

ANALIZA MNENJ OBISKOVALCEV O TURISTIČNI PONUDBI V LJUBLJANI

TJAŠA REDEK¹, UROŠ GODNOV²

POVZETEK: *Turisti se pri odločanju vedno močneje opirajo na spletne vire podatkov o značilnostih turistične ponudbe v določenem kraju in predvsem o kvaliteti oziroma značilnostih konkretnih ponudnikov. Namen članka je preučiti značilnosti turistične ponudbe Ljubljane, predvsem posameznih vrst le-te (hoteli, gostinska, kulturna ponudba, itd.) s pomočjo vsebinske in čustvene analize tekstovnih ocen z največjega spletnega turističnega portala. S pomočjo metod tekstovnega rudarjenja smo analizirali slabih 9 tisoč ocen različnih ponudnikov v Ljubljani. Rezultati čustvene analize kažejo, da so ocenjevalci najbolj pozitivno spregovorili o hotelih, sledijo gostinska ponudba, kulturno-zgodovinska ponudba in šele na koncu mestno jedro. Analizo dopolnimo z vsebinsko analizo tekstovnih komentarjev. Pričujoča analiza predstavlja prvo tovrstno oceno slovenskega turističnega mesta Ljubljana. Metodološko članek temelji na najnovejši metodologiji tekstovnega rudarjenja in kot prvi analitično povzame vsebino mnenj turistov za Ljubljano. Analiza opozori tudi na možnosti uporabe razpoložljivih programskih orodij ob uporabi razpoložljivih spletnih virov v podporo poslovnemu odločanju.*

Ključne besede: *turizem, Tripadvisor, Ljubljana, tekstovno rudarjenje*

JEL klasifikacija: Z32, C89, M3

DOI: 10.15458/85451.29

UVOD

TripAdvisor.com je poleg spletnih strani, kot sta Booking.com in Expedia.com, ena od ključnih spletnih strani, kjer potencialni turisti iščemo informacije o kraju, kamor želimo potovati, o tem, kaj je vredno ogleda, katere so dobre, katere slabe restavracije v izbranem kraju in kateri hotel je »vreden svoje cene«. Po podatkih Bassiga (2013) več kot 90 odstotkov potrošnikov svoje odločitve sprejema s pomočjo Tripadvisorja ali podobnih strani, zato ne preseneča, da je mogoče med turističnimi ponudniki pogosto slišati: »Če te ni na Tripadvisorju, te NI.« Zato so informacije, ki jih lahko turistični ponudniki pridobijo s spletnih portalov izjemno dragocen vir podatkov, ki omogočijo med drugim tudi pozicioniranje podjetja na trgu z vidika kvalitete in cene; predvsem pa omogočijo ponudnikom tudi, da prepoznajo, kaj so njihove glavne prednosti in kaj slabosti in seveda skušajo le-te odpraviti.

¹ Univerza v Ljubljani, Ekonomska fakulteta, Ljubljana, Slovenija, e-pošta: tjasa.redek@ef.uni-lj.si

² Univerza na Primorskem, Fakulteta za management, Koper, Slovenija, e-pošta: uros.godnov@fm-kp.si

V Sloveniji postaja t.i. »mestni turizem« vedno bolj pomemben, med lokacijami pa je seveda, podobno kot v številnih drugih državah, najbolj v ospredju prestolnica, Ljubljana. Popotnike mesta privlačijo, ker predstavljajo dinamično mešanico kulturnega, gospodarskega in družbenega razvoja in omogočajo posamezniku da doživi »koncentrirano« izkušnjo (UNWTO, 2012). Urbani oziroma mestni turizem je začel pridobivati na pomenu predvsem v osemdesetih, ko je nekaj vzporednih trendov (predvsem povečanje zanimanja za mesta, kulturo in zgodovino ter razvoj transporta) omogočilo razvoj produkta »krajšega mestnega oddiha« (*angl.* city break). Urbani turizem napreduje zelo hitro, saj ima zelo širok spekter ciljnih kupcev; mesta nudijo pravzaprav »nekaj za vsakogar«, tako z vsebinskega kot tudi cenovnega vidika. Po podatkih UNWTO (2012) naj bi število urbanih turistov vsako leto do leta 2030 naraslo za okoli 3,3 %, oziroma naj bi se vsako leto število prihodov povečalo za dobrih 40 milijonov.

V Evropi so najbolj popularne večje prestolnice. London je bil v letu 2015 ocenjen celo za vodilno urbano destinacijo v svetu, medtem ko je bil Pariz tretji. Med najbolj popularnimi desetimi so se nato izmenjavala azijska, ameriška mesta ter mesta iz Bližnjega vzhoda (CNN, 2015). Med evropskimi mesti so seveda v ospredju London, Pariz, Berlin in Rim (European capital city tourism, 2012). Ljubljana postaja v zadnjih nekaj letih vedno bolj popularna destinacija za krajše obiske. Ocene celo kažejo, da je bila med evropskimi mesti ravno Ljubljana tista, ki je beležila najhitrejšo rast obiskov (European capital city tourism, 2012).

Glede na pomen družbenih omrežij pri izboru turističnih destinacij nasploh, lahko sklepamo, da so le-ta in že omenjene platforme (Tripadvisor.com, Booking.com, itd.) pomembne tako pri odločanju o izboru ciljne destinacije pri turistih, ki izbirajo med različnimi mesti, kot tudi pri izboru namestitve in obiskanih lokacij znotraj izbrane urbane destinacije.

Namen članka je celovito analizirati turistično ponudbo v Ljubljani skozi tako številske kot tudi predvsem tekstovne (vsebinske) ocene gostov. S pomočjo najnovejših metod tekstovnega rudarjenja na približno 9000 tekstovnih ocenah, ki smo jih zbrali z najbolj znanega turističnega portala, članek analizira osnovne tipe turistične ponudbe: namestitve (hoteli), gostinsko ponudbo ter kulturno-zgodovinsko ponudbo. Zaradi pomena starega mestnega jedra kot ene temeljnih privlačnih točk Ljubljane smo posebej analizirali še to. Pri analizi nas je zanimalo predvsem, kako turisti ocenjujejo različne vidike turistične ponudbe, in sicer tako z vidika številske ocene, ki jo podajo, kot tudi z vidika ocene, ki jo lahko prebere bralec »med vrsticami«. Načeloma bi morale imeti destinacije, ki so številsko boljše ocenjene, tudi boljše vsebinske ocene. Če temu ni tako, nas je zanimalo, kako močna je sploh povezava med njima in zakaj prihaja do razlik. Na koncu nas je zanimala tudi vsebina ocen. Kateri so tisti elementi ponudbe za vsakega od navedenih tipov, o katerih gosti največ diskutirajo, in jih lahko tako privzamemo kot tiste, ki so najbolj pomembni?

Članek prispeva k razvoju literature na področju turizma, konkretno vloge socialnih omrežij v turizmu z več vidikov ter k aplikaciji tekstovnega rudarjenja (*angl.* text mining) v turizmu. Gre za prvo oceno turistične ponudbe na podlagi tako široke podatkovne baze in eno redkih analiz, ki poleg čustvene analize dopolni analizo spletnih ocen še z vsebinsko

analizo (*angl.* content analysis) (glej npr. tudi Bassig, 2013, Bjorkelund et al., 2012, Simms in Gretzl, 2013, Jeong in Jeon, 2008). Članka, ki bi celovito analiziral urbano turistično ponudbo oziroma značilnosti le-te skozi vsebino ocen na različnih platformah, v literaturi še ni zaslediti, na voljo so le zelo omejene analize na vzorcu ocen ali z zelo specifičnim namenom (npr. Vasquez, 2011, na 100 ocenah, Tamajon in Valeinete, 2015 za Barcelono). Metodološko pričujoča analiza temelji na najnovejših metodah tekstovnega rudarjenja (npr. Feinerer in Hornik, 2015) in je prva takšna analiza za Ljubljano. Predstavlja tudi prvo aplikacijo tovrstno analize v slovenskem turizmu. Kot prvi tako pričujoči tekst omogoča primerjalno analizo in odgovor na vprašanje, kateri tipi ponudbe so v Ljubljani boljši in kateri slabši, tako z vidika številskih ocen, kot tudi z vidika tekstovnih ocen. Analiza kot prva pokaže tudi, o čem diskutirajo ocenjevalci Ljubljane in tako identificira glavne pozitivne in negativne elemente posameznega tipa ponudbe. V obeh primerih prispeva s tem tudi k razumevanju vloge družbenih omrežij pri potrošniškem odločanju ter k razumevanju vloge potrošnika kot vira informacij v procesu nakupnega obnašanja. S tem širi literaturo, ki ta vidik že analizira (npr. Jacobsen, 2015, Bassig, 2013, Gretzel et al., 2007). Prispeva tudi k razvoju diskurza o vlogi informacij z družbenih omrežij ter spleta nasploh za konkurenčno analizo (npr. tudi Gemar in Jimenez-Quintero, 2015). Z vidika uporabne vrednosti pa podaja dragocen primerjalni vidik ter zbirni pogled v vsebino komentarjev, s pomočjo katerih lahko turistični ponudniki v Ljubljani ugotovijo, kaj so temeljne ugotovitve za posamezni analizirani tip ponudbe. Hkrati pa omogoča vsakemu ponudniku, da v tej povprečni oceni najde tudi sebe oziroma se pozicionira (ob predpostavki, da vsak seveda pozna svoje ocene na spletu). Nenazadnje pa članek opozarja na do sedaj hitro razvijajoče se področje na področju empirične analize, in s tem odpira vrsto raziskovalnih izzivov.

1 LJUBLJANSKI TURIZEM V ŠTEVILKAH

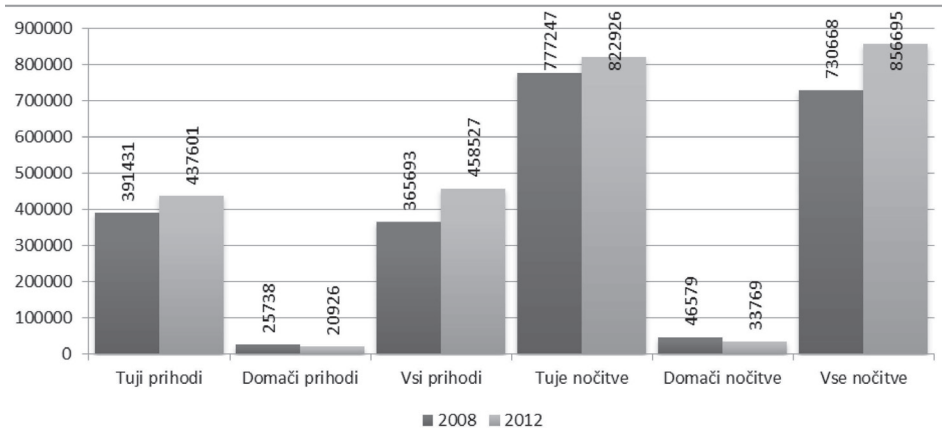
Turizem je v Sloveniji relativno pomembna panoga. Po podatkih WTTC (2015) je turizem v letu 2014 neposredno prispeval 3,5 % k bruto domačem proizvodu, s posrednimi vplivi pa celo 12 %. Neposredno je bilo v turizmu zaposlenih 31 tisoč ljudi, skupaj pa, upoštevaje tudi posredne vplive, celo 103 tisoč ljudi. Turizem predstavlja tudi 8 % našega izvoza, samo januarja 2015 turistični sektor zabeležil 546 tisoč nočitev (Statistični urad, 2015). Navkljub majhnosti Slovenija ponuja številne različne možnosti: od klasičnega preživljanja prostega časa ob morju do planinarjenja, kmečkega turizma, zdraviliškega turizma ter seveda tudi obiska mest.

Ljubljana je z vidika turistov zagotovo eno najbolj privlačnih mest v Sloveniji. Če citiramo uporabnika s spleta: »(...)Kakorkoli že, ugotovili boste, da je Ljubljana eno najbolj navdušujočih, sofisticiranih in vizualno navdušujočih manjših glavnih mest v Evropi. Postavite ga na svoj radar raje malo prej kot kasneje.«

Ta podatek potrjujejo tudi številke. Ljubljana je namreč med evropskimi prestolnicami v zadnjih letih zelo hitro pridobivala na popularnosti. Leta 2012 je bila tako tretja v Evropi po rasti števila gostov, leta 2015 pa se je uvrstila na sedmo mesto na lestvici European best destination (European Best Destination, 2015). Zadnji statistični podatki, ki so sicer

za leto 2012, kažejo, da je Ljubljana zabeležila 458 tisoč prihodov turistov in 856 tisoč nočitev, od tega, pričakovano, daleč prevladujejo tujci, ti predstavljajo okoli 95 % vseh prihodov in okoli 96 % vseh nočitev (Statistični letopis Ljubljane, 2013, Slika 1).

Slika 1: Število prihodov in nočitev turistov v Ljubljani, 2008 in 2012*

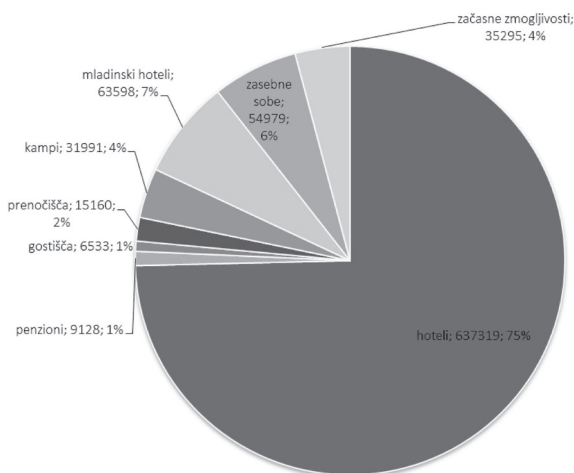


Prihod turistov: število posameznikov, ki so prišli. Nočitev: število noči, ki so jih turisti, ki so prišli, prespali v Ljubljani. Npr. če je posameznik prišel in trikrat prespal, je to 1 prihod in 3 nočitve (metodološka pojasnila v Vprašalniku za statistično raziskovanje Prihodi in prenočitve turistov, 2015).

Vir: Statistični letopis Ljubljane, 2013.

Od leta 2009 je namreč število prihodov turistov v Ljubljano naraslo za 24 %, pri čemer je mogoče to rast pripisati predvsem 27-odstotnemu porastu števila tujih turistov, domači so namreč pokazali manj interesa za glavno mesto in je njihovo število upadlo (Slika 1). Najbolj je poraslo število prihodov turistov iz Islandije in Ukrajine (za skoraj 400 %, a je potrebno upoštevati, da je bilo prvih leta 2008 le 178, drugih pa 1499). Od številčno bolj močnih skupin je bilo mogoče velik porast zabeležiti tudi med izraelskimi ter kitajskimi gosti. Med številčno najmočnejšimi skupinami pa je število nemških gostov, ki jih je bilo leta 2008 26 tisoč, do leta 2012 naraslo za 34 %, število italijanskih prihodov pa je poraslo za 22%, in sicer s 43 na 53 tisoč. V povprečju so turisti v Ljubljani ostali slaba 2 dneva, tuji nekoliko več kot domači (1,9 in 1,6). Prenočevali so predvsem v hotelih, mladinskih hotelih zasebnih sobah in kampih (Slika 2).

Slika 2: Število prenočitev domačih in tujih gostov skupaj po tipih namestitev* ter odstotek prenočitev v posameznem tipu kot delež vseh nočitev



*Začasne zmogljivosti zajemajo študentske domove in internate.

Vir: Statistični letopis Ljubljane, 2013.

Ljubljano turisti obiščejo zaradi več razlogov, pogosto zato, ker je cenovno dostopna, šarmantna in je utrip v njej sproščen (Rebernik, 2014). Kot kažejo podatki s Tripadvisorja, turisti večinoma Ljubljano zelo dobro sprejmejo, pogosto to, kar vidijo in obiščejo, presega njihova pričakovanja. Če povzamemo eno od ocen s TripAdvisorja: »Rezervirala sva si 2 noči, kar ni dovolj. Ta kraj je tako čudovit in obiskati je potrebno kup stvari. Hodila sva naokoli ves dan in uživala v krasnem starem mestnem jedru, lepo ohranjenih stavbah, živahnem dogajanju v restavracijah in barih. Restavracije so izjemne. In, ja, svoje bivanje sva podaljšala.«

2 RAZISKOVALNA VPRAŠANJA IN METODOLOGIJA

Namen analize je bil celovito analizirati značilnosti turistične ponudbe v Ljubljani skozi vse dimenzije ocen, ki so na voljo, tako številske ocene kot tudi tekstovne ocene. Analizirali smo 4 tipe ponudbe: namestitev¹, gostinsko ponudbo, kulturno ponudbo in staro mestno jedro kot svoj tip ponudbe. Vsakega od tipov smo analizirali najprej ločeno, in sicer tako številske ocene kot vsebino, nato pa še primerjalno. Cilj analize je bil poiskati odgovor na naslednja vprašanja:

- 1) Kakšno je povprečno zadovoljstvo obiskovalca Ljubljane, in sicer nasploh in po posameznih tipih ponudb, če zadovoljstvo merimo s številsko oceno in pa če merimo zadovoljstvo s čustveno vrednostjo komentarjev?
- 2) Ali obstaja sistematična povezava med številsko oceno in čustveno vrednostjo tekstovne ocene? Drugače rečeno, ali lahko trdimo, da so vrste ponudbe, ki imajo boljšo številsko oceno tudi sistematično bolj vsebinsko ocenjena?

1 V analizo so bili zajeti ljubljanski hoteli (v mestu in ne v okolici). Le ti so imeli zadostno število ocen, ki je omogočalo kvalitetno analizo, hkrati pa se v njih opravi tudi največji del nočitev (Slika 2).

- 3) Kaj so ključni vsebinski poudarki, ki jih lahko zasledimo v ocenah za posamezne tipe ponudb? Sklepamo namreč lahko, da so tisti vidiki, ki se jih največ omenja, gostom tudi najbolj pomembni.
- 4) Katere značilnosti predstavljajo prednosti Ljubljane in katere ne? To vprašanje je povezano tudi z vprašanjem, kako lahko tekstovno rudarjenje, če bi bili analizirani posamezni ponudniki (in ne več ponudnikov skupaj, kot v tem primeru), uporabimo v konkurenčni analizi.

Metodologijo, uporabljeno v članku, v osnovi razdelimo na dva dela, in sicer pridobivanje podatkov ter analizo. Zajem podatkov je bil narejen s pomočjo programa R. Analiza metodološko temelji na več pristopih, ki predstavljajo nekatere od najnovejših dosežkov s področja tekstovnega rudarjenja in analize jezika (*angl.* natural language processing). V osnovi uporabljamo tri različne metodološke pristope, dve s področja tekstovnega rudarjenja (čustveno analizo (*angl.* sentiment analysis), analizo vsebin (*angl.* content analysis ali pa *angl.* topic modelling), ki ju kombiniramo s klasično statistično analizo. V nadaljevanju podajamo kratek opis obeh uporabljenih metod tekstovnega rudarjenja.

Čustvena analiza temelji na analizi vsake besede v komentarju. Tekom razvoja čustvene analize so se postopoma pripravljali t.i. leksikoni, ki vsaki besedi dodajo številsko čustveno vrednost. Prvi slovarji so bili zelo preprosti (te uporabljajo nekateri standardizirani programi za čustveno analizo) in so omogočali dodelitev zgolj treh vrednosti (-1, 0 in 1, torej negativno, nevtralno, pozitivno), pri čemer niso omogočali ocene čustvene vrednosti v kontekstu stavka kot celote (Hu in Liu, 2004, Liu, 2012). Najnovejši slovarji pa omogočajo oceno čustva, ki ga izraža posamezna beseda, na lestvici celo od -5 do 5, pri čemer je moč in smer čustva določena s pogojno verjetnostjo, torej od besed, ki so v okolici. Moderni algoritmi skupaj z novimi slovarji tako zaznajo sarkazem, cinizem, itd., česar prvi niso zmogli (več o metodologiji v Godnov in Redek, 2015a). V analizi bo uporabljena t.i. metoda AFINN (glej Nielsen, 2011, in Hansen et al., 2011), ki omogoča trenutno najbolj poglobljeno analizo in variacijo čustvene vrednosti besede in ki, kot rečeno, omogoča analizo besed na lestvici od -5 do 5.

Analizo čustvene vrednosti bomo nagradili še z analizo prevladujočega čustva v celotnem komentarju, ki s pomočjo naprednega NRC leksikona ter ob uporabi pogojne verjetnosti (torej zazna morebitni sarkazem) oceni, katero čustvo (veselje, pričakovanje, sreča, razočaranje, itd.) prevladuje v oceni kot celoti (več o metodologiji glej npr. Mohammad, 2015). Tovrstna analiza dopolni čustveno analizo, ker številčno čustveno oceno dopolni še s prevladujočim čustvom.

Druga metoda s področja tekstovnega rudarjenja je t.i. analiza vsebin, kjer smo uporabili tri različne pristope. Analiza vsebin se namreč pripravlja z več različnimi orodji. Najbolj osnovno orodje je iskanje ključnih besed, ki se jih določi s štetjem, pri čemer se upošteva koren besed, hkrati pa se izločijo vsebinsko nepomembne besede (vezniki, predlogi, ipd.). Čeprav je metodologija zelo preprosta, vseeno učinkovito povzame bistvene vsebinske poudarke. Osnovni namen ključnih besed je namreč ravno analitičen povzetek vsebinskih poudarkov. Sledi analiza s pomočjo korelacije oziroma moči povezav med besedami, kjer je mogoče že izločiti temeljne vsebinske vzorce. Analiza temelji na funkcijah iz »tm« paketa v programu R (glej npr. Feinerer, 2015). Poleg omenjene rudimentarne metode analiza v članku temelji tudi na bistveno bolj sofisticirani metodi iskanja vzorcev oziroma sistematičnih povezav v tekstu, kar sodi v t.i. »topic models« (Gruen in Hornig, 2011),

katerih temelj je iterativni proces, ki na podlagi pogojne verjetnosti identificira glavne teme ter besede, ki so povezane z vsako od tem (glej npr. tudi Godnov in Redek, 2015a, Blei, 2012, Chen, 2011). Omenjene metodologije omogočajo učinkovito analizo vsebine tekstovnih komentarjev. Ob predstavitvi rezultatov bodo, kjer je to smiselno zaradi lažje oziroma bolj nazorne interpretacije rezultatov, podana še dodatna metodološka pojasnila.

3 REZULTATI

Za namene analize mnenj o Ljubljani je bilo analiziranih skupaj slabih 9 tisoč ocen gostov, pri čemer je slaba tretjina ocen predstavljala ocene hotelov, dobrih 40 % je bilo ocen mesta, mestnega jedra in znamenitosti, slaba četrtnina ocen restavracij in lokalov, ostalo pa so predstavljale ocene galerij, muzejev (slednjih je žal malo) (Tabela 1).² V povprečju je bila številska ocena zelo visoka, 4,36. Povprečna tekstovna ocena je zajemala 77 besed, medtem ko je vsebovala najdaljša ocena kar 1122 besed. V poprečju so bile najdaljše ocene s področja namestitve, najkrajše pa ocene starega mestnega jedra.

Tabela 1: Opis vzorca in glavne opisne statistike po kategorijah*

		N	Minimum	Maksimum	Povprečje	St. odklon
Skupaj	Ocena	8878	1	5	4,36	0,822
	N besed	8878	1	1122	76,95	69,801
	Sentiment AF	8878	-20	162	10,36	8,7104
Namestitvev	Ocena	2778	1	5	4,29	0,827
	N besed	2778	7	1122	113,97	90,988
	Sentiment AF	2778	-15	162	14,78	10,7381
Gostinska ponudba	Ocena	2110	1	5	4,38	0,915
	N besed	2110	1	565	74,71	60,738
	Sentiment AF	2110	-20	69	10,78	7,7556
Mestno jedro	Ocena	3860	1	5	4,39	0,756
	N besed	3860	6	558	51,19	37,666
	Sentiment AF	3860	-11	59	6,981	5,5676
Kultura	Ocena	130	1	5	4,35	0,897
	N besed	130	19	398	86,9	68,884
	Sentiment AF	130	-1	37	9,546	7,879

*Ocena: številska ocena, ki jo poda pisec ocene (na lestvici 1-5). N besed – število besed v oceni. Sentiment AF – izračunana čustvena vrednost po AFINN metodologiji.

Vir: Lastna analiza.

² Želimo poudariti, da namen analize ni bil analizirati značilnosti posameznega ponudnika, čeprav metodologija to omogoča. Namen je bil analizirati povprečno stanje posameznega tipa ponudbe, s tem omogočiti vpogled v značilnosti le-te ter omogočiti primerjavo med njimi (v segmentih, kjer je to smiselno). Hkrati pa analiza predstavlja, kaj metodologija omogoča. Analiza posameznega ponudnika bi omogočila podrobno konkurenčno analizo.

V nadaljevanju sistematično predstavljamo ugotovitve, in sicer v vrstnem redu kot si sledijo raziskovalna vprašanja.

Zadovoljstvo obiskovalcev Ljubljane. Najprej nas je zato zanimalo, kakšno je povprečno zadovoljstvo obiskovalca Ljubljane, pri čemer smo zadovoljstvo merili s številskimi ocenami, poleg tega pa tudi z oceno čustvene vrednosti posameznih vsebinskih ocen.

Slika 4 prikazuje rezultate analize (v Tabeli 1 so tudi podatku o minimalnih, maksimalnih vrednostih in standardnih odklonih). V povprečju so ocenjevalci z Ljubljano zadovoljni, saj je povprečna številka ocena 4,4 (na lestvici 1 do 5). Če sodimo po številski oceni, so ocenjevalci najbolj zadovoljni z mestnim jedrom, sledi gostinska ponudba, najmanj pa so zadovoljni z namestitvijo. Razlika je sicer majhna, vendar pa Kruskal-Wallis test v vseh treh primerih (tako za številsko oceno, kot tudi za oba sentimenta) zavrne z veliko značilnostjo ($p < 0,0001$ v vseh treh primerih), da so porazdelitve enake.

Povezava med številsko in čustveno oceno. Analiza vsebine tekstovnih ocen namesto zgolj številске ocene poda bistveno boljše informacijo o tem, kaj turist dejansko meni o neki lokaciji. Najprej se lahko vtis o tem, ali je nekdo zadovoljen z neko destinacijo ali ne, meri preko analize čustvene vrednosti teksta, ki odraža, kakšno »čustvo« veje iz komentarjev. Če je nekomu nekaj res všeč, bo skozi besede, ki jih uporablja, to tudi pokazal. Besede so tiste, ki izražajo neko čustvo. »Dobro« in »super« ni enako, pri čemer pa imata lahko oba komentarja npr. isto številsko oceno, npr. 5. Torej sama številka ocena ni dovolj za pridobitev celovite informacije o mnenju posameznika o neki lokaciji. Številka ocena je zgolj diskretna vrednost, ki zavzema vrednosti od 1 do 5; ocenjevalec je pravzaprav zelo omejen v svoji izbiri. Hkrati pa sta lahko dve številsko odlični oceni vsebinsko povsem različni. Ni namreč vseeno, če uporabnik napiše, da je bilo »dobro« ali pa napiše »odlično, izjemno, vsekakor bom še prišel«. Za vse, ki uporabljajo Tripadvisor kot vir informacij, ki pomagajo ustvariti sliko o neki lokaciji in se tako odločiti, so dragocene predvsem tiste informacije, ki jih preberemo »med vrsticami«. Torej, ne samo gola vsebina, temveč tudi kakšen vtis dobimo, ko ta tekst preberemo. Kot primer navajamo dve oceni v Tabeli 2. Obe sta imeli številsko oceno 5, vendar pa je druga (ki smo jo zaradi dolžine skrajšali) bistveno bolj čustvena, pozitivnost veje iz uporabljenih besed. Da se ton ne bi izgubil skozi prevod ali pa da ga ne bi spremenili zaradi izbora besed, puščamo tekst nepreveden.

Tabela 2: *Razlika med številsko in čustveno oceno*

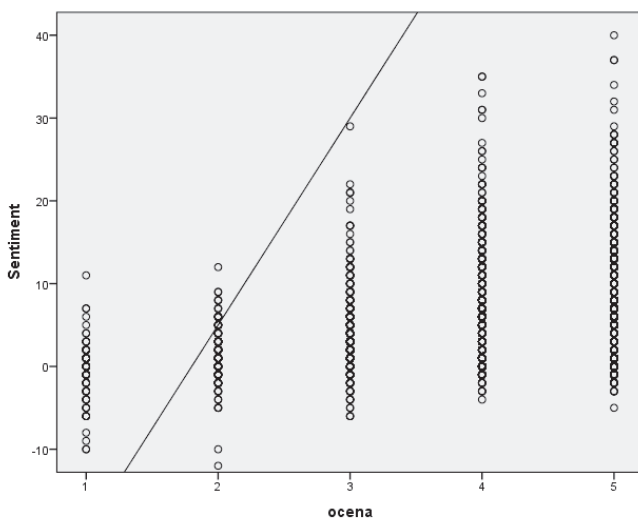
	Številska ocena	Čustvena vrednost (Sentiment)
We spent just one night here in mid-June. The hotel is small but very comfortably appointed and exceptionally clean. The service is a notch above. In particular, we appreciated the complimentary mini-bar and wi-fi. It's so refreshing to stay in a hotel and not be nickel-and-dimed for all of the extras. The location is convenient to the old town (5 minutes walk), but it is on a rather busy street. This wasn't a problem for us (no noise issues) but if you're looking to have the old town atmosphere right out the front door, then you might be disappointed. Our only real regret is that we were not able to stay longer to explore this enchanting city.	5	-2

	Številska ocena	Čustvena vrednost (Sentiment)
What is the best? 1.Location 2.Staff 3.Bed 4.Amenities 5.Breakfast 6.Impression at check-in and check-out. (...)It is for sure the best four star hotel we have stayed in. If somebody ask me what the best is in hotel ***, I would answer „everything“. (...) kindness and those wonderful donuts made specially for us. We would like to thank also to *** for his kindness and hospitality and wonderful baker who make magic donuts, but most of all we thank to general manager ***. We really respect and admire him and his team and are looking forward to visit *** again.	5	162

Vir: Lastna analiza.

Čustvena vrednost tekstovne ocene ter številka vrednost ocene torej nista nujno močno povezani. Da res nista, potrди tudi Slika 3, ki prikazuje razsevni diagram med številsko oceno ter čustveno oceno. Povezava sicer je, je tudi pozitivna, vendar pa ni močna. Korelacijski koeficient je 0,218, ki pa je visoko značilen ($p=0,000$).

Slika 3: Razsevni diagram med številsko oceno ter AFINN sentimentom

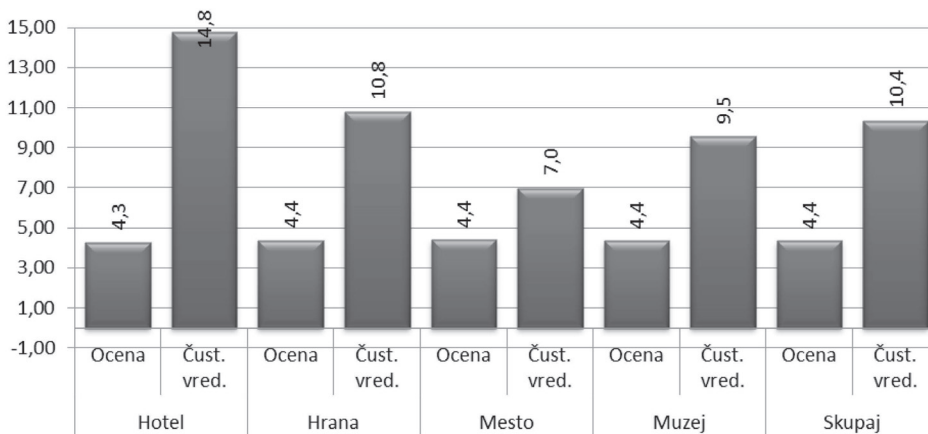


Vir: Lastna analiza.

Omenjeni rezultati potrjujejo, da je smiselno bolj natančno raziskati tudi razlike v čustveni vrednosti teksta, seveda pa v drugi fazi tudi vsebino. Tudi ocene čustvenih vrednosti tekstovnih ocen kažejo, da so gosti v povprečju kar zadovoljni (Slika 4, Tabela 1). Vendar pa, če so številčne ocene kazale rahlo prednost mestnega jedra pred ostalimi ocenami, pa ocene čustvene vrednosti komentarjev kažejo, da so gosti pravzaprav najbolj zadovoljni z namestitvijo. Najboljših pet ocen, če jih gledamo z vidika čustvene vrednosti, je ocen

namestitev (najboljša ocena je deloma zajeta v Tabeli 2). Bolj natančno: med najboljšimi 50 ocenami z vidika čustvene vrednosti je samo 6 ocen, ki niso s področja namestitve, od tega je ena ocena mestnega jedra, ostalo so ocene gostinske ponudbe.

Slika 4: Številčne ocene in čustvene vrednosti

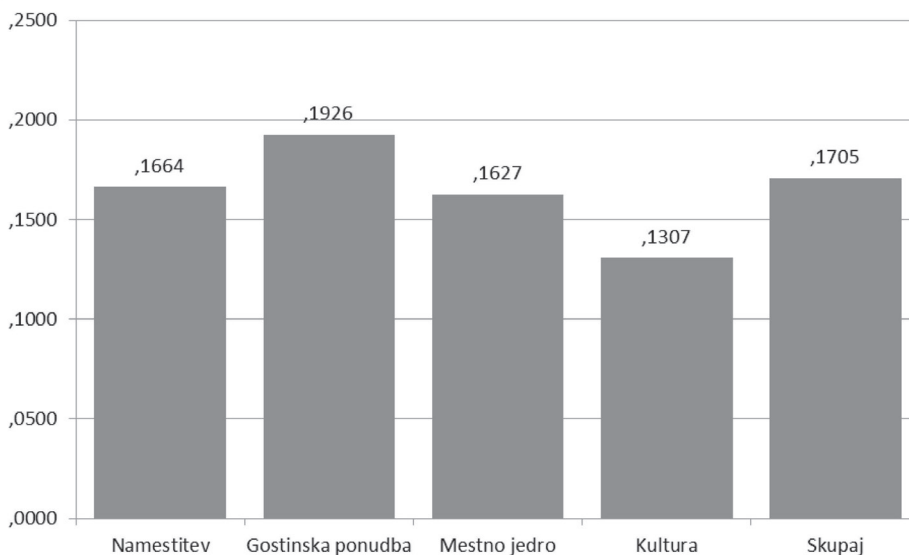


Vir: Lastna analiza.

Ker pa je čustvena vrednost povezana z dolžino ocene (glej npr. Godnov in Redek, 2015a, 2015b), smo v nadaljevanju čustveno vrednost preračunali na dolžino tekstovne ocene. Analiza stanja v tako slovenskem kot tudi hrvaškem turizmu (Godnov in Redek, 2015a, 2015b) je pokazala, da so daljše ocene v povprečju bolj negativne. To je pričakovano: jezen in razočaran uporabnik si bo vzel čas in podrobno opisal težave. Seveda pa daljša ocena, tudi kadar je ta pozitivna, vpliva na povprečno vrednost, in sicer v pozitivnem smislu. Vendar pa z naraščanjem števila besed v komentarju, narašča tudi seštevek sentimentov. Zato je tudi s tega vidika smiselno, da se sentiment standardizira glede na dolžino.

Tudi v analizi ocen Ljubljene ugotavljamo podobno: korelacija med številom besed in sentimentom na besedo je negativna in značilna, torej so daljše ocene bolj negativne. Korelacija sicer ni močna, koeficient je v primeru AFINN metode $-0,293$, vendar pa je značilnost zelo visoka ($p=0,000$). Tudi korelacija med številom besed in številsko oceno je negativna in visoko značilna, vendar šibka ($\text{corr}=-0,1$, $p=0,000$). Res pa je, da je bila v tem vzorcu najbolj pozitivna ocena tudi najdaljša (že delno citirana v tekstu, vsebovala je kar 1122 besed, kar je približno 15-krat več od povprečne. Vendar ta odnos, da so boljše ocene tudi daljše, v povprečju ne velja.

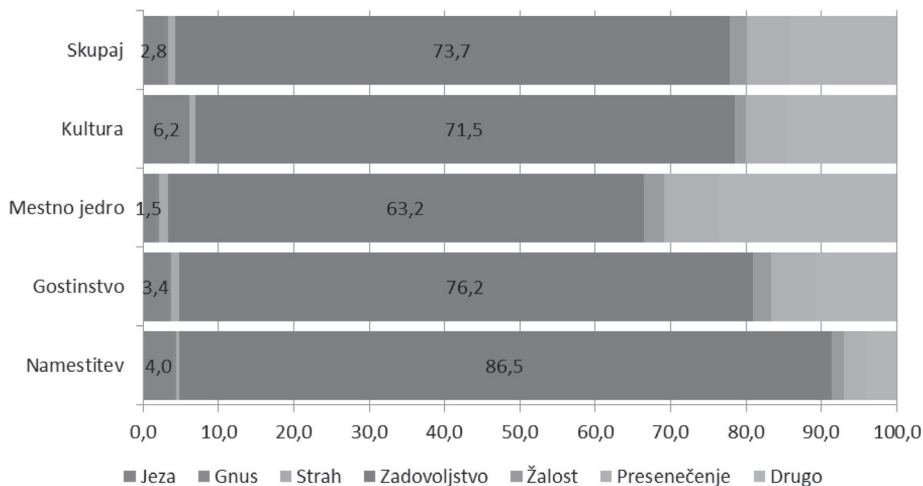
Slika 5: Povprečna čustvena vrednost po kategorijah, upošteva standardizacijo glede na dolžino komentarja



Vir: Lastna analiza.

Tudi v tem primeru je torej dolžina negativno povezana z oceno. Standardizacija tako poda bolj objektivni vpogled. Rezultati kažejo, da je v primeru standardizacije najbolj ocenjena gostinska ponudba, namestitev in mestno jedro sta ocenjeni v povprečju precej slabše (Slika 5).

Slika 6: Struktura vsebine komentarjev glede na prevladujoče čustvo



*Drugo označuje ocene, ki jih ni bilo mogoče jasno razvrstiti pod eno čustvo.

Vir: Lastna analiza.

Komentarje smo ocenili še z vidika prevladujočega čustva. Analiza prevladujočega čustva v tekstu se izvede s pomočjo obsežnega leksikona NRC (National Research Council Canada) ter algoritma Naive Bayes, ki pa ne ocenjuje samo posamezne besede, ampak oceni stavek kot celoto. Rezultati (Slika 6) kažejo, da so gostje največkrat izražali zadovoljstvo v tekstovnih ocenah namestitve, dobrih 86 % vseh ocen je namreč izražalo zadovoljstvo. Sledila je gostinska ponudba, kjer je okoli 76 % vseh ocen izražalo zadovoljstvo. Z vidika prevladujočega čustva v negativnem smislu se je najslabše odrezala kulturna ponudba, saj je največji del ocen izražal jezo. Vendar pa je bil ta odstotek vseeno zelo nizek (6%). Analiza pokaže tudi, da obstaja značilna razlika v porazdelitvi čustev po skupinah ($p=0,000$).

Gosti so v povprečju z vsemi analiziranimi vidiki ponudbe v Ljubljani zadovoljni. Številčne ocene, ki sicer lahko zavzemajo vrednosti od 1 do 5, kažejo, da so najbolj zadovoljni z mestnim jedrom, vendar pa so razlike v številčnih ocenah pravzaprav zelo majhne. Hkrati pa se čustvene ocene, torej to, kar bralec ocene razbere »med vrsticami« močno razlikujejo od številčnih ocen, tako po razlikah med posameznimi vrstami ponudb kot tudi po razvrščanju vrst ponudb. Gosti so tako najbolj navdušeno spregovorili o hotelih, sledi 'hrana' (restavracije, lokali), nato kulturno-zgodovinska ponudba in šele na koncu mestno jedro. Torej velja, da je vtis, ki ga bralec dobi, ko bere to množico komentarjev, v povprečju najboljši, kadar prebira ocene hotelov. Bistveno slabši pa je vtis, ko bere komentarje pri znamenitostih, ki jih lahko vidi v starem mestnem jedru.

Hoteli so najbolj ocenjeni. Recenzija, ki je bila najbolj nabita s pozitivnimi čustvi, tako navaja: *»Kaj je najboljše? Lokacija. Osebj. Postelja. Dodatna ponudba. Zajtrk. Vtis, ki ga dobiš ob prijavi in odjavi. (...) Hotel je zagotovo najboljši hotel s štirimi zvezdicami v Evropi. (...) za svoje goste zelo skrbijo, so zelo prijazni in dajo občutek dobrodošlosti, kar pogrešam pri številnih hotelih v Evropi (...).«* V podobnem tonu navdušeni uporabnik opisuje ljubljanski hotel.

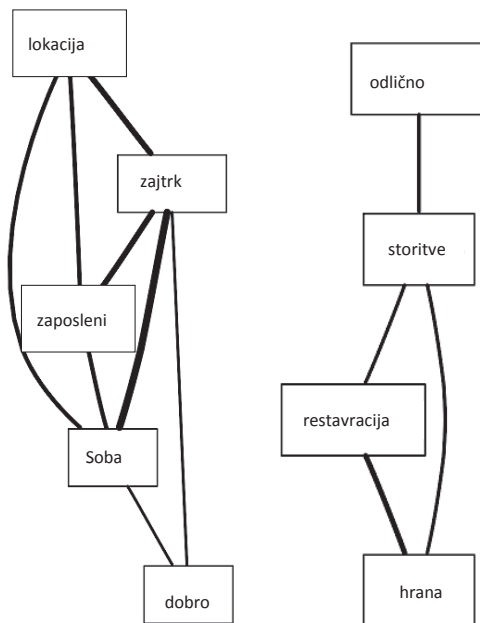
Dobro so bile ocenjene tudi restavracije. Z največ navdušenja po čustveni vrednosti je bila sprejeta restavracija, v kateri, kot pravi avtor, so vedno izjemno dobro postreženi, ne glede na to, ali pridejo v poslovni obleki ali v jeansu. Doda pa še: *»Odkar smo prvič prišli v Restavracijo (ime izbrisano), je bila storitev vedno izjemna. Priporočila osebja so izjemna, deležni pa ste dobrega odnosa ne glede na to, ali ste v poslovni obleki ali v jeansu. Hrana v tej restavraciji je po mojih izkušnjah najbolj sistematično izjemna, pa sem imel možnost jesti v številnih dobrih restavracijah (tudi z Michelinovo zvezdico) po vsem svetu. Jedi so pogosto enostavne, a izjemno okusne in vedno lepo prezentirane (...).«*

Vsebinski poudarki v tekstovnih ocenah. Z vidika uporabnika je poleg same ocene in pa čustvene ocene, ki se jo lahko prebere med vrsticami, zelo pomembna tudi vsebina. Vsebinska je izjemno pomembna ne samo za potencialne obiskovalce, temveč predvsem tudi za managerje različnih tipov ponudbe. Iz tekstovnih ocen lahko s pomočjo metod tekstovnega rudarjenja razberemo glavne vsebinske poudarke in tako povzamemo vsebinske poudarke nekaj 100 ali celo nekaj 1000 dolgih tekstovnih ocen, kar bi sicer zahtevalo veliko časa. Vsebine ocen smo povzeli s pomočjo analize ključnih besed, LDA analize ter analize povezav med besedami.

Slika 7 povzema ključne povezave, ki so najbolj značilne v tekstu. V analizo smo vključili naenkrat vse ocene in skušali izločiti ključne vsebinske povezave. S pomočjo funkcij v *tm* paketu programa R smo pripravili matriko ključnih besed ter dokumentov, kjer se te

ključne besede pojavljajo, nato pa je izračunana korelacija, ki pove, kako pogosto se besede nahajajo skupaj v istem dokumentu. Rezultati lepo kažejo to, kar potrdijo tudi ostale metode vsebinske analize. Če ocenjujemo posamezen tip ponudbe, ocene običajno ocenijo tiste temeljne vidike. Pri hotelu so to soba, osebje, zajtrk. Pri teh je pomembna prijaznost osebja, čistoča sobe in seveda kvaliteta hrane. Gosti v Ljubljani so v povprečju zelo zadovoljni s hoteli, pohvalijo tako sobe, lokacijo, kot predvsem tudi osebje, prijaznost, hrano v hotelu, čistočo (glej tudi analizo ključnih besed, Tabela 3). Hitra analiza ključnih besed kaže namreč, da recenzije v povprečju govorijo največ ravno o sobah, zajtrku, osebju. V povezavi s temi besedami se pogosto pojavi beseda »dobro«, ki je za sobo pravzaprav druga najbolj pogosta. To seveda opozarja na to, da so nam gostom, tudi ko podajamo svoje mnenje naprej, sobe resnično pomembne. Gostje se torej v svojih ocenah dotikajo predvsem temeljnih značilnosti ponudbe. Zelo redko je v oceni mogoče zaslediti diskusijo o tem, ali ima hotel savno, fitness, kakšna je kava v bližnji okolici, ali je ob hotelu lep park, itd.

Slika 7: Najmočnejši vsebinski vzorci v tekstovnih ocenah



Vir: Lastna analiza.

V mestu, kot je Ljubljana, za razliko od ostalih lokacij (npr. morje), je izraziteje v ospredju še lokacija hotela (ob morju si običajno takšnega izberemo oziroma so ti hoteli običajno ob obali). Tudi v tem vzorcu so besede »lokacija, mesto, Ljubljana« tako visoko pozicionirane v ocenah hotelov, in sicer so med prvimi desetimi najbolj pogostimi. V Ljubljani je lokacija hotela pomembna predvsem zato, ker si turisti želijo ogledati staro mestno jedro z vsemi znamenitostmi. Zato so večinoma tudi prišli. Kaj jih najbolj pritegne? Grad in Tromostovje, izpostavljajo, da je lep razgled in lep pa je tudi sprehod na grad. Poudarjajo tudi, da je mestno jedro nasploh lepo in prijetno. Vendar pa tudi najboljše recenzije niso navdušujoče, v povprečju so bistveno bolj hladne od ocen hotelov. Tako ima najvišjo

čustveno vrednost ocena mesta, ki v večjem delu sicer zelo natančno in korektno opiše številne znamenitosti Ljubljane, a brez presežkov, pri tam pa zaključi, da tudi »če je vreme slabo, lahko najdete suho zatočišče v številnih muzejih in galerijah. Kakorkoli že, ugotovili boste, da je Ljubljana eno najbolj navdušujočih, sofisticiranih in vizualno navdušujočih manjših glavnih mest v Evropi. Postavite ga na svoj radar raje malo prej kot kasneje.«

V ocenah se je izrazil tudi vzorec ocenjevanja restavracij (Slika 7). Tudi tu so ključni osebje, hrana in pa sama postrežba. Sicer pa so gosti dobro ocenili restavracije. Bolj podrobna analiza tem, ki se pojavijo v ocenah, kaže, da prevladujeta dve temi: ena je bolj splošna in hvali restavracijo samo, ki jo priporočajo zaradi dobre (odlične) hrane. Poleg tega pa je izrazita tudi prava »gurmanska« tema z ocenami vin, jedi (hodov), osebja v restavraciji. Glede na to, da je v Ljubljani kar nekaj restavracij, kamor zahajajo »foodiji«³, tak poznavalski pristop v ocenah ne preseneča (analiza ključnih besed, Tabela 3).

Kar nekaj navdušenja požanje tudi grad. Tako najbolj navdušeni obiskovalec pravi, da je »grad atrakcija št.1 v mestu. (...) Je odlična lokacija za obisk tako v dobrem kot tudi v slabem vremenu. (...) Sprehod skozi park do gradu je sicer strm, vendar super. (...) Zanimive so tudi slike in kapelica, kjer pari lahko sklepajo zakonske zveze.(...). V celoti ocenjujem, da je bil ogled vreden 8 evrov, kolikor je bila vstopnina, čeprav bi lahko večino stvari, ki sem jih, videl tudi brezplačno.« No, pa niso vsi ravno navdušeni. Najbolj kritična ocena je izpostavila, da je »v Evropi veliko gradov in da je ta zagotovo najslabši, kar sem jih obiskal. Grad je mišmaš različnih arhitekturnih obdobj s teraso in kavarno na sredini. Uničili so ga.«

Tabela 3: Analiza ključnih besed

Mestno jedro	Gostinska ponudba	Namestitev	Kultura
grad	hrana	soba	muzej
mesto	dobra	zajtrk	Ljubljana
Ljubljana	restavracije	dobro	ogled
center	postrežba	osebje	zgodovina
star	super	hotel	mesto
lep	Ljubljana	mesto	hrana
super	vino	lokacija	resnično
most	kraj	Ljubljana	super
reka	odlično	super	zanimivo
sprehod	meni	lepo	Slovenija
kraj	kvalitetno	bivanje	dobro
lep	obrok	čisto	time
obiskati	prijazen	prijazni	vodič
dobro	osebje	odlično	videti
razgled	natakar	dobro	obisk
restavracije	večerja	udobno	razstava

Vir: Lastna analiza.

³ Izraz se uporablja za opis posameznika, ki rad uživa v dobri hrani, gurmana.

4 SKLEP IN RAZPRAVA

Spletne platforme, kakršne so Tripadvisor, Booking, Expedia, itd. so postali izjemno pomemben vir informacij, ki pomagajo potencialnemu gostu oblikovati sliko turistične destinacije (*angl.* tourist destination image, glej npr. Gunn, 1988). Gost je tako postal nekdo, ki na eni strani oblikuje svojo podobo turistične destinacije zase, hkrati pa pomembno vpliva na oblikovanje te podobe drugim. S tem se je proces potrošniškega odločanja precej spremenil, saj smo postali potrošniki pravzaprav eden najbolj aktivnih členov v verigi oblikovanja podobe o proizvodu (ne le turističnem, spletne ocene so pomemben vir informacij za več kategorij izdelkov, glej npr. Rick, 2013).

V Sloveniji je turizem precej pomembna panoga, saj neposredno prispeva okoli 3,5 % k BDP, s posrednimi vplivi skupaj pa kar okoli 12 %. Zato je seveda izjemno pomembno, da se sektor uspešno razvija. Glede na to, da s prebiranjem spletnih komentarjev postajamo posamezniki pravzaprav bolj racionalni potrošniki, pravi mali »homo economicus«, ki se trudimo maksimirati svojo korist in torej dobiti največ za svoj denar, je pomembno, da se podobno racionalno obnašajo tudi ponudniki. In da zato skušajo izluščiti iz ocen bistvo, torej to, kar je njihova konkurenčna prednost in pa seveda odpraviti glavne težave.

Razprava. V članku smo analizirali okoli 9000 spletnih ocen, ki se dotikajo različnih tipov ponudbe ljubljanskega turizma. Nasploh lahko rečemo, da so gosti z Ljubljano zadovoljni, v povprečju je številska ocena 4,36. Upošteva je zgolj številsko oceno so gosti najvišje ocenili mestno jedro, sledi gostinska ponudba, nato pa namestitev in kulturna ponudba. Ker pa številska ocena ne odraža dobro tistega, kar preberemo med vrsticami ali pa preberemo neposredno v oceni, smo v nadaljevanju analizirali še vsebino samih ocen, in sicer s čustvenega vidika in z vidika vsebine same. Čustvene ocene kažejo, da pravzaprav številske ocene ne podajo dobre informacije o zadovoljstvu gostov. Ocena, ki jo izražajo čustva, in jo lahko skozi nabor besed preberemo med vrsticami, lahko kaže ali zadovoljstvo ali navdušenje ali jezo ali bes. Moč čustev, ki je zajeta v oceni in je bila izmerjena s čustveno analizo, kaže, da so gosti najbolj zadovoljni z gostinsko ponudbo (standardizirana vrednost), sledi namestitev, nato mestno jedro in kultura. Če se osredotočimo samo na čustva, ki vejejo iz posameznih ocen, je v povprečju zopet mogoče ugotoviti, da je večina gostov Ljubljano zapustila zadovoljnih, saj v povprečju 73 % ocen izraža zadovoljstvo. Če bi ocenjevali tipe ponudbe le po prevladujočem čustvu, bi bili zmagovalci hoteli, saj kar 86 % ocen izraža zadovoljstvo.

Vsebinska analiza kaže, da gosti ocenjujejo predvsem osnovne kategorije nekega tipa ponudbe. Pri hotelu tako sobo, zajtrk, osebje, čistočo. Pri restavracijah so to hrana, postrežba in vino. Ocene se sicer razlikujejo v poudarkih, vendar pa je fokus na glavnih elementih sistematičen. To je seveda tudi tista informacija, ki jo bo povprečni potencialni potrošnik potreboval in želel. Hkrati pa je to zelo dragocena informacija za vse managerje. Namreč, kot kaže, se dodatna ponudba redko pojavlja kot plus v tekstovnih ocenah. Zato je z vidika uspeha izjemno pomembno vlaganje v temeljne elemente in odprava vseh morebitnih slabosti. Na primer, pri restavracijah se kot najbolj negativne besede ali besedne zveze pojavljajo »visoka cena, drago«, ogoljufan. V hotelih pa »ni sveže«, drago, umazan, slabo.

Rezultati kažejo tudi, da je pravzaprav po vseh načinih ocenjevanja v ozadju sistematično kulturni vidik, predvsem muzeji in galerije. Deloma zagotovo zato, ker je bilo teh ocen najmanj, bile pa so tudi kratke. Vendar tudi upoštevanje teh omejitev pušča kulturo na repu. Zato je z vidika razvoja ljubljanskega turizma potrebno narediti več tudi na tem področju, predvsem za bolj zahtevne goste.

Prispevki k razvoju področja. Članek kot prvi v Sloveniji analizira ocene turistične ponudbe na spletu za slovensko destinacijo s pomočjo tekstovnega rudarjenja v tako širokem obsegu oziroma s tako številnimi metodološkimi prijemi (*angl.* computational linguistics). Hkrati je tudi eden redkih v svetovnem merilu. Literatura s področja turizma sicer vedno pogosteje analizira vsebino spletnih strani, kot so Tripadvisor, Booking, Expedia in podobne, vendar pa se ali osredotočajo zlasti na čustveno oceno (npr. Kasper in Vela, 2012), ali temeljijo na manualni obdelavi in zato na manjšem številu analiziranih tekstovnih ocen (npr. Jeong in Jeon, 2008), se osredotočijo na uporabo ene metodologije (npr. Roseti e tal., 2015), pri čemer je vsebinsko modeliranje (*angl.* topic modelling ali content modelling) zelo redko. In nenazadnje, številni članki, ker gre pač za razvijajoče se področje znotraj računalništva in programiranja, uporabljajo tekstovne ocene, ker le-te omogočajo lep in nazoren prikaz rabe neke metodologije, pri čemer pa je v ospredju prikaz metode in ne dejanska analiza z vidika turistične panoge (npr. tudi Kasper in Vela, 2012). Pričujoči članek tako uporabi različne razpoložljive tehnike s področja tekstovnega rudarjenja z namenom celovite vsebinske analize izbranih raziskovalnih vprašanj.

Nadalje članek prispeva k razvoju znanstvenega diskurza na področju marketinga in turizma, predvsem tudi vloge družbenih omrežij z vidika vplivanja na potrošnika. Namreč, raziskave kažejo, da se je proces nakupnega odločanja z vzponom interneta in družbenih omrežij precej spremenil, posameznik ni več prejemnik informacij, ampak predvsem postaja, s tem ko deli svoje izkušnje, tudi dejavnik vpliva na odločanje drugih (Rick, 2013). Prvi korak k razumevanju možnih smeri in intenzivnosti vpliva je seveda vsebina samih ocen.

Ker je kupec oziroma ocenjevalec postal pomemben dejavnik soodločanja, je z vidika turističnih ponudnikov izjemno pomembno, da poznajo ocene, ki se jih dotikajo. S strateškega vidika je izjemno pomembno, da ponudnik pozna, kaj je ocenjeno pozitivno, kaj se izpostavlja kot slabo. Tudi literatura že izpostavlja pomen analize spletnih ocen z vidika konkurenčne analize. Ob predpostavki, da posamezniki, ki delujejo v neki dejavnosti, poznajo svojo konkurenco, so tovrstne informacije tudi izjemno dragocen vir podatkov za konkurenčno analizo (npr. Memarzadeh in Chang, 2015, Gemar in Jimenez-Quintero, 2015). Tako na ravni posameznega ponudnika kot tudi panoge, regije, mesta. Posploševanje (npr. konkretno teh) rezultatov pa seveda ni možno, saj so ocene odvisne od številnih specifik, ki so vezane ali na lokacijo, tipe ponudbe, itd.

Tovrstna analiza pokaže, da je z vidika ponudnika pomembno več stvari. Najprej, tovrstna analiza omogoča vpogled v vsebino ocen, kar je brez dvoma pomembno za izboljšanje ponudbe in s tem dolgoročni poslovni uspeh. Drugič, pokaže se, da številna ocena ni merodajna. Če predpostavimo, da posamezniki (potencialni potrošniki) berejo tekstovne ocene in se na podlagi njih odločajo (in Bassig, 2013, kaže da več kot 90% ljudi to počne), je pomembno, kakšen je ton napisanih ocen (tega poda čustvena vrednost) in pa seveda vsebina. Posamezniki (potencialni potrošniki) smo nagnjeni k branju daljših ocen, ki pa

so ponavadi bolj negativne (Godnov in Redek, 2015a). Predvsem pa se morajo turistični ponudniki zavedati, da številna ocena ni dovolj, da se morajo posvetiti tudi analizi vsebine.

Omejitve raziskave. Članek posega na področje, ki se trenutno hitro razvija, je pa precej neraziskano, s čimer pa odpira tudi vrsto raziskovalnih izzivov. Eden ključnih izzivov je objektivnost ocen in pa problem »lažnih« ocen (angl. fake reviews) (npr. Mayzlin in ostali, 2014). Pogosto se pojavi očitke, da določeni ponudniki investirajo resurse v izboljševanje ocen, veliko pa se jih vsaj aktivno odziva na očitke, s tem ko posameznim ocenjevalcem podajajo odgovore. Vendar pa meniva, da je ta problem prisoten bolj na področju analize posameznega ponudnika, manj pa v primeru takšne širše analize mesta, regije, ipd., saj je podatkovna baza bistveno večja in pa hkrati se morebitne manipulacije skrijejo v povprečjih. Naslednji problem je kvaliteta uporabljenih leksikonov. Pričujoča analiza temelji na trenutno najbolj naprednih leksikonih, vendar pa se področje ves čas razvija, zato bo morda analiza čez eno leto pokazala drugačne poudarke. Potencialni problem predstavlja tudi prevod, vendar pa študije kažejo, da so prevodi izjemno kvalitetni (več o problemu strojnega prevajanja v Fiederer and O'Brian, 2009) in ne predstavljajo večjega problema pri zanesljivosti rezultatov. Velik izziv predstavlja tudi segmentacija potrošnikov na podlagi kategorij ocenjevalcev po spolu, starosti, interesih, kar bo dragocen input v analizo panoge, potrošnika pa tudi ponudnika na področju turizma.

Članek je tako predstavil rezultate aplikacije tekstovnega rudarjenja na tekstovnih ocenah. S tem je na eni strani izpostavil temeljne ugotovitve o različnih tipih ponudbe v Ljubljani, na drugi strani pa je ilustriral uporabnost tekstovnega rudarjenja v poslovni analizi, konkretno v turizmu.

LITERATURA IN VIRI

Bassig, M. (2013). 'TripAdvisor Reveals Key Travel Trends in TripBarometer.' <http://www.reviewtrackers.com/tripadvisor-reveals-key-travel-trends-tripbarometer/> (Pridobljeno 2. maja, 2015).

Bjorkelund, E., Burnett, T.H & Nørøvåg, K. (2012) A Study of Opinion Mining and Visualization of Hotel Reviews. Presented at iiWAS2012 Conference, 3-5 December, 2012, Bali, Indonesia. <https://www.idi.ntnu.no/~noervaag/papers/iiWAS2012.pdf> (Pridobljeno 5. junija, 2015).

Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), pp.77-84. <http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf> (Pridobljeno 7. maja, 2014).

Chen E. (2011). Introduction to Latent Dirichlet Allocation. *Microsoft research*. <http://blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/> (Pridobljeno 7. maja, 2014).

CNN. (2015). 10 most popular cities for travelers in 2015. <http://edition.cnn.com/2015/06/03/travel/mastercard-top-10-destination-cities-2015/> (Pridobljeno 3. decembra, 2015).

European Best Destination. (2015). <http://www.europeanbestdestinations.com/top/europe-best-destinations-2015/> (Pridobljeno 3. decembra, 2015).

European capital city tourism. (2012). Report: Analysis and finding. Roland Berger Consultants. https://www.rolandberger.com/media/pdf/Roland_Berger_European_Capital_City_Tourism_20120127.pdf (Pridobljeno 10. januarja, 2016).

Feinerer, I. (2015). Package »tm«. <https://cran.r-project.org/web/packages/tm/tm.pdf> (Pridobljeno 10. oktobra, 2015).

Feinerer, I. & Hornik, K. (2015). Introduction to the RKEA Package. <https://cran.r-project.org/web/packages/RKEA/vignettes/kea.pdf> (Pridobljeno 10. oktobra, 2015).

- Fiederer, R. & O'Brien, S. (2009). Quality and Machine Translation: A realistic objective? *The Journal of Specialized Translation*, 11, 52–74.
- Garay Tamajón, L. & Cànoves Valiente, G. (2015). Barcelona seen through the eyes of TripAdvisor: actors, typologies and components of destination image in social media platforms. *Current Issues in Tourism* (published on-line July 2015).
- Gemar, G. & Jimenez-Quintero, J.A. (2015). Text mining social media for competitive analysis. *Tourism and management studies*, 11(1), 84-90.
- Global report on city tourism. (2015). http://dtxqt4w60xqpw.cloudfront.net/sites/all/files/pdf/am6_city_platma.pdf (Pridobljeno 10. januarja, 2016).
- Godnov, U. and Redek, T. (2015a). Application of text mining in tourism: case of Croatia. Fakulteta za management, UP, in Ekonomska fakulteta, UL. Mimeo.
- Godnov, U. and Redek, T. (2015b). 'Croatian tourism evaluation: the lessons from the on-line peer assessments using text mining.' Presented at the CONFET2015 conference, May 2015, Pula, Croatia.
- Gruen, B. in Hornik, K. (2011). Topicmodels: An R Package for Fitting Topic Models. <http://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf> (Pridobljeno 4. maja, 2014).
- Gunn, C. (1988). *Vacationscapes: Designing tourist regions*. New York: Van Nostrand Reinhold.
- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. National Conference on Artificial Intelligence. <https://www.cs.uic.edu/~liub/publications/aaai04-featureExtract.pdf> (Pridobljeno 4. maja, 2014).
- Jacobsen, G.D. (2015). Consumers, experts, and online product evaluations: Evidence from the brewing industry. *Journal of Public Economics*, 126, 114-123.
- Jeong, M., & Jeon, M.M. (2008). Customer reviews of hotel experiences through consumer generated media. *Journal of hospitality and leisure marketing*, 17(1/2), 121–138.
- Kasper, W. & M. Vela. (2012b). Sentiment Analysis for Hotel Reviews. In Jassem, K., Fuglewicz, P., Piasecki, M. & Przepiorkowski, A. (Eds.), *Proceedings of the Computational Linguistics-Applications Conference (CLA-2011)* (pp. 45–52). Jachranka, Oct. 2011.,
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. May 2012. <http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf> (Pridobljeno 4. maja, 2014).
- Memarzadeh, F. & Chang, H.J.J.C. (2015). Online Consumer Complaints About Southeast Asian Luxury Hotels, *Journal of Hospitality Marketing & Management*, 24(1), 76–98. DOI: 10.1080/19368623.2014.893222.
- Mohammad, S. (2015). NRC Word-Emotion Association Lexicon. <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> (Pridobljeno 8. oktobra, 2015).
- Rebernik, A. (2014). Turisti so nad Ljubljano navdušeni, tudi zato, ker ne vedo, kam prihajajo. http://www.siol.net/trendi/potovanja/novice/2014/07/turizem_ljubljana.aspx (Pridobljeno 10. oktobra, 2015).
- Rick, T. (2013). The consumer decision making process has permanently shifted. <http://www.torbenrick.eu/t/r/kre> (Pridobljeno 8. maja, 2015)
- Rossetti, M., Fabio, S., Cao, L. & Zanker, M. (2015). Analysing User Reviews in Tourism with Topic Models. V Tussyadiah, I., Inversini, A. (ur.), *Information and Communication Technologies in Tourism 2015* (str. 47–58). Cham: Springer.
- Statistični letopis Ljubljane. (2013). Mestna občina Ljubljana. <http://www.ljubljana.si/si/ljubljana/ljubljana-v-stevilkah/statisticne-publikacije/> (Pridobljeno 10. oktobra, 2015)
- Statistični Urad Republike Slovenije. (2013). Nastanitvene zmogljivosti, prihodi in prenočitve, Slovenija. Metodološko pojasnilo. <http://www.stat.si/statweb/Common/PrikaziDokument.aspx?IdDatoteke=7779> (Pridobljeno 10. januarja, 2016)
- Tripadvisor. (2015). Ocene uporabnikov.
- Vasquez, C. (2011). Complaints online: The case of TripAdvisor. *Journal of pragmatics*, 43(6), 1707–1717.
- Vprašalnik za statistično raziskovanje Prihodi in prenočitve turistov, za leto 2015. Statistični Urad Republike Slovenije, 2015. <http://www.stat.si/StatWeb/Common/PrikaziDokument.aspx?IdDatoteke=8794> (Pridobljeno 10. januarja, 2016)