

ALATI ZA ANALIZU SENTIMENTA

SENTIMENT ANALYSIS TOOLS

Amelia Kovačević¹, Željko Kovačević²

¹Nacionalna i sveučilišna knjižnica u Zagrebu, Hrvatske bratske zajednice 4, 10000 Zagreb

²Tehničko veleučilište u Zagrebu, Vrbik 8, 10000 Zagreb, Hrvatska

SAŽETAK

Analiza sentimenta (ili istraživanje mišljenja) tehnika je obrade prirodnog jezika koja se koristi za određivanje jesu li podaci pozitivni, negativni ili neutralni. S obzirom na izuzetnu važnost ove tehnike pri donošenju poslovnih odluka koje uvelike ovise i o evaluaciji mišljenja krajnjih korisnika i kupaca, u ovom radu dan je pregled nekoliko najčešće korištenih alata koji se mogu iskoristiti u te svrhe. Oni pomoću ključnih riječi pretražuju određeni internetski sadržaj poput društvenih mreža, foruma, blogova i slično. Zatim, korištenjem ugrađenih algoritama i ključnih riječi analiziraju sentiment, te dodjeljuju pozitivne, negativne ili neutralne konotacije. U radu se obrađuju tri pojma različitih kategorija. Njihova je svrha tek prikazati sličnosti i različitosti alata te vrste pojmova koji se analiziraju. Svrha nije pružiti općenitu analizu i sud o alatima već na konkretnim primjerima prikazati kako oni funkcioniraju te što se sve njima može postići u istraživanju mnijenja tržišta.

Ključne riječi: analiza sentimenta, istraživanje mišljenja, alati, evaluacija, subjektivnost

ABSTRACT

Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative, or neutral. Given the exceptional importance of this technique in making business decisions that largely depend on the evaluation of the opinions of end users and customers, this paper provides an overview of several commonly used tools that can be used for these purposes. They use keywords to search certain

Internet content such as social networks, forums, blogs and the like. Then, using built-in algorithms and keywords, they analyze sentiment and assign positive, negative, or neutral connotations. The paper deals with three topics of different categories. Their purpose is only to show the similarities and differences of the tools and the types of concepts that are analyzed. The purpose is not to provide a general analysis and judgment of the tools but to show by concrete examples how they work and what can be achieved by them in market opinion research.

Keywords: sentiment analysis, opinion mining, tools, evaluation, social media, subjectivity

1. UVOD

1. INTRODUCTION

Sav tekst objavljen na internetu poput blogova, komentara, foruma prenosi stavove koji pobuđuju osjećaje. [1] Logična posljedica toga jest i interes za izrečenim stavovima. Taj fenomen naglašeniji je pojavom foruma, a posebice društvenih mreža i blogova. Prije smo za mišljenje pitali osobe koje poznajemo, proveli anketu, a sada nam je Internet otvorio vrata i uveo čitav svijet u naš život. Ispitivanje mišljenja zastupljeno je kod istraživanja stavova o nekom proizvodu klasificiranjem recenzija proizvoda na internetu, praćenjem promjena u stavovima opće populacije prema političkim kandidatima, traženja savjeta i recenzija o raznim proizvodima, mišljenjima o političkim idejama, osobama, uslugama i sličnim pojavnostima u svakodnevnom životu. [2]

Pružateljima usluga i proizvoda, političkim kandidatima i sl. uvelike je olakšano upoznavanje ciljane publiku, odnosno koju vrstu sadržava preferiraju, kada su aktivni na kojim medijima; saznajemo koju su hashtagovi najpopularniji i privlače najviše prometa. Korisnici se sve rjeđe obraćaju proizvođaču, odnosno pružatelju usluga s pritužbama. Umjesto toga, objavljuju svoje negativne komentare online. Tako ne možemo razriješiti spornu situaciju na korist kupca ni svoju korist, već dobivamo negativne kritike što utječe na naše tržište. Analizom stavova tržišta može se spriječiti ta štetu reagirajući na njihove objave, prate se trendovi, napredak konkurencije.

Analiza sentimenta (engl. *sentiment analysis*; *opinion mining*) poddisciplina je računalne lingvistike koja se ne bavi prenesenim sadržajem već izrečenim stavom. [2] Može se podijeliti u nekoliko dijelova:

1. utvrđivanje subjektivnosti dokumenta – je li tekst činjeničan (objektivan) ili obojen sentimentom (subjektivan)
2. utvrđivanje orijentacije dokumenta, odnosno polariteta – pozitivan ili negativan stav
3. utvrđivanje jačine sentimenta – je li dokument slabo, srednje ili jako pozitivan, primjerice

Klasifikacija sentimenta (engl. *sentiment classification*) područje je analize sentimenta kojim se utvrđuje je li tekst koji sadrži sentiment pozitivan ili negativan. Sentiment se može analizirati na razini riječi, fraze, rečenice ili dokumenta. Na razini analize riječi koje nose sentiment, one mogu biti temeljno ili komparativno subjektivne – same po sebi ili u odnosu na drugu riječ. Određivanje sadrži li rečenica sentiment ili ne naziva se klasifikacija subjektivnosti (engl. *subjectivity classification*). Kod klasifikacije sentimenta govorimo o analizi na razini cijelog teksta. [3]

2. ANALIZA SENTIMENTA

2.1. SENTIMENT ANALYSIS

U radu će se prikazati rezultati pretraživanja triju pojmova.

To su:

- Elon Musk (osoba),
- how to tie a tie (frazu) i
- Procter & Gable (proizvođač).

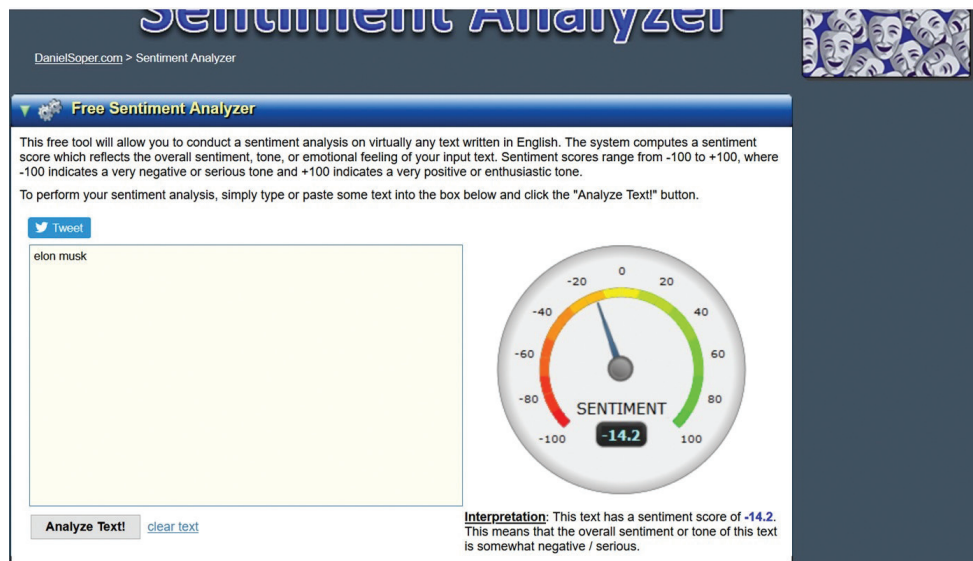
Pojmovi su odabrani iz triju kategorija. Za osobu je odabran Elon Musk jer se uz njega miješaju pozitivna i negativna stajališta – inovativan je, među najbogatijim je osobama na svijetu, njegov rad na komercijalnim letovima u svemir su kontroverzni kao još neki čimbenici njegovog života. „How to tie a tie“ leksički je zanimljiva fraza za proučavanje. Procter & Gamble svjetski je poznata trgovačka marka. Njihovi proizvodi su u širokoj upotrebi, ali uz njih se vežu i negativne konotacije (primjerice, pregršt PEG-ova u sastojcima proizvoda). Tvrtka je zanimljiva i zbog naziva, jer se uz pravilan oblik Procter & Gamble [4] koriste još i inačice Procter Gamble te Procter and Gamble.

Svi alati pretraženi su 4. ožujka 2021. Pretraženi su istovremeno kako ne bi došlo do nekog događaja koji bi utjecao na stavove o korištenim pojmovima čime rezultat istraživanja ne bi bio mjerodavan. Svaki od alata analiziran je u kontekstu mogućnosti koje pruža. Odabrana su četiri alata – Sentiment Analyzer, Social Buzz, Sentiment Viz i Brand24. Neki alati su pouzdani, ali ograničenih mogućnosti u odnosu na ostala postojeća rješenja, dok su neki namijenjeni agencijama koje rade analizu tržišta za svoje klijente. To je vidljivo i samim opsegom opisa pojedinih alata naprosto što kod nekih postoji više mogućnosti za opis i bolja dokumentacija.

2.1. SENTIMENT ANALYZER

2.1. SENTIMENT ANALYZER

Od prikazanih alata Sentiment Analyzer [5] nudi najmanje mogućnosti. (Slika 1) Analizira riječ ili tekst i prikazuje izračun sentimenta. Rezultati analize prikazani su putem mjerača s kazaljkom. Ne nude se nikakve dodatne mogućnosti, već samo ono što se vidi iz prikazanih slika. Upisani tekst analizira se bez mogućnosti korištenja Booleovih operatora. Nebitno je koriste li se navodnici i kod primjerak Procter and Gamble, veznik „and“, znak „&“ ili išta od navedenoga. Rezultati su jednaki.



Slika 1 Početni (i jedini) prikaz alata Sentiment Analyzer.

Figure 1 Initial (and only) screen of the Sentiment Analyzer.

Tablica 1. Rezultati pretraživanja zadanih pojmova u Sentiment Analyzeru.

Table 1. Search results of the given topics in Sentiment Analyzer

	Elon Musk	How to tie a tie	Procter (and/&) Gamble
Vrijednost sentimenta	-14,2	-100	-14,2

Ovaj alat analizira tekstove pisane na engleskom jeziku, prvenstveno američkom budući da koristi američki nacionalni korpus (American National Corpus, ANC) koji se sastoji od pisanih podataka i transkripata razgovora koji datiraju od 1990. godine nadalje, a uključuje dokumente, elektroničku poštu, tweetove, internetske podatke, s anotacijama i sl.

Na stranici se napominje da pruža općeniti rezultat, a ne pojedinih izvora pa ih niti ne navodi. Stranica se također ograđuje od preciznosti budući da bi čovjek drugačije ocijenio sentiment nego računalo i njihovi bi se rezultati podudarali u 80% slučajeva. [5]

Ovo je jedini alat koji je prikazao isti rezultat u decimalu za dva pojma (Tablica 1), što je vrlo neobično. Test je ponovljen 26.4.2021., ali rezultati su i dalje identični onima u tablici. Također, neobično je da upit o načinu vezanja kravate polučuje stopostotni negativni sentiment. Teoretski je gotovo nemoguće da je sav pretražen korpus o vezanju kravate isključivo negativan.

Sljedeći alati nude dodatne mogućnosti uz pretplatu kojih je uglavnom više vrsta. Nude filtre, slanje obavijesti elektroničkom poštom, generiranje izvještaja u raznim formatima, pregled aktivnosti prema vremenskom razdoblju i

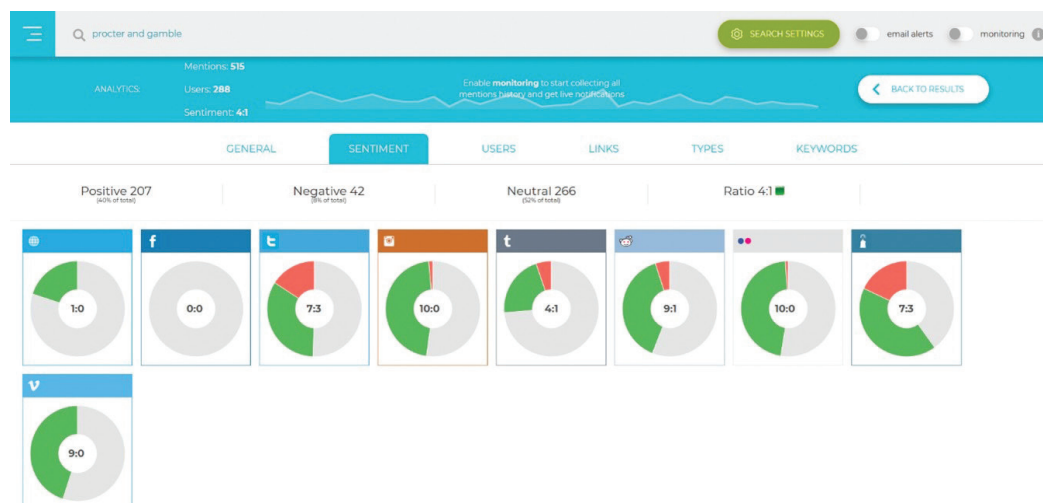
zemljopisnom položaju, analizu u kojem okružju se najčešće govori o pojmu/proizvodu/usluzi, korištene hashtagove i slično. Svi alati koji slijede imaju različite rezultate za pojmove Procter Gable, Procter and Gamble, Procter & Gamble (ispravan oblik naziva tvrtke [6]).

2.2. SOCIAL BUZZ

2.2. SOCIAL BUZZ

Ovaj alat [7] pretražuje jedanaest društvenih mreža. (Slika 2) U besplatnoj inačici moguće je spremati dva pretraživanja i primati mailove s obavijestima, a prikazuje se samo prvih 20 objava po društvenoj mreži. Dostupne su dodatne opcije za pretraživanje po ključnim riječima – točne ključne riječi (Booleove navodnike) i ključne riječi koje ne želimo da nam se prikazuju (Booleov minus). Ponuđeni su i filtri, prema tipu objava (poveznica, status, fotografija, video), sentimentu (pozitivan, negativan, neutralan), izvoru pojedinačno. Rezultati se mogu sortirati prema datumu i popularnosti.

Korištenje Social Buzza uz pretplatu nudi, očekivano, dodatne mogućnosti – izvoz podataka, API integraciju i napredna analitika. Pretražuje i novosti, blogove, forume, komentare i video zapise.



Slika 2 Jedan od zaslona alata Social Buzz.

Figure 2 One of the Social Buzz screenshots.

Omogućava i uvid u to tko sve dijeli pojedine stranice s domene. Dostupna su dodatna filtriranja i sortiranja rezultata, označavanje postova kao favorita, brisanje nevažnih objava, metriku sentimenta, kompleksne izvještaje. Podaci se čuvaju cijelo vrijeme trajanja pretplate.

Zanimljivo je usporediti odaziv kod triju oblika naziva tvrtke P&G (Tablica 2). Najveći broj rezultata dobiva se za oblik s veznom „and“, koji nije službeni oblik naziva. Međutim, kod pravilnog oblika najviše rezultata se doista odnosi na tu tvrtku, a ne na druge pojmove (prezime Procter često se pojavljuje vezano uz jedan sudski spor). Začuduje mali broj spominjanja naziva tvrtke Procter Gamble koji iznosi oko desetine broja drugih dvaju oblika. Ovaj alat našao je dosta tekstova o vezanju kravate. Sentiment je svima, slučajno, isti. Na 4 pozitivna sentimenta jedan je negativan – 4:1

2.3. SENTIMENT VIZ: TWEET SENTIMENT VISUALIZATION

2.3. SENTIMENT VIZ: TWEET SENTIMENT VISUALIZATION

Sentiment Viz nudi zanimljive funkcionalnosti po kojima je najdinamičniji, ali i najograničeniji od prikazanih na samo jednu društvenu mrežu. [8] Prikazuje rezultate Twittera dodjeljujući im kružice. (Slika 3.) Prelaskom miša (bez klikanja) preko kružica, prikazuje se tekst poruke. Riječi označene plavom bojom i podebljanim kurzivom procjenjuju se takve kakve jesu. Narančasta boja procjenjuju se kao negacija. Moguće je

istovremeno postavljanje upita za više ključnih riječi odjednom i njihova kombinacija. Pri tome se koriste Booleovi operatori.

Rezultati su prikazani u mreži (Slika 3) podijeljenoj u četiri dijela, s dodatnim podjelama. Rezultati su prikazani grafički prema tome koliko daleko datira objava, je li pozitivn(ij) a ili negativn(ij) a i slično, dakle ne samo prema sentimentu koji ima dva pola već i prema osjećajima. Pogledom na grafički smještaj objava lako je razaznati kontekst. Dodatne kategorije prikaza smještene su u svoje kartice (tabove). Objave su prikazane točkama čija boja nosi informaciju. Prelaskom miša preko točaka vidljiva je cijela objava s dodatnim, vrlo detaljnim informacijama za profesionalne korisnike.

Rezultati se mogu pregledavati koristeći kartice, odnosno tabove:

- **Sentiment.** Svaki tweet prikazan je kružicom. Lijevo su plavi tweetovi s negativnim sentimentom, a desno pozitivni i zeleni. Stariji tweetovi prikazani su tamnijim kružicama i u dnu su prikaza dok su aktivni pri vrhu i kružići su svjetlije boje. Prelaskom mišem preko kružica prikazuje se tekst i vrijednost pojedinih pojmova. Klikom na kružić dobivaju se dodatne informacije.
- **Tema.** Objave su organizirane u skupinama (klasterima) prema njihovoj temi koja se može identificirati ključnim riječima navedenim iznad skupina. Oni koji ne pripadaju skupinama (engl. *singletons*) prikazani su pojedinačno s desne strane.

Tablica 2. Rezultati Social Buzza.

Table 2. Social Buzz results.

	Elon Musk	How to tie a tie	Procter Gamble	Procter & Gamble	Procter and Gable
Broj spominjanja	458	525	52	416	515
Broj korisnika	348	379	285	235	288
Sentiment	4:1	4:1	4:1	4:1	4:1

Tablica 3. Rezultati Sentiment Viz.

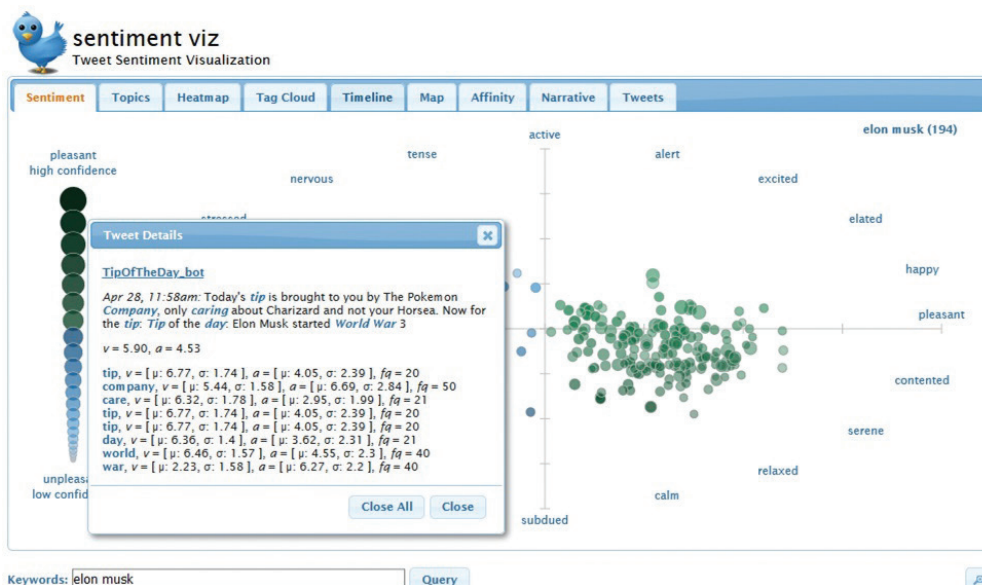
Table 3. Sentiment Viz results.

	Elon Musk	How to tie a tie	Procter Gamble	Procter & Gamble	Procter and Gamble
Sentiment viz	196	234	168	177	122

- **Naracija.** Odabir izvornog tweeta prikazuje kronološki redoslijed tweetova koji se nadovezuju na njega.
- **Tweet.** Rezultati se mogu sortirati prema: datumu, autoru objave, zadovoljstvu, uzbuđenju i samoj objavi. Moguć je odabir redoslijeda njihovog prikazivanja.
- **Toplinska karta.** Zadovoljstvo ili uzbuđenost dijele kartu na iscrtanu mrežu koja se sastoji od po 8 polja okomito i vodoravno. Svaki kvadrat dobiva određenu boju – crvenu za iznadprosječan broj tweetova, plavu za manji broj, a bijelu za polje koje ih nema. Broj se pojavljuje prelaskom mišem preko polja.

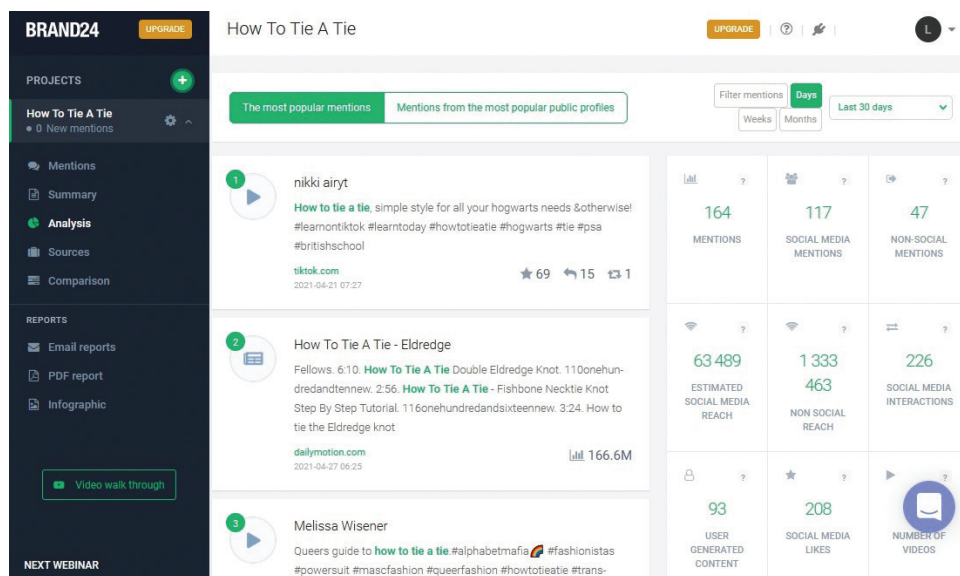
- **Tag oblak.** Prikazuju se učestale riječi u poljima koja predstavljaju određene osjećaje – Uznemiren, Sretan, Opušten i Nesretan. Riječi koje se pojavljuju češće prikazane su većim slovima. Prelaskom miša preko riječi prikazuje se podatak o učestalosti pojavljivanja.

Oblik „Procter & Gamble“ daje najbolji rezultat, 177 objava i to baš navedene tvrtke. Znak „&“ daje više rezultata, ali oni se ne odnose na tvrtku (opet često na prezime Procter). Oblik bez ikakvog poveznika je sljedeći s brojem objava, vjerojatno jer autori objava skraćuju naziv. Oblik naziva s veznikom daje najslabiji rezultat. Začudo, „Elon Musk“ je na drugom mjestu po objavama na Twitteru iza vezanja kravate. (Tablica 3)



Slika 3 Prikaz početne kartice Sentiment Viz.

Figure 3 Sentiment Viz initial tab screenshot.



Slika 4 Zaslona analize rezultata alatom Brand24.

Figure 4 Brand 24 results analysis screenshot.

Tablica 4. Rezultati analize pojmovna alatom Brand24.

Table 4. Brand24 results analysis.

	Elon Musk	How to tie a tie	Procter Gamble	Procter and Gable	Procter & Gamble
Mentions	18.640	112	25	403	185
Social media mentions	768	73	1	5	12
Non-social media mentions	17.872	39	24	398	173
Estimated social media reach	2,481.887	1.493	4	16	26.000
Non-social media reach	318,493.562	1,106.504	80.571	8,728.008	7,176.687
Social media interactions	0	0	0	0	0
Positive mentions	3.637	12	7	0	2
	42%	39%	70%	0%	100%
Negative mentions	5.188	19	3	0	0
	58%	61%	30%	0%	0%
AVE	\$18.41M	\$7.739K	\$388	\$522,25K	\$372,51K

2.4. BRAND24

2.4. BRAND24

Ovaj alat [9] za praćenje sentimenta jedan je od najpopularnijih i od svih prikazanih najprofesionalniji je. (Slika 4)

Koriste ga tvrtke kao što su Ikea, Panasonic, Intel, Vichy, GlaxoSmithKline. Tri su modela pretplate, o kojima, logično, ovisi i mogućnosti korištenja.

Prati javno dostupne objave od dana postavljanja projekta. Najranije prikupljanje podataka koje je moguće je unazad najviše 30 dana. Potrebno je odabrati kanale koji se promatraju – Facebook, Instagram, Twitter, novosti, video zapisi, bilteni, podcastovi, forumi, blogovi, internetske stranice. Postavljaju se ključne riječi za svaki projekt, između 5 i 20. Ažuriranje rezultata vrši se u realnom vremenu, odnosno u roku od nekoliko minuta, svakih sat vremena ili svakih 12 sati. Na mjesečnoj razini moguća je analiza do 5, 100 ili 250 tisuća spominjanja. Izvoz podataka moguć je u formatima pdf, xml, excel ili kao infografika. U najvišoj pretplati, namijenjenoj agencijama za analizu tržišta, nude se i konzultacije. Rezultate je moguće usporediti s rezultatima konkurencije. Dostupna je i aplikacija za platforme iOS i Android.

Grafikon volumena rasprave prikazuje nagle promjene u opsegu spominjanja branda. Problemi se mogu uočiti na vrijeme i pravodobno reagirati kako bi se spriječila šteta, odnosno zadržala pozitivna slika u javnosti.

Ovaj alat nudi jednu novinu u odnosu na ostale obrađene, a to je AVE (*Advertising value equivalent*) – vrijednost prikupljenih članaka i spominjanja, odnosno iznos koji bi se trebalo platiti da se ostvari postignut doseg. Brand24 izračunava koliko bi koštalo reklamiranje branda/ usluge da bi se postigao ostvareni doseg što je vjerojatno svima zanimljiv podatak.

Od ukupnog broja objava, odnosno spominjanja pretraživanih pojmova, najveći dio otpada na sadržaj izvan društvenih mreža kod svih analiziranih pojmova. (Tablica 4) Međutim, ovdje je najviše vidljiva dominacija pojma „Elon Musk“ pred ostalim obrađenim pojmovima. Zanimljivo je da od ukupnog broja spominjana, tek mali postotak otpada na društvene mreže. Doseg objava također je daleko manji kod društvenih medija. Stoga ne iznenađuje iznenađuje odsutnost dijeljenja na društvenim mrežama, vjerojatno jer ih je tako malo. Iznadpolovični negativni sentiment vezan uz pretraživanje pojma o vezanju kravate i nije iznenađujući budući da ga korisnici spominju u potrazi z savjetom budući da trebaju pomoć s time.

3. ZAKLJUČAK

3. CONCLUSION

Analiza sentimenta ključna je za istraživanje javnog mnijenja te za razvoj branda, radilo se o proizvodu, usluzi, tvrtki ili o osobi. U tu svrhu koriste se prilagođeni alati koji korištenjem različitih metodologija rade analizu sentimenta i brojčanim vrijednostima ukazuju na utjecaj određenog pojma. Rad je predstavio neke od najčešće korištenih alata u tom području, njihove mogućnosti i na koji način mogu pomoći u analizi tržišta. Kako svi prikazani alati koriste različitu metodologiju analize sentimenta, tu je i mogući prostor za poboljšanje kojemu se planiramo posvetiti u nastavku istraživanja. Primjerice, ukoliko svaku od metodologija predstavimo kao zasebni klasifikator koji rezultira vlastitom evaluacijom ciljanog sentimenta, uvjereni smo da je onda također moguće po uzoru na ansambl metode strojnog učenja [10] [11] dobiti i dobre predikcije rezultata, te time skratiti buduće vrijeme potrebno za cjelovitu analizu sentimenta.

4. REFERENCE

4. REFERENCES

- [1.] A. Devitt i K. Ahmad, »Sentiment Polarity Identification in Financial News: A Cohesion-based Approach,« u Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prag, Republika Češka, 2007.
- [2.] A. Esuli i F. Sebastiani, »Determining Term Subjectivity and Term Orientation for Opinion Mining,« u 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, Italija, 2006.
- [3.] B. Liu, »Sentiment Analysis and Subjectivity,« u Handbook of Natural Language Processing, Chapman and Hall/ CRC, 2010, pp. 1-38.
- [4.] »Consumer Terms of Use for Procter & Gamble Web Sites,« P&G, [Mrežno]. Available: https://www.pg.com/en_US/terms_conditions/index.shtml. [Pokušaj pristupa 01 04 2021].

- [5.] D. Soper, »Sentiment Analyzer,« [Mrežno]. Available: <https://www.danielsoper.com/sentimentanalysis/default.aspx>. [Pokušaj pristupa 04 03 2021].
- [6.] »P&G,« 01 04 2021. [Mrežno]. Available: https://www.pg.com/en_US/terms_conditions/index.shtml.
- [7.] »Social Buzz,« [Mrežno]. Available: <https://www.social-searcher.com/social-buzz/>. [Pokušaj pristupa 04 03 2021].
- [8.] »Sentiment Viz,« [Mrežno]. Available: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/. [Pokušaj pristupa 03 04 2021].
- [9.] »Brand24,« [Mrežno]. Available: brand24.com. [Pokušaj pristupa 04 03 2021].
- [10.] D. O. a. R. Maclin, »Popular Ensemble Methods: An Empirical Study,« *Journal of Artificial Intelligence Research*, svez. 11, pp. 169-198, 1999.
- [11.] D. T.G., »Ensemble Methods in Machine Learning,« *Computer Science*, svez. 1857, 2000.

AUTORI · AUTHORS



• Amelia Kovačević

Diplomirala je informatologiju te hrvatski jezik i književnost na Filozofskom fakultetu Sveučilišta u Zagrebu. Zaposlena je u Nacionalnoj i sveučilišnoj knjižnici u Zagrebu u zvanju

višeg knjižničara. Trenutno pohađa doktorski studij pri Filozofskom fakultetu Sveučilišta u Zagrebu, a područja interesa su joj obvezni primjerak, edukacija odraslih, igrifikacija i razvoj web stranica.

• **Željko Kovačević** - nepromjenjena biografija nalazi se u časopisu *Polytechnic & Design* Vol. 8, No. 3, 2020.

Korespondencija · Correspondence

zeljko.kovacevic@tvz.hr