

ISTRAŽIVANJE BRANDA KORIŠTENJEM ANALIZE TEKSTA I KONTEKSTA

BRAND RESEARCH USING TEXT AND CONTEXT ANALYSIS

Sara Slamić Tarade

Tehničko veleučilište u Zagrebu

SAŽETAK

U ovom radu prikazala sam proces modeliranja teme (topics modeling) zasnovan na alatima za analizu tekstualnih sadržaja primjenom obrade prirodnog jezika za potrebe komunikacije u odnosima s javnosti ili marketingu. U radu su obrađene osnovne karakteristike analize teksta i konteksta primjenom algoritama strojnog učenja i obrade prirodnog jezika na tekstualnim sadržajima skinutih sa web stranica tehnikom web „grebanja“. Korištenjem odgovarajućih programskih alata napravila sam analizu pojavnosti i važnosti ključnih pojmova vezan uz brandove sportske obuće. Slikovito sam prikazala proces obrade navedenih sadržaja i rezultate koji su dobiveni provedenom analizom. Rezultati pokazuju da je moguće bez ikakvog prethodnog označavanja teksta dobiti informacije o kojim temama se radi vezanim za brandove i u kojem kontekstu. Prikazana metodologija istraživanja omogućuje da obradom velike količine tekstualnih sadržaja iz raznih digitalnih izvora automatiziranim postupkom dođemo do korisnih informacija o značaju brandova od interesa.

Ključne riječi: *analiza teksta, NLP metode, brand, algoritmi*

ABSTRACT

In this paper, I have presented the topic modeling process based on tools for textual content analysis using natural language processing for public relations or marketing purposes. The paper deals with the fundamental characteristics of text and context analysis using machine learning algorithms and natural language processing on textual content downloaded from web pages using the web scraping technique.

Using appropriate software tools, I analyzed the incidence and importance of key terms related to sports footwear brands. I have pictorially presented the process of processing the mentioned contents and the results obtained by the conducted analysis. The results show that it is possible to acquire information on which topics to refer to brands and in which context without any prior labeling of the text. The presented research methodology enables us to obtain useful information about the importance of brands of interest by processing a large amount of textual content from various digital sources in an automated way.

Keywords: *text analysis, NLP methods, brand, algorithms*

1. UVOD

1. INTRODUCTION

Analiza sadržaja ima za cilj izvlačenje relevantnih informacija vezanih uz neki problem, brand, stanje na tržištu, aktivnosti konkurencije i sl. Kada se govori o analizi sadržaja uglavnom se misli na nestrukturirane podatke iz kojih nije moguće eksplicitno uporabom standardnih upita dobiti informaciju za razliku od strukturiranih sadržaja i podataka u pravilu organiziranih kao tablice. Nestrukturirani sadržaji nalaze se u različitim medijima kao što su knjige, članci, novinski sadržaji i vijesti, mišljenja korisnika, filmovi, video sadržaji i drugo i moguće ih je susreti kako u klasičnim papirnatim oblicima, filmskim i drugim analognim zapisima s jedne strane i u digitalnom obliku s druge strane. Sadržaji u digitalnom obliku danas imaju primat i količina tih nestrukturiranih sadržaja mjeri se danas u desetinama Zetabajta.

U ovom radu naglasak će biti na digitalnim sadržajima. Ovisno o vrsti digitalnih sadržaja moguće je koristiti različite metodologije i tehnike same analize.

Cilj ovog rada je priprema za analizu sadržaja vezanih za položaj nekih brendova na tržištu u kontekstu definiranja komunikacijskih strategija i temelji se na tematskom modeliranju (engl. topics modeling) kao priprema za dublje semantičke analize koje se zasnivaju na inovativnoj metrici ocjene značaja brendova (engl. Semantic Brand Scores) [1] primjenom tehnika i metoda obrade prirodnog jezika (engl. NLP - Natural Language Processing) i teorije grafova (mreža). Navedena metrika nije razmatrana u ovom radu. Za provedbu analize koriste se različiti izvori tekstualnih dokumenata, poput članaka u novinama, web stranica, e-mailova, Tweetova, Facebook objava, mrežnih foruma, blogova i sl.. S obzirom da su većina ovih tekstualnih sadržaja spontani izrazi potrošača, moguće je smanjiti pristranost kojoj mogu doprinijeti ciljane ankete.

Analiza teksta temelji se na kvalitativnim metodama za prikupljanje, predobradu, klasifikaciju i interpretaciju tekstualnih sadržaja. Ove analize mogu relativno brzo, iz različitih tekstualnih sadržaja, dati relevantne informacije o nestrukturiranim podacima, odnosno dati odgovore koji su predmet analize.

U današnjim okolnostima kada digitalni sadržaji rastu gotovo eksponencijalno, značenje nestrukturiranih sadržaja predstavlja glavni izazov u raznim područjima, od socijalnih političkih, poslovnih do akademskih istraživanja. Klasično ručno čitanje, označavanje tema, konteksta i tona i analiza sadržaja nije više moguća jer bi to predugo trajalo s obzirom na količine podataka, a rezultati bi stoga bili često zastarjeli i neprimjenjivi i nije moguće osigurati njihovu skalabilnost. Iz tog razloga primjenjuju se tehnike i metode automatizirane analize sadržaja kroz analizu teksta i konteksta.

Navedena tematika ovog rada obrađena je kroz sljedeće točke: analiza teksta u kojoj će se definirati glavni elementi te analize, analiza konteksta koji daje ton sadržaju odnosno temi koja se istražuje, metodologija istraživanja i provedba na sadržajima iz različitih izvora, diskusija rezultata i zaključak.

2. ANALIZA TEKSTA

2. TEXT ANALYSIS

Analiza tekstova (engl. text mining) predstavlja razumijevanje kulturnih, ekonomskih, socijalnih ili ideoloških aspekata sadržanih u suštini nekog teksta i kako su oni povezani s određenim kontekstom. Analiza tekstova podrazumijeva opisivanje karakteristika teksta i interpretaciju u cilju dobivanja odgovora na neka konkretna pitanja. Metodologija analize teksta danas se široko koristi kako u akademskim istraživanjima tako i raznim studijama medija, marketinga i komunikacije, kulture, filozofije i ekonomije.

U današnje doba ekstremne količine tekstualnih podataka, potrebno je objediniti razna znanja i vještine koje će omogućiti transformaciju složenih i velikih podataka (engl. big data) u korisne i upravljive informacije. Zbog toga je proces analize sadržaja teksta nezamisliv bez pomoći računalnih sustava i alata koji omogućuju automatizaciju procesa, pojednostavljuju promatranje obrazaca i kategorizaciju sadržaja prema prethodno definiranim kategorijama. Ta metodologija omogućuje izdvajanje različitih značajki sadržaja kao što su imena, zemlje, e-mail adrese, autori, brandovi i drugo.

Analiza tekstualnih sadržaja može se provoditi s više aspekata ovisno o području i ciljevima istraživanja. Prema [2] „Tekstualna analiza kvalitativna je metoda koja se koristi za ispitivanje sadržaja u medijima i popularnoj kulturi, poput članaka iz novina, televizijskih emisija, web stranica, igara, videa i oglašavanja.“ Ova kvalitativna metodologija ispituje strukturu, sadržaj i značenje teksta i povezanost s povijesnim i kulturnim kontekstom u kojem je nastao i utemeljena je na semiotičkim, lingvističkim i interpretacijskim. S druge strane analizu teksta može se gledati kroz prizmu promišljanja teksta (engl. text mining) utemeljenoj na interdisciplinarnoj metodologiji pronalaženja informacija koristeći strojno učenje, statističke metode i računalnu lingvistiku [3]. U ovom radu neću se baviti kulturološkim aspektom analize teksta već samo u kontekstu primjene u odnosima s javnosti i marketingu. Promišljanje teksta možemo definirati na više načina kao što su izvlačenje informacija (engl. information extraction), promišljanje podataka

ili otkrivanje znanja (engl. knowledge discovery) [3]. Sama analiza sadržaja može se promatrati kao potkategoriju analize teksta u cilju sustavne analize teksta koristeći tehnike označavanja (engl. abelling) elemenata teksta radi izvlačenja kvantitativnih pokazatelja.

3. ANALIZA KONTEKSTA

3. CONTEXT ANALYSIS

Alat komunikacije je jezik koji kroz različite tekstualne sadržaje izražava cjelovitu ideju komunikatora. Komunikacija tekstem između komunikatora vrlo često se ostvaruje kontekstom koji se može razvrstati na ko-tekst ili suradnički tekst, situacijski kontekst i kulturološki kontekst [4] i na tu temu lingvisti su napravili mnoštvo različitih istraživanja o odnosu teksta i konteksta.

Prema [5] „kontekst je a) govorna ili tekstualna cjelina koja točno određuje smisao pojedine riječi (ili druge jezične jedinice) koja je i sama njezin dio, ili b) tekst ili dio teksta u kojem se neka riječ pojavljuje u konkordanciji ili korpusu“.

Tekst se može opisati modelom kojeg čine osnovne leksičke jedinice riječi i interpunkcijski znakovi. Za označavanje tih tekstualnih jedinica koristi se izraz „token tekstualnih jedinica“. Tokeni se u nekom tekstu nalaze u određenom redosljedom pa se tekst može modelirati pomoću konačnog niza svih riječi i interpunkcijskih znakova. U tekstovima ima više različitih informacija strukturnog karaktera kao što su naslovi, odlomci i drugo, ali i značajke prikaza po kojima se razlikuju jedni dijelovi teksta od drugih.

Cjelinu tekstualnog modela uvijek se može postaviti u nekakav kontekst koji pobliže određuje smisao teksta. Kontekstualna analiza se može definirati određenim pravilima koja omogućuju identificiranje i po potrebi označavanje nekih dijelova koji prethode nekom pojmu (engl. left context) ili slijede nakon njega (engl. right context). Ako se stvar pojednostavi često neki pojam označen kao imenica uz sebe veže neki pridjev koju mu daje odgovarajući kontekst. Elementi kontekstualnih pravila mogu biti pozitivni i negativni pa to omogućuje analizu sentimenta vezanog za neki pojam, događaj ili brand [6].

Analiza sentimenta izlazi iz okvira ovog članka jer je to poseban problem koja se temelji na primjeni niza složenih algoritama strojnog učenja i obrade teksta prirodnog jezika pa stoga neće biti razmatran.

4. METODOLOGIJA ISTRAŽIVANJA

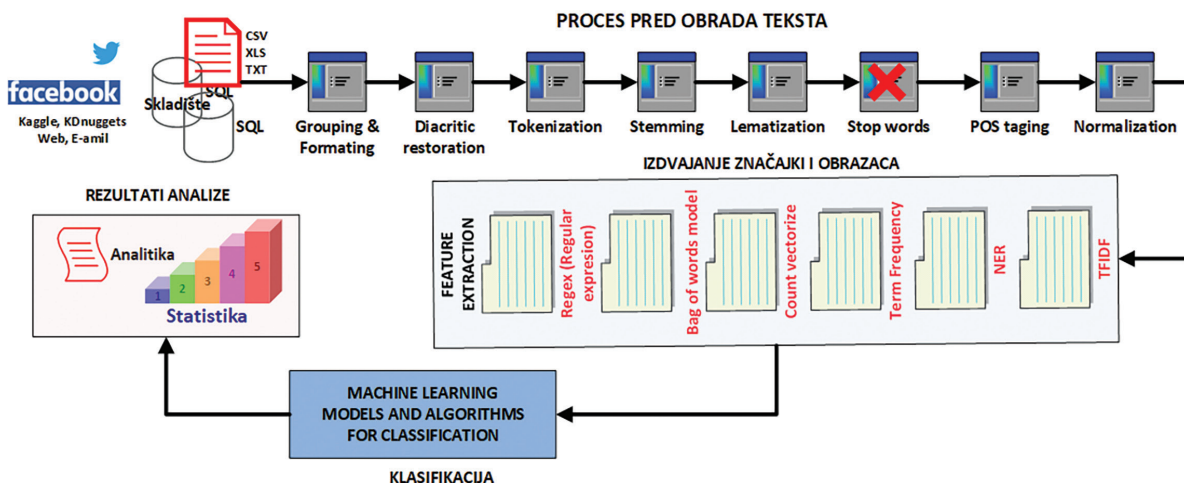
4. RESEARCH METHODOLOGY

Sadržaji se mogu prikupljati iz bilo kojih izvor tekstualnih dokumenata, poput članaka u novinama, web stranica, e-mailova, Tweetova, Facebook postova, mrežni foruma, blogova. Kako se radi o ogromnim količinama podataka taj posao nije moguće obavljati ručno već se i on mora automatizirati. Konačni cilj je prikupljanja sadržaja je stvaranje istraživačkog korpusa koji se sastoji od niza dokumenata. Analiza prikupljenih podataka temelji se na obradi teksta prirodnog jezika (NLP) koja podrazumijeva da je dokument odgovarajuća čestica teksta pripremljena za obradu. Dokumenti mogu biti pojedinačne rečenice, paragrafi i druge veće cjeline teksta.

Za prikupljanje sadržaja s weba (engl. news, službene stranice, e-mail) proces automatiziranog prikupljanja temelji se na takozvanom „grebanju“ sadržaja (engl. web scraping). Ova tehnika se može koristiti pod uvjetom da je pristup sadržajima otvoren i da pružatelj sadržaja nije ograničio pristup tzv. robotskim agentima. Etički bi bilo zatražiti dozvolu za dohvaćanje sadržaja na taj način.

Sadržaji s društvenih mreža (Twittera, Facebook) mogu se automatizirano prikupljati korištenjem tzv. API-a (engl. Application Programming Interface) koji mogu preuzimati poruke ili postove u stvarnom vremenu. Za korištenje API-a potrebno je registrirati se kao korisnik na Twitter ili Facebook i zatražiti ključ za pristupanje sadržajima.

Podaci sa Twitter skidaju se kao tokovi podataka (engl. stream) korištenjem navedenih API-a u stvarnom vremenu i omogućuju skidanje većeg broja sadržaja odjednom za razliku od korištenja standardnih web servisa pomoću REST API-a (tehnika pristupa web servisima). Slično vrijedi i za skidanje sadržaja sa Facebook-a i drugih društvenih mreža.



Slika 1 Proces analize teksta

Figure 1 Text analysis process

Treći pristup dohvaćanja sadržaja je korištenje različitih organizacija koje se bave prikupljanjem i obradom podataka s ciljem stvaranja znanja o podacima i koje najčešće registriranim korisnicima daju besplatan pristup njihovim bazama podataka i različitih tekstualnih sadržaja. Jedna od tih organizacija je Kaggle koji ima više od 19.000 skupova podataka iz različitih područja (marketing, komunikacije, poslovni svijet itd.) i 200.000 javnih bilježnica (notebooks) u kojima su prezentirani neki rezultati analize podataka ili tekstova iz različitih izvora. Drugi poznatiji pružatelj podatkovnih skupova je KDnuggets koji ima također bogate baze podataka različitih tekstualnih sadržaja pripremljenih za računalne analize i strojno učenje.

4.1 OBRADA TEKSTUALNIH SADRŽAJA PRIMJENOM NLP ALGORITAMA

4.1 PROCESSING TEXTUAL CONTENT USING NLP ALGORITHMS

Zahvaljujući razvoju strojnog učenja obrada zapisa prirodnog jezika (NLP) postala je vrlo značajna u svi područjima primjene od poslovanja, marketinga, komunikacije, medicine itd. U zadnjih nekoliko godina zahvaljujući razvoju tehnologija i algoritama strojnog učenja postignuti su značajni rezultati koji se široko koriste u različite svrhe.

Većina tih algoritama oslanja se na razne unaprijed pripremljene baze podataka ili sadržaja ili korpuse tekstualnih zapisa, no nažalost svi oni su razvijeni uglavnom za engleski jezik ili jezike velikih nacija. Korištenje tih algoritama na obradu tekstualnih zapisa na hrvatskom jeziku još uvijek je u začecima uglavnom iz razloga nedostatka kvalitetnog stručnog vokabulara (rječnika) i gramatičkih pravila u elektroničkom obliku. Za ovaj rad upotrijebljene su različite biblioteke za programske jezike Python ili R koje implementiraju algoritme obrade teksta pisanog na engleskom jeziku.

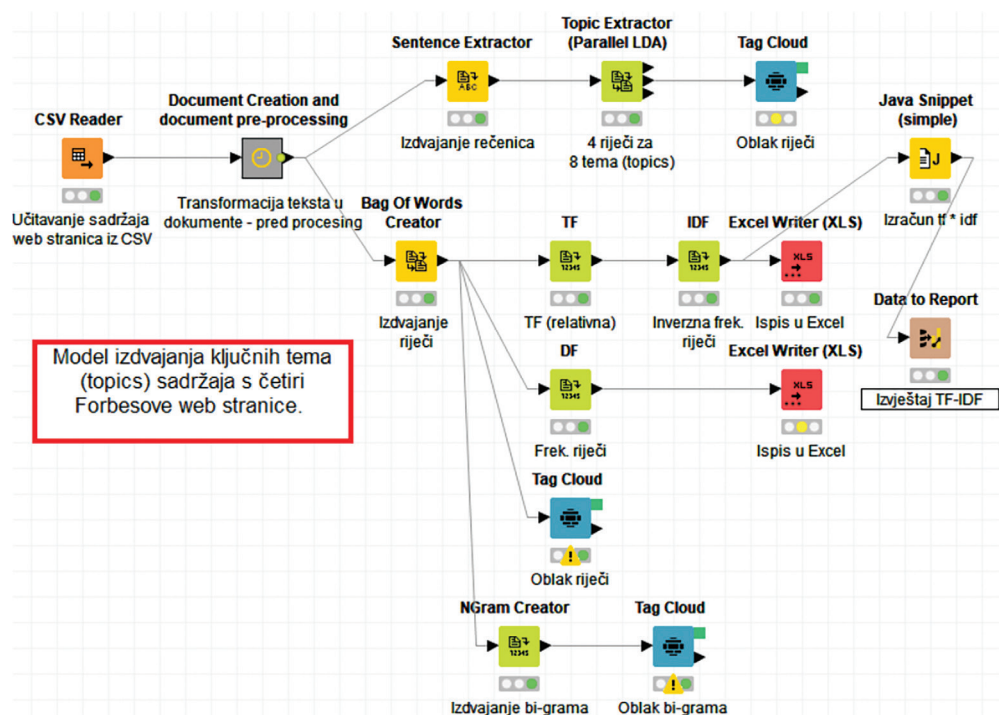
Korištenjem NLP algoritama nestrukturirane tekstualne podatke iz različitih izvora, najčešće pisanih kao slobodan tekst, pretvaramo u informacije i strukturirane podatke pogodne za daljnje analize kao što su različite klasifikacije, statističke analize i slično [7][8].

Proces analize teksta prikazan na slici 1 sastoji se od sljedećih cjelina: predobrada teksta (engl. preprocessing), izdvajanje značajki (engl. feature extraction), klasifikacija informacija (engl. ML model classification) i prezentacija rezultata analize.

4.2 ISTRAŽIVAČKI MODEL

4.2 RESEARCH MODEL

Za provedbu istraživanja korištena je Open Source aplikacija KNIME Analytics Platform [9] pomoću koje je napravljen model istraživanja.



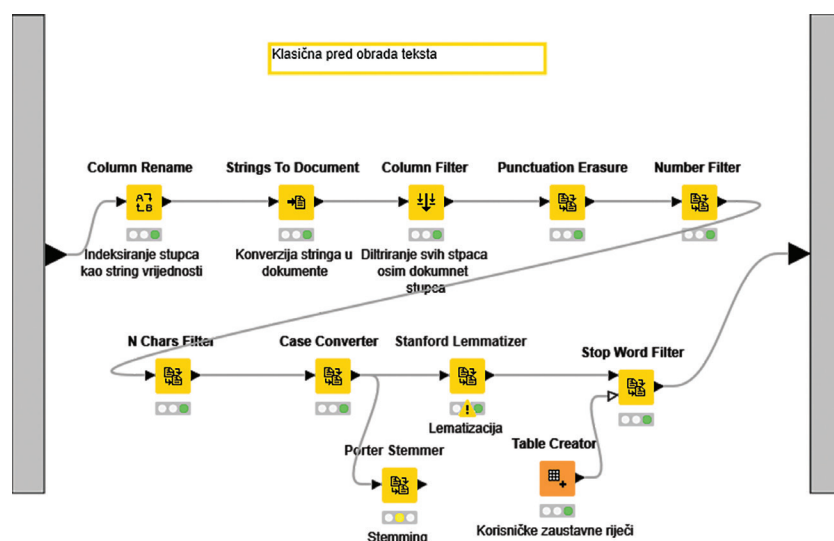
Slika 2 KNIME model istraživanja
Figure 2 KNIME research model

Snaga ove aplikacije je u brojnim proširenjima za provedbu statističkih analiza, strojnog učenja i obradu teksta. Pored toga aplikacija omogućuje i korištenje skripti pisanih u programskim jezicima Python i R radi proširenja mogućnosti modela. Aplikacija omogućuje razvoj modela bilo koje složenosti i testiranje na osobnom računaru, a u slučaju potrebe za velikim računalnim resursima zbog velike količine sadržaja koji se obrađuju moguće je isti model izvoditi na serverima tvrtke KNIME uz plaćanje.

Za potrebe istraživanja u ovom radu skinuti su tekstualni sadržaji s četiri Forbesove web stranice [10] s tematikom vezanom uz brandove sportske obuče.

Sadržaji su skinuti pomoću tehnike web „grebanja“ (engl. web scraping) i nakon toga izdvojeni su tekstovi zapisani u HTML tagu <body>, a zatim pohranjeni u CSV datoteku. Na slici 2 prikazan je model istraživanja razvijen pomoći aplikacije KNIME koji obuhvaća učitavanje tekstualnih sadržaja iz CSV datoteke, pred obradu te analizu primjenom tematskog modeliranja pomoću koje su izvučeni ključni pojmovi vezani uz neke brandove.

Sirove tekstualne sadržaje prije bilo kakve analize nužno je pripremiti i obraditi na način koji će biti razumljiv računalnim algoritmima koji moraju izvlačiti ciljane značajke koje se mogu dobiti iz teksta.



Slika 3 KNIME model pred obrade teksta
Figure 3 KNIME model text preprocessing

Proces pred obrade u KNIME aplikaciji (modul Document Creation and document pre-processing na slici 1) prikazan je na slici 3.

Najznačajniji koraci pred obrade teksta su: filtriranje sadržaja, izbacivanje interpunkcija i numeričkih podataka, pretvaranje svih tekstualnih jedinica (engl. „tokena“) pisanih malim slovima, normalizacija koja uključuje uklanjanje sufiksa i prefiksa iz korijena riječi (engl. stemming) i preslikavanje riječi na korijenski oblik (engl. lemmatization) i na kraju uklanjanje zaustavnih riječi koje nemaju neko bitno značenje u tekstu (engl. stopwords). Također ovisno o vrsti analize moguće je napraviti i gramatičko označavanje riječi (POS Tagger). Primjer nekih pojmova nakon pred obrade teksta prikazan je u tablici 1.

Iz tablice vidimo da kada izbacimo zaustavne riječi nestaju pojmovi 'for' i 'this' naravno i druge slične riječi kao 'I', 'am', 'the', 'a' i sl., a nakon izbacivanja sufiksa neke riječi se svode na osnovni oblik npr. 'personal' se pretvara u 'person', a ako bi u tekstu imali još riječi sa sličnim korijenom sve bi one bile pretvorene u 'person' i u određivanju frekvencije pojavnosti sve se broje kao jedna teista riječ. Ako gramatički označimo riječi tada možemo vidjeti koji su pojmovi imenice, glagoli ili pridjevi što nam može dodatno olakšati semantičku analizu teksta.

Nakon početne obrade teksta pristupilo se izvlačenju tematskih značajki obrađivanog teksta i dobiveni rezultati su diskutirani u sljedećoj točki.

Tokeni (nakon izbacivanja brojki i pretvaranja svih riječi u mala slova)	Izbacivanje zaustavnih riječi (stopwords)	Izbacivanje sufiksa i prefiksa (Porter Stemmer)	Gramatičko označavanje riječi (POS Tagger)
edit	edit	edit	edit[NN(POS)]
sneakers	sneakers	sneaker	sneakers[NNS(POS)]
for	women	women	for[IN(POS)]
women	spring	spring	women[NNS(POS)]
this	top	top	this[DT(POS)]
spring	picks	pick	spring[NN(POS)]
top	trendy	trendi	top[JJ(POS)]
picks	athletic	athlet	picks[NNS(POS)]
trendy	shoes	shoe	trendy[JJ(POS)]
athletic	forbes	forb	athletic[JJ(POS)]
shoes	personal	person	shoes[NNS(POS)]
forbes	shopper	shopper	forbes[NNS(POS)]
personal	contributor	contributor	personal[JJ(POS)]
shopper	opinions	opinion	shopper[NN(POS)]

Tablica 1. Primjeri pojmova nakon različitih koraka pred obrade teksta

Table 1. Examples of terms after preprocessing steps

5. DISKUSIJA REZULTATA

5. DISCUSSION OF RESULTS

Kako pokazuje model na slici 2 (modul DF) prvo su izračunate frekvencije pojavljivanja ključnih riječi vezanih uz neke brandove. Na slici 4. prikazano je 15 ključnih riječi s najvećom frekvencijom. Rezultati nam pokazuju da pojmovi vezani za brandove kao što su 'nike', 'sneaker', 'jordan' i 'nba' imaju najveće frekvencije pojavnosti (engl. prevalency) u istraživanim sadržajima.

Vidim da u prvih 15 riječi imamo i brand 'adidas' ali se on u analiziranim tekstovima puno manje spominje. Poznato je da je Nike kreirao brand Air Jordan za bivšeg igrača NBA lige i zato se ti pojmovi ističu. Slikovitije to možemo vidjeti na slici 5. koja predstavlja oblak riječi (engl. words cloud) gdje su pojmovi s većim značenjem više naglašeni.

Tablica 2. Rezultati LDA analize

Table 2. Results of LDA analysis

Tema_0		Tema_1		Tema_2		Tema_3	
Ključna riječ (term)	Težina	Ključna riječ (term)	Težina	Ključna riječ (term)	Težina	Ključna riječ (term)	Težina
jordan	88	sneaker	61	jordan	36	nike	79
nike	48	features	36	forbes	32	sneaker	48
million	40	shoe	34	nike	28	air	41
forbes	24	pair	33	annual	20	shop	32
Tema_4		Tema_5		Tema_6		Tema_7	
Ključna riječ (term)	Težina	Ključna riječ (term)	Težina	Ključna riječ (term)	Težina	Ključna riječ (term)	Težina
jordan	32	jordan	68	sneakers	45	nba	114
shoe	29	nike	64	sneaker	35	nike	74
sports	28	company	32	shoe	27	million	53
signature	24	adidas	30	forbes	24	players	52

Rezultati LDA analize dobiveni iz modula Topic Extractor (slika 2) prikazani su u tablici 2. Ovi rezultati potvrđuju prijašnje tvrdnje da su najčešći pojmovi vezani za brandove 'nike', 'jordan', 'sneaker'. Tako se pojam 'jordan' pojavljuje kao najznačajniji u temama 0, 2, 4 i 5, a pojam 'nike' se pojavljuje u temama 0, 2, 3 i 5.

5.1 OGRANIČENJA PROVEDENIH ISTRAŽIVANJA

5.1 LIMITATIONS OF THE RESEARCH

Analiza je provedena samo na četiri teksta koji se sastoje od nekoliko stotina rečenica i par tisuća riječi. Da bi se dobila što relevantniji rezultati potrebno je obraditi desetine gigabajta sadržaja (i do tisuću različitih tekstualnih sadržaja) no to zahtijeva jake računalne resurse. Provedena istraživanja na ovom malom uzorku pokazala su cjeloviti metodološki pristup i razvijeni KNIME model se za velike količine podataka može obraditi na zakupljenim serverima proizvođača softvera KNIME.

6. ZAKLJUČAK

6. CONCLUSION

U ovom radu dane su neke osnovne karakteristike analize teksta i konteksta primjenom algoritama strojnog učenja i obrade teksta prirodnog jezika na velikim izvorima tekstualnih sadržaja. Razlučeno je što je to sadržaj što je tekst i što je kontekst na osnovu nekih teorijskih razmatranja iz literature. Opisan je elementarni proces pred obrade teksta i pripremu istog za ciljne klasifikacije.

Kao primjer pokazani su samo rezultati analize teksta temeljem pojavnosti nekih ključnih riječi uzetih kao uni-grami i bi-grami. Primjeri pred obrade teksta i frekvencijske i LDA analize napravljeni su pomoću aplikacije KNIME koja za modeliranje koristi niz različitih gotovih biblioteka i modula. Ti rezultata su pokazani grafikonima frekvencije pojavljivanja pojmova vezanih uz neke brandove u tekstu koji je direktno skinut s web-a tehnikom „grebanja“. Relevantnost dobivenih rezultat frekvencijskom analizom pojavnosti pojmova potvrdila sam provedenom LDA analizom. Ove temeljene analize mogu se proširiti korištenjem složenijih algoritama za klasifikaciju informacija izvučenih iz teksta. Rezultati analize tematskog modeliranja pokazali su da je moguće iz neoznačenih tekstova izvući korisne informacije o značaju i pojavnosti brandova.

Kako bi se dobile još bolje informacije o značaju i položaju branda na tržištu u daljnjim istraživanjima treba primijeniti složenije algoritme za ocjenu branda gdje se koriste algoritmi iz teorije grafova (mreža) za semantičku procjenu branda (Semantic Brand Scores). Daljnje istraživanja također treba provesti na puno većim podatkovnim skupovima.

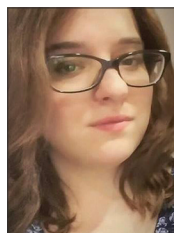
7. REFERENCE

7. REFERENCES

- [1.] Andrea Fronzetti Colladon, 2018, The Semantic Brand Score, Journal of Business Research Volume 88, Pages 150-160, <https://doi.org/10.1016/j.jbusres.2018.03.026>, 8.5.2020.

- [2.] Elfriede Fürsich, 2018, Communication Textual Analysis and Communication, DOI: 10.1093/OBO/9780199756841-0216, <https://www.oxfordbibliographies.com/view/document/obo-9780199756841/obo-9780199756841-0216.xml#obo-9780199756841-0216-bibItem-0001>, 14.06.2020.
- [3.] Hotho, Andreas & Nürnberger, Andreas & Paass, Gerhard. (2005). A Brief Survey of Text Mining. LDV Forum - GLDV Journal for Computational Linguistics and Language Technology. 20. 19-62.
- [4.] Lihong Shen, 2012, Context and Text, Theory and Practice in Language Studies, Vol. 2, No. 12, pp. 2663-2669, ISSN 1799-2591.
- [5.] Definicija, <https://jezikoslovac.com/word/x8ry>
- [6.] Dina WONSEVER, Jean-Luc MINEL, 2001, Contextual Rules for Text Analysis, Conference Paper, Lecture Notes in Intelligence Artificial, DOI: 10.1007/3-540-44686-9_52.
- [7.] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saied Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys Kochut. 2017. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques . In Proceedings of KDD Bigdas, Halifax, Canada, arXiv:1707.02919v2.
- [8.] Falguni N. Patel, Neha R. Soni, 2012, Text mining: A Brief survey, International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970), Volume-2 Number-4 Issue-6.
- [9.] KNIME Analytics Platform, <https://www.knime.com/software-overview>, 10.6.2020.
- [10.] url1 do url4: <https://www.forbes.com/sites/forbes-personal-shopper/2019/05/28/best-sneakers-for-women-this-spring-our-top-picks-for-trendy-athletic-shoes/#2e6d6d7e7b95>, <https://www.forbes.com/sites/kurtbadenhausen/2019/08/28/the-nbas-richest-shoe-deals-lebron-kobe-and-durant-are-still-no-match-for-michael-jordan/#7a7dc3e83d02>, <https://www.forbes.com/sites/kurtbadenhausen/2020/05/03/michael-jordans-1-billion-nike-endorsement-is-the-biggest-bargain-in-sports/#527dd51b6136>, <https://www.forbes.com/sites/seandeveney/2020/05/25/new-documentary-tells-story-of-nike-air-jordans-warts-and-all/#5abf4ed93a80>, 18.6.2020.
- [11.] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I (January 2003). Lafferty, John (ed.). "Latent Dirichlet Allocation". Journal of Machine Learning Research. 3 (4–5): pp. 993–1022. doi:10.1162/jmlr.2003.3.4-5.993, 12.6.2020.

AUTOR · AUTHOR



• Sara Slamić Tarade

Rođena je 27.01.1992. u Zagrebu, gdje završava osnovnu školu i srednju strukovnu školu (Poljoprivredna škola Zagreb). Godine 2018. završava Edward Bernays Visoku školu

za komunikacijski menadžment te stječe titule stručne specijalistice za odnose s javnošću. Godine 2020. upisuje sveučilišni diplomski studij na Sveučilište Sjever u Varaždinu, smjer Odnosi s javnostima. Njezin akademski i znanstveni istraživački interes uključuje primjenu algoritama strojnog učenja za obradu teksta prirodnog jezika u području odnosa s javnošću i marketinga.

Korespondencija · Correspondence

sara.slamic@tvz.hr