponekad komentaram na neki članak o hotelu. S druge strane, ako je mišljenje negativno, posjetitelj hotela ima često potrebu upozoriti potencijalne posjetitelje, ali i izraziti svoju frustraciju vezanu za uslugu.

Baš uz pomoć analize sentimenta hoteli će pobliže vidjeti koji dio njihove usluge treba poboljšanje. Jasno će vidjeti jesu li gosti nezadovoljni sa zaposlenicima hotela, hranom, spavaćim sobama, dodatnim uslugama hotela i slično. Isto tako, vidjet će što je ono najbitnije u njihovoj usluzi što privlači njihove posjetitelje. U (Hargreaves, 2015) preporučuje se da menadžment hotela počne pratiti negativne komentare tijekom vremena i mjeriti napredak u smanjenju negativnog sentimenta.

2.1. Tijek analize

Prvenstveno je bitno definirati mišljenje i odvojiti ga od definicije činjenice. Mišljenje je subjektivno izražavanje emocija ljudi, dok činjenica označava objektivni izraz o nečemu. Iz tog razloga, analiza sentimenta kao dio NLP-a (Natural Language Processing) mora riješiti dva problema: klasifikaciju subjektivnosti te klasifikaciju polarnosti. Dakle, analiza mora razlučiti subjektivno od objektivnog, te isto tako pozitivno od negativnog kod izražavanja mišljenja.

Prvi korak analize na primjeru rečenice "Hotel je bio odličan!" nastaje od koraka nazvanog tokenizacija, odnosno podjele rečenice po znakovima, u ovom slučaju razmacima. Na taj način dobiva se riječ po riječ nekog teksta.

Sljedeći korak je čišćenje podataka, u ovom slučaju posebnih znakova, a to je uskličnik.

Treći korak je izlučivanje tzv. Stop riječi, a to je u ovom slučaju "je" i "bio". Nakon toga ostaje "Hotel odličan".

Četvrti korak je korak klasifikacije, koji prikazuje značenje riječi kao pozitivno, negativno ili neutralno. Za pozitivan rezultat analiza može davati +1, za negativan -1, a za neutralan 0.

Ovdje se za klasifikaciju mogu koristiti rječnici ili metode strojnog učenja (primjerice metoda potpunih vektora, Naivan Bayesov klasifikator, klasifikator maksimalne entropije) (Serrano-Guerrero et al., 2015).

Posljednji korak je kalkulacija, odnosno zbrajanje pozitivne ili oduzimanje negativne ocjene. Ako je ocjena veća od 0, analizirani izraz je pozitivan.

2.2. Vrste mišljenja

Mišljenja se mogu sagledati kao izravna, komparativna, eksplicitna i implicitna.

Izravna mišljenja daju izravno mišljenje o nekom proizvodu ili usluzi, dok komparativna mišljenja uspoređuju jedan proizvod ili uslugu s drugom. Najčešće, komparativna mišljenja izražavaju sličnosti ili razlike dvaju predmeta upotrebom superlativnog oblika pridjeva ili priloga.

Eksplicitno mišljenje je ustvari izričito mišljenje izraženo u subjektivnoj rečenici, koje može biti pozitivno ili negativno (primjerice "Ručak u hotelu je svaki dan bio savršen."). S druge strane, implicitno mišljenje je mišljenje koje se podrazumijeva u objektivnoj rečenici (primjerice, "Ručnika u hotelskoj sobi nije bilo dovoljno."). Implicitna mišljenja mogu u sebi sadržavati metafore, koje su najteža vrsta mišljenja za analizu jer podrazumijevaju kombinaciju više semantičkih informacija.

3. Pristupi i izazovi u analizi sentimenta

Postoje mnoge vrste analize sentimenta, od kojih su najvažnije sljedeće:

- sitnozrnasta analiza,
- detekcija emocija,
- analiza temeljena na aspektu te
- analiza namjere.

Kod sitnozrnaste analize naglasak je na detaljnijem i preciznijem polarnosti mišljenja. Odnosno, može se koristiti opis poput "vrlo pozitivan", "vrlo negativan" i slično. Neki sustavi povezuju polarnost s nekim emocijama, kao što je ljutnja, tuga ili briga, te pozitivnim osjećajima kao što je kombinacija više osjećaja – ljubav, radost ili entuzijizam.

Detektiranje emocija pokušava otkriti emocije ljutnje, frustracije, sreće, zadovoljstva i cijeli spektar emocija unutar nekog izraza, odnosno mišljenja. Međutim, ovdje postoji jedan nedostatak jer se neke riječi negativne konotacije koriste kao pozitivna emocija. Primjerice u rečenici "Ovaj hotel je mrak." ili "Ovaj hotel ubija." značenje obje riječi je negativno, ali emocija je, ustvari, pozitivna.

Analiza aspekta ukazuje na pojedine značajke proizvoda ili usluge koja se pruža. Dakle, ova analiza stavlja naglasak na neku značajku kojom se nositelj mišljenja bavi, a od velike je važnosti za analizu.

Analiza namjere ukazuje točno što korisnik želi naglasiti svojom recenzijom, iako ne govori konkretno je li usluga bila dobra ili loša. Primjerice, mišljenje u rečenici "Čekao sam na recepciji hotela 10 minuta." je negativno, a njegova namjera je ukazati na lošu uslugu osoblja u hotelu ili možda nekvalitetnu organizaciju.

Zbog raznih ograničenja i izazova kod analize sentimenta, posljednjih nekoliko godina radi se na razvoju preciznijih klasifikatora sentimenta. Neki od izazova su: precizno razdvajanje subjektivnih od objektivnih rečenica, problem konteksta, usporedbe u rečenicama, ironija i sarkazam, emotikoni, negacije, itd.

Za što bolje razumijevanje recenzije, potrebno je obratiti posebnu pažnju na navedene izazove. Bitno je razlikovati prikriveno mišljenje poput ironije ili sarkazma od stvarnog mišljenja.

Isto tako, sve više recenzija uz običan tekst sadrži i emotikone koji su postali dio svakodnevne komunikacije. Potrebno je odrediti kako pojedini emotikon izražava mišljenje korisnika, najčešće je li to pozitivno ili negativno mišljenje. Negacije utječu na polarnost drugih riječi u rečenici. Opseg negacije može biti ograničen na sljedeću riječ nakon negacije ali se može proširiti i do udaljenijih riječi (Farook et al., 2017).

Dati preciznu i točnu ocjenu neke recenzije nije lagan zadatak za analizu sentimenta. Štoviše, moguće je da će analiza ponekad pogriješiti, ali s druge strane, koristeći ovu analizu rezultat će biti točan u 70-80% slučajeva. Uz to, korisnici analize vrlo brzo će vidjeti poboljšanje odnosno, ova analiza će im uštedjeti vrijeme i novac.

4. Povezana istraživanja

U (Stecanella, 2016) analizirano je više od milijun recenzija putnika na stranici TripAdvisor, a rezultat analize bio je uglavnom pozitivan. Analiziranje je provedeno na temelju recenzija danih samo za putovanja u velike gradove poput Londona, Pariza, New Yorka, Bangkoka, Pekinga i Madrida. Oko 82% recenzija koje su bile napisane, analiza je označila kao pozitivne recenzije. Konkretnije, London je dobio najgore ocjene odnosno recenzije i pritom je kao glavni razlog nezadovoljstva navedena loša hrana i prljavost smještaja. Daljnjom analizom dvije riječi su se često ponavljale, a to su „žohari“ u Bangkoku i „kroasani“ u Parizu. Pritom, recenzije vezane uz kroasane su bile negativne, jer se odnosilo na ponudu hrane u hotelu.

U (Bucur, 2015) analizira se učinkovitost algoritma putem izvlačenja mišljenja korisnika specifično vezanih za hotelske usluge. Rad naglašava prednost online rezervacije upravo zbog velikog broja recenzija prijašnjih korisnika. Napravljeni sustav za analizu sentimenta sastoji se od modula za prikupljanje (web robot koji posjećuje web stranicu i izvlači željeni sadržaj) i modula za analizu sadržaja. Modul za analizu sadržaja obrađuje spremljene recenzije i kreće s procesom otkrivanja mišljenja. Uključuje procesni modul, modul rudarenja mišljenja i leksičku bazu podataka SentiWordNet. Sustav je testiran na 3,000 recenzija s web stranice TripAdvisor koje su ručno podijeljene u pozitivne i negativne unutar dvije datoteke. Dakle, 1.500 recenzija je imalo oznaku pozitivnih recenzija i isti broj je imao oznaku negativnih recenzija. Preciznost u klasifikaciji na pozitivna i negativna mišljenja bila je između 72% i 76,5%.

Sustav BESAHOT kojemu je cilj omogućiti prikupljanje recenzija hotela s raznih web stranica predstavljen je u (Kasper i Vela, 2011). Ciljani korisnici su hotelijeri i trenutno se proučavaju samo recenzije u Njemačkoj na njemačkom jeziku. Radi se o interaktivnoj web aplikaciji koja radi na temelju GWT okvira. Sustav funkcionira tako da prikuplja i sprema podatke, te ih analizira. Sustav je u mogućnosti analizirati određeni tekst i raspodijeliti na više tema, ako u recenziji pronade dvije različite teme. Primjerice, ako recenzija glasi „Soba je uredna i osoblje je profesionalno“, BESAHOT uspijeva odvojiti urednost soba kao jednu temu, te uslužnost osoblja kao drugu temu.

Za analizu sentimenta hotelskih kompleksa u Las Vegasu u (Philander i Zhong, 2016) korišteni su podaci prikupljeni s društvene mreže Twitter. Za ukupno 34 hotelska kompleksa ukupno je analizom obuhvaćeno 31,550 Twitter poruka (928 po hotelu). Rječničkim pristupom utvrđeno je da kod svih hotela preteže pozitivni sentiment. Najpoznatija imena, kao što su Aria, Bellagio i Venetian, nalaze se među pet najbolje ocijenjenih kompleksa.

Drugačijim pristupom, nadgledanim strojnim učenjem, u (Yordanova i Kabakchieva, 2017) recenzije hotela (586 recenzija s Booking.com i 347 s TripAdvisor.com) klasificirane su u pozitivnu odnosno negativnu klasu. Korištena metoda stabla odlučivanja davala je vrlo dobre rezultate – točnost klasifikacije se kretala između 80.8% i 86.7%.

Problemom računalnog raspoznavanja autentičnih od izmišljenih recenzija hotela bavi se

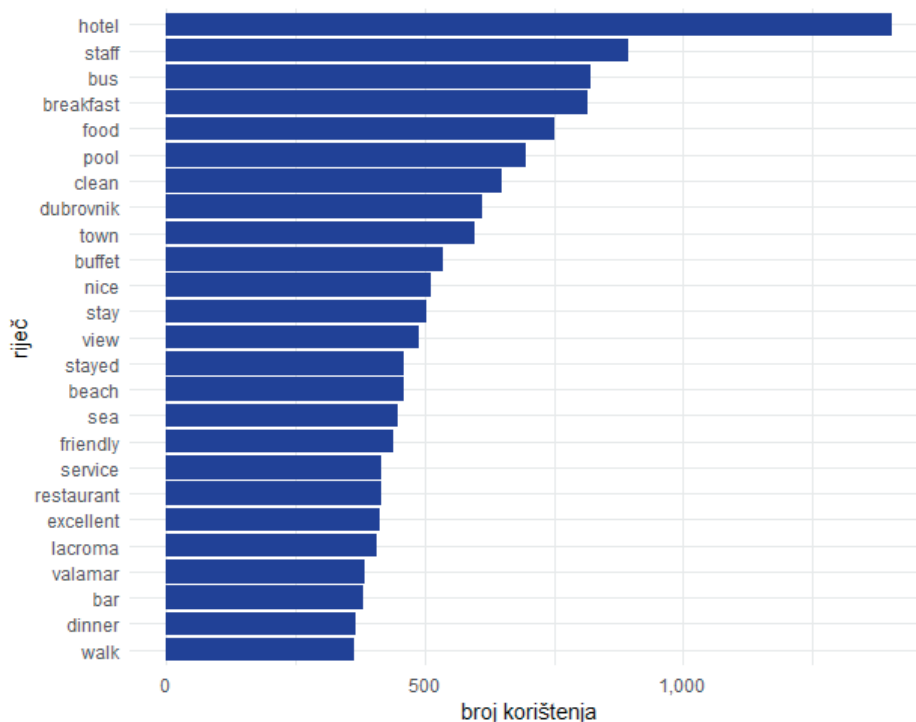
istraživanje provedeno nad 900 autentičnih i isto toliko lažnih recenzija (Banerjee i Chua, 2017). Razmatrane su četiri karakteristike – razumljivost, specifičnost, pretjerivanje i nemarnost u tekstu. Kao zanimljivost, otkriveno je da u autentičnim pozitivnim recenzijama ima više negativnih emocija nego u njihovim lažnim ekvivalentima.

5. Istraživanje

Za istraživanje odabrane su recenzije hotela Valamar Larcoma u Dubrovniku objavljene na platformi TripAdvisor. Hotel ima oko 2,300 recenzija, od čega je 1,564 na engleskom jeziku. Recenzije pisane engleskim jezikom ciljane su zbog dostupnosti rječnika sentimenata upravo na tom jeziku te su dohvaćene pomoću programskog alata Octoparse. Uz pojedinu recenziju ne stoji točan datum kad je ona napisana (izuzetak su one pisane u tekućem mjesecu, gdje postoji dan i mjesec, ali nedostaje godina) već samo mjesec i godina. Zato su, radi lakše obrade, datumi umjetno konstruirani kao da su pisani 15. u mjesecu. Najstarija recenzija je iz lipnja 2009. godine, a najnovija iz lipnja 2019. godine. Kompletna obrada podataka napravljena je u programskom okruženju R.

Slika 1. prikazuje 25 riječi s najvećim frekvencijama pojavljivanja u recenzijama. Stop riječi (riječi koje ne donose nikakvo relevantno značenje, primjerice the, and, which, at...) prethodno su filtrirane. Riječi nisu svodene na osnovni oblik, stoga su primjerice stay i stayed tretirane kao dvije različite riječi.

Slika 1: Najčešće riječi u recenzijama



Izvor: vlastito istraživanje

Tablica 1. prikazuje najčešće korištene fraze koje se sastoje od dvije uzastopne riječi. Zanimljivo da se u tri od deset fraza spominje riječ bus.

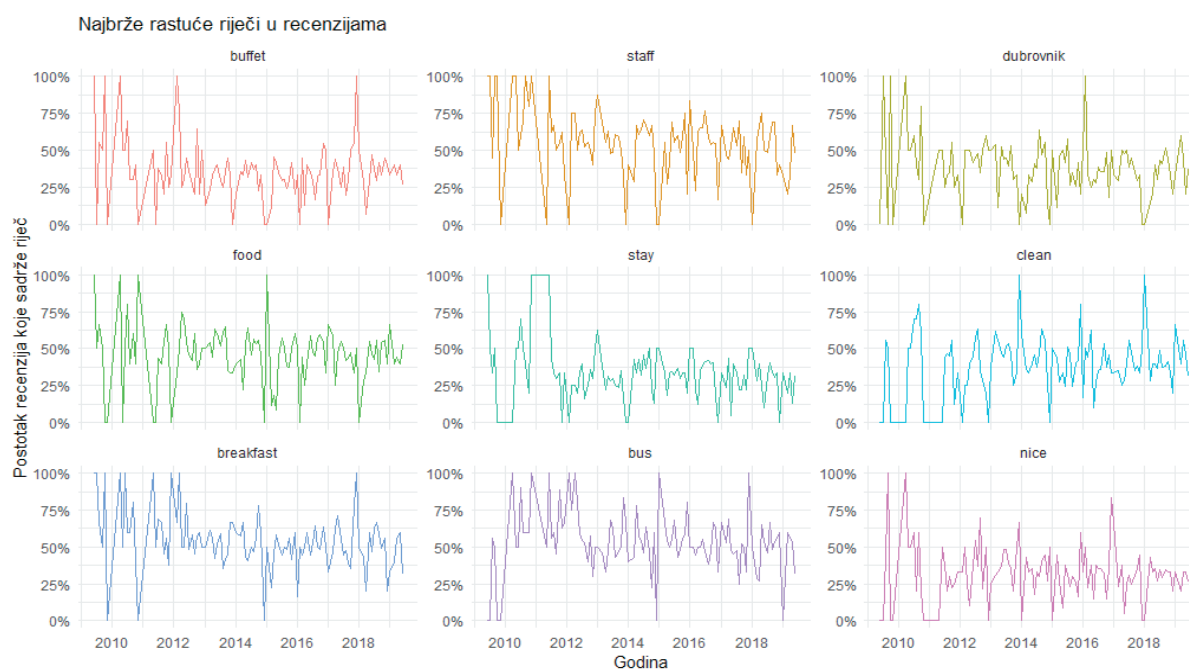
Tablica 1. Najčešće fraze od dvije riječi

Fraza	Pojavljivanja
valamar lacroma	258
sea view	242
bus stop	168
bus ride	147
breakfast buffet	136
indoor pool	130
half board	126
6 bus	120
outdoor pool	111
valamar president	100

Izvor: vlastito istraživanje

Slika 2. prikazuje kako se pojedine riječi odnosno teme pojavljuju s obzirom na vrijeme. Prikazano je devet najbrže rastućih (s obzirom na frekvenciju pojavljivanja u recenzijama) riječi od onih koje se pojavljuju barem 500 puta. Ovo može poslužiti za razumijevanje promjena u ekosustavu hotela, primjerice neke usluge ili renoviranja i slično.

Slika 2. Najbrže rastuće riječi

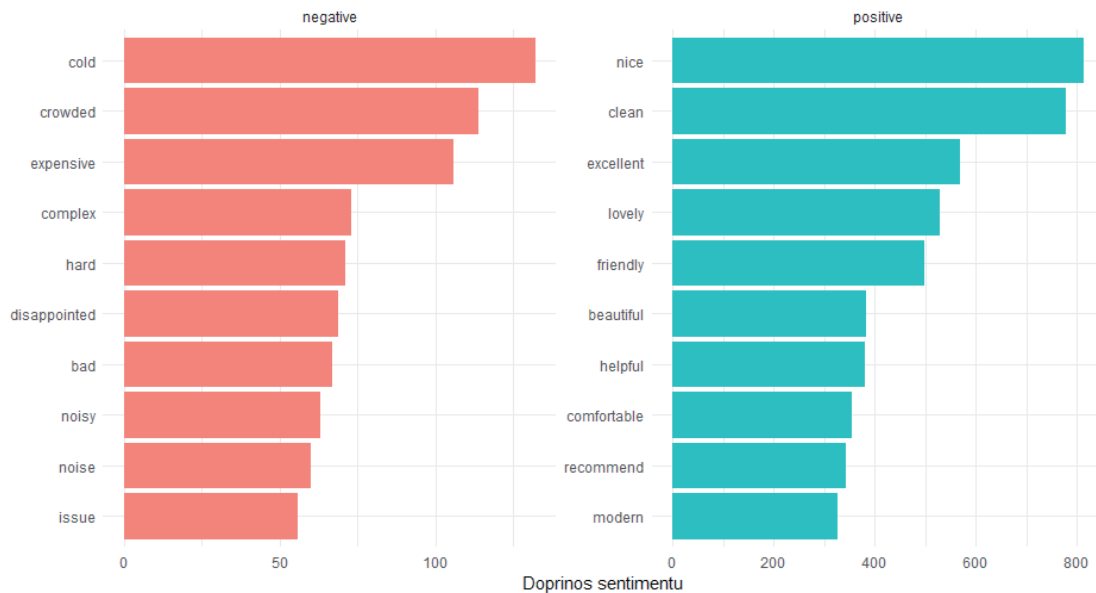


Izvor: vlastito istraživanje

Analiza sentimenta provedena je najprije pomoću rječnika Bing u kojem su riječi koje nose sentiment označene samo kao pozitivne ili negativne. Slika 3. prikazuje riječi s najvećim do-

prinosom (mjereno učestalošću riječi) pozitivnom odnosno negativnom sentimentu. Vrijedi uočiti da najčešće pozitivne riječi višestruko nadmašuju najčešće negativne riječi. Riječ nice pojavljuje se 813, clean 779, a excellent 569 puta. S druge strane, riječ cold pojavljuje se 132, crowded 114, a expensive 106 puta.

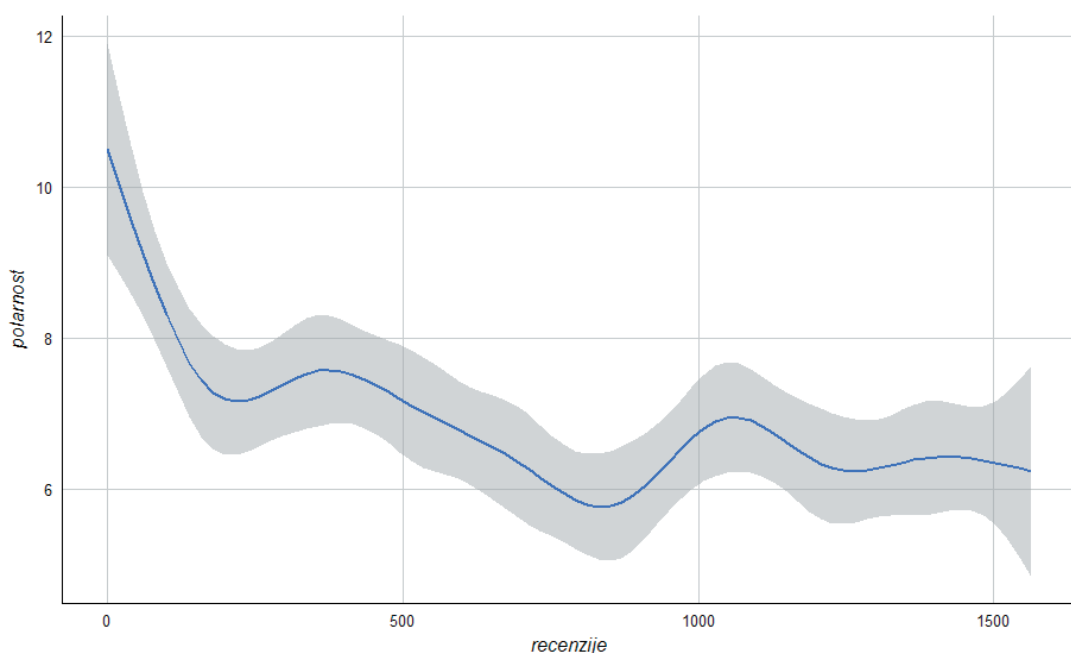
Slika 3. Riječi s najvećim doprinosom sentimentu



Izvor: vlastito istraživanje

Slika 4. prikazuje promjenu sentimenta tijekom vremena. Recenzije su poredane kronološki od najstarije (na lijevoj strani apscise) prema najnovijoj (na desnoj strani apscise). Na slici je vidljiv izrazito pozitivni sentiment (polarnost) u početnim recenzijama. Treba istaknuti da je krivulja cijelo vrijeme značajno iznad nule, što pokazuje da uvjerljivo dominiraju pozitivne recenzije.

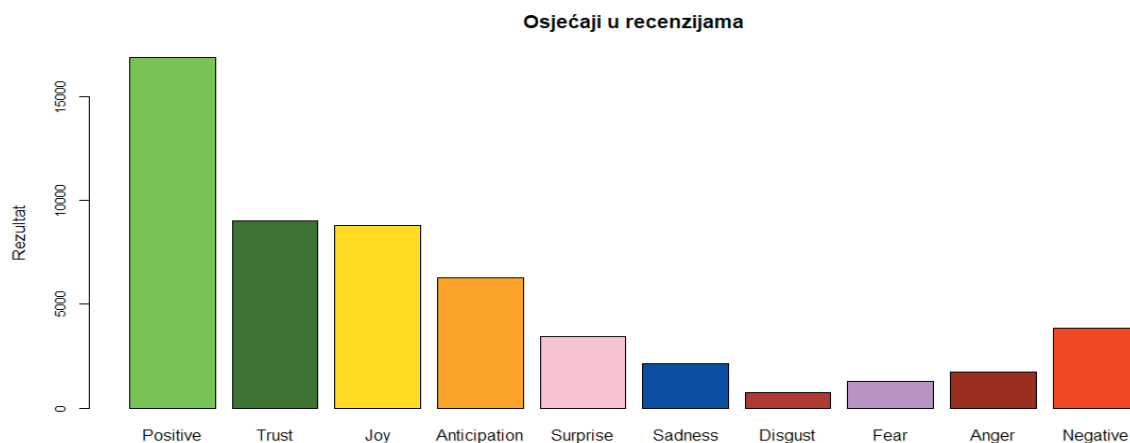
Slika 4. Promjena sentimenta tijekom vremena



Izvor: vlastito istraživanje

Nakon rječnika Bing, analiza je rađena pomoću rječnika NRC. Ovaj rječnik, osim mogućnosti prepoznavanja pozitivnih i negativnih riječi, može prepoznati i jedan od osam konkretnih osjećaja. Pojedine osjećaje u recenzijama pokazuje slika 5. Primjenom NRC, kao i Bing rječnika, jasno je vidljivo da u recenzijama pretežu pozitivni osjećaji, dok je negativnih značajno manje. Najmanje zastupljeni je osjećaj gađenja (disgust), sasvim očekivano budući da se radi o hotelu s četiri zvjezdice.

Slika 5. Osjećaji u recenzijama



Izvor: vlastito istraživanje

Problem rječnika je izolirano promatranje pojedinih riječi. Negiranjem riječi mijenja se polarnost mišljena. U rečenici jedne recenzije "The staff are not friendly and welcoming ..." analiza rječnikom dat će pozitivan sentiment zbog riječi *friendly* i *welcoming*, a u stvarnosti značenje je suprotno. Samo riječ *not* pojavljuje se 488 puta, a to je tek jedna od riječi negiranja. Tablica 2 pokazuje najčešće negirane riječi pomoću *not*. Stupac 'Vrijednost' sadrži brojkom izraženu snagu pozitivnosti ili negativnosti u rječniku AFFIN – trećem korištenom u ovom istraživanju. Riječi koje nose sentiment mogu poprimiti jednu od cjelobrojnih vrijednosti između -5 (najnegativnije) i 5 (najpozitivnije). Za razliku od rječnika Bing, koji samo daje informaciju je li riječ pozitivna ili negativna, AFFIN dodatno pridružuje 'snagu' riječi – tako je primjerice, *easy* (+1) manje pozitivna riječ od *great* (+3).

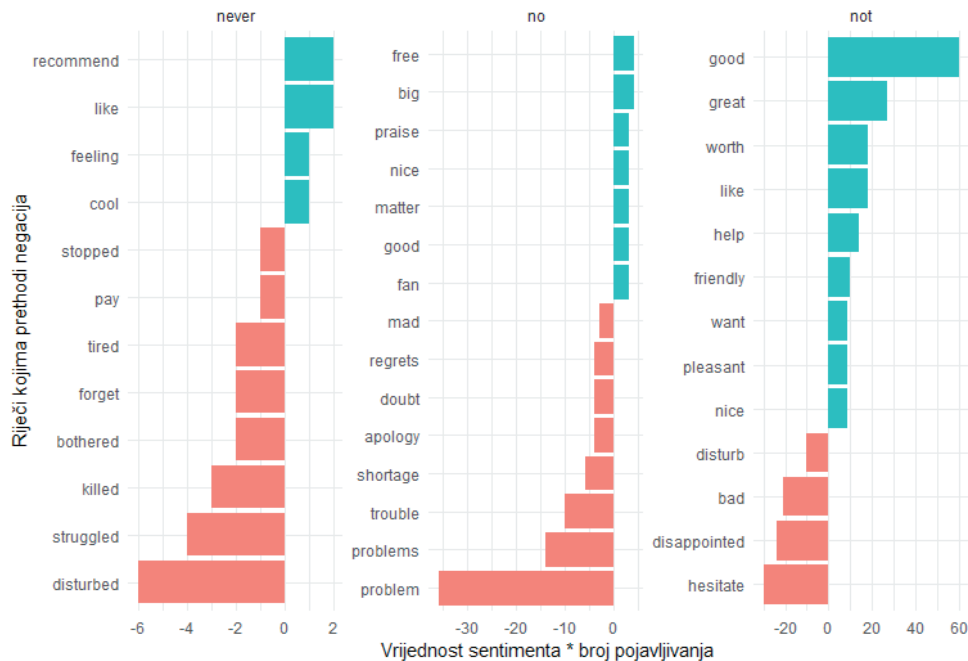
Tablica 2: Pojavljivanja riječi negiranih s not

Riječ	Vrijednost	Pojavljivanja
good	3	20
hesitate	-2	15
disappointed	-2	12
great	3	9
like	2	9
want	1	9
worth	2	9
bad	-3	7
easy	1	7
help	2	7

Izvor: vlastito istraživanje

Slika 6. prikazuje na kojim se riječima najviše griješlo u izračunu sentimenta zbog negiranja s not, no i never. Riječi good, great, worth i like doprinose pozitivnom sentimentu, a zapravo su negativne jer su negirane. Isto tako, riječi hesitate, disappointed i bad smatrane su negativnim, ali su zbog negiranja zapravo pozitivne.

Slika 6. Osjećaji u recenzijama



Izvor: vlastito istraživanje

Na temelju izračunatog sentimenta, pronađene su najpozitivnija i najnegativnija recenzija (Slika 7.).

Slika 7. Najpozitivnija i najnegativnija recenzija

"SophieDonalson", "15-Aug-14", "Newcastle upon Tyne, United Kingdom", "Amazing hotel!", "This hotel is really fantastic! It is very modern with brilliant facilities. Rooms were nice, large and had everything you needed in them. Pool was gorgeous, always got sunbeds and the indoor pool turned out to be great for the cloudy morning we had. Breakfast was amazing!!! Everything you could possibly wish for! Omelettes were fab and the lady who made them was such good fun! Pool bar also served very good and reasonable food! Hotel was about a 15/20 minute walk along the coast to Lapad where there were lots of nice bars and restaurants and a 15/20 minutes bus journey to the old town! Couldn't have asked for more and will definitely be going back!"

"Califreqflyer", "15-May-14", "San Francisco, California", "Ok hotel, terrible staff and service", "I stayed here while I was attending a conference in the hotel. From the start, the service from the hotel staff was terrible. We did not know that our taxi driver required cash and upon getting to the hotel, I asked if the hotel had a cash machine. The staff (Vilanka) on duty was extremely annoyed and rude. There was a cash machine onsite, however it did not work and she did not inform me that it wasn't working. I also had a hard time finding the fitness center (the hotel is very maze-like and confusing) and asked the staff at the desk for directions. They actually ignored me for quite some time until I insisted a second time for help! Finally, the night before checkout, I asked for a 4:15 am wakeup call and for a taxi to come at 4:40. Surprise surprise, wakeup call at 4:25 (10 minutes late) and no taxi. I waited until 4:50, again asking Vilanka where the taxi was. She completely ignored me and rolled her eyes. Because of her, I almost missed my flight. Beware of the terrible staff and service!"

Izvor: vlastito istraživanje

6. Zaključak

Povratne informacije gostiju o (ne)zadovoljstvu uslugama oduvijek su bile važne menadžmentu u sektoru turizma i ugostiteljstva. One su, u obliku recenzija, danas javno dostupne na internetu svim zainteresiranim stranama (što uključuje i konkurenciju).

Relativno mali broj recenzija pojedina osoba može sama obraditi i izvući iz njih korisne informacije. Međutim, s porastom broja recenzija nastoji se proces automatizirati. Ovdje se ne misli na odgovaranje predstavnika hotela na recenziju, već na teme koje se spominju u dužem vremenskom periodu – godinama. Osoba koja čita recenzije na dnevnoj bazi teško može sagledati što gosti hvale i što ih muči u dužoj vremenskoj perspektivi. Analiza sentimenta može biti vrijedna menadžmentu hotela i u usporedbi s drugim konkurentskim hotelima, ili u usporedbi s drugim hotelima istog hotelskog poduzeća.

U ovom radu prikazane su neke mogućnosti analize sentimenta pomoću tri raspoloživa rječnika (Bing, AFFIN i NRC). Obradeno je više od 1,500 izvornih recenzija, a rezultati su vizualizirani kako bi se mogli prepoznati trendovi i pronaći iznimke, što može dati podlogu za daljnje analize. Primjerice, mogle bi se izdvojiti recenzije koje su označene osjećajem straha te vidjeti što izaziva strah kod gostiju u hotelu. Ili, moglo bi se tijekom vremena pratiti smanjuje li se ili povećava ljutnja. Mogućnosti analize sentimenta su brojne, no treba biti oprezan u provođenju iste – u radu je prikazan problem kako se dolazi do pogrešnih rezultata kad se ne vodi računa o negacijama koje mijenjaju polarost mišljenja.

LITERATURA

1. Banerjee, S., i Chua, A. Y. (2017). Authentic versus fictitious online reviews: A textual analysis across luxury, budget, and mid-range hotels. *Journal of Information Science*, 43(1), 122-134.
2. Bucur, C. (2015). Architecture Of A Sentiment Analysis Platform. *Annals-Economy Series*, 154-159.
3. Farooq, U., Mansoor, H., Nongillard, A., Ouzrout, Y., & Qadir, M. A. (2017). Negation Handling in Sentiment Analysis at Sentence Level. *JCP*, 12(5), 470-478.
4. Hargreaves, C. A. (2015). Analysis of hotel guest satisfaction ratings and reviews: an application in Singapore. *American Journal of Marketing Research*, 1(4), 208-214.
5. Internet World Stats, <https://www.internetworldstats.com/stats.htm> (pristupljeno: lipanj 2019.)
6. Kasper, W., i Vela, M. (2011). Sentiment analysis for hotel reviews. In *Computational linguistics-applications conference*, Vol. 231527, 45-52.
7. Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. New York, Cambridge University Press.
8. Philander, K., i Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55(2016), 16-24.
9. Schuckert, M., Xianwei L., i Rob L. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing* 32.5, 608-621.
10. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., i Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, 18-38.
11. Stecanella, B. (2016). *Machine Learning over 1M hotel reviews finds interesting insights*, <https://monkeylearn.com/blog/machine-learning-hotel-reviews-insights/> (pristupljeno: lipanj 2019.)
12. Yordanova, S., i Kabakchieva, D. (2017). Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning. *International Journal of Computer Applications*, 158(7).

*Summary***SENTIMENT ANALYSIS OF HOTEL REVIEWS**

A large number of product and service reviews is generated on Internet on a daily basis. These reviews can serve as a valuable source of information to service providers on customer's (non) satisfaction with the aim of improving the offer. The main problem is the painstakingness of the review analysis. That is why there is intention to automate this process, and that is a purpose of sentiment analysis – a method of applying computer-based natural language processing that identifies, extracts and analyzes subjective information in the text. There are two fundamental approaches to sentiment analysis: machine learning and lexical approach. In this paper, a lexical approach was used using the three dictionaries: Bing, AFFIN and NRC. NRC dictionary reveals not only the positive and negative aspect of the sentence, but also the specific feelings such as anger, confidence, disgust and surprise. For a selected hotel in Dubrovnik, a sentiment analysis was conducted based on more than a 1,500 reviews published on TripAdvisor. The analysis reveals the most frequent words and phrases as well as words that contribute to positive and negative sentiment. Negations – words that change sentence polarity - are taken into consideration. Research results also show a change of sentiment over time.

Keywords: *opinion mining, sentiment analysis, machine learning, hotel reviews*