
Analýza zákaznickeho sentimentu na sociálnych sieťach

Mária Szivósová¹

Abstrakt

Náš každodenný život bol a je stále ovplyvnený tým, čo si ľudia myslia. Myšlienky a názory, ktoré majú ostatní, vždy ovplyvnia naše vlastné názory. Explózia Web 2.0 má za následok zvýšenú aktivitu podcastingu, blogovania, značkovania, čo prispieva k RSS, sociálnemu bookmarkingu a sociálnym sieťam. V dôsledku toho došlo k erupcii záujmu čerpania zdrojov z obrovských dát pre určenie stanoviska. Analýza sentimentu alebo čerpanie je výpočtová úprava stanovísk, emócií a subjektivity textu. V tomto článku sa pozrieme na rôzne problémy pri aplikácií analýzy sentimentu. Budeme diskutovať o detailoch rôznych prístupov na vykonanie výpočtovej úpravy stanovísk a emócií.

Kľúčové slová

sentiment, analýza sentimentu, Twitter

Abstract

Our day-to-day life has always been influenced by what people think. Ideas and opinions of others have always affected our own opinions. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. In this article, we take a look at the various challenges and applications of Sentiment Analysis. We will discuss in details various approaches to perform a computational treatment of sentiments and opinion.

Key words

sentiment, sentiment analysis, Twitter

JEL classification

M15

1 Úvod

Vďaka vzniku online marketingu, firmy prezentujúce sa na sociálnych sieťach došli k názoru, že potrebujú sledovať o čom sa ich zákazníci rozprávajú a aké majú názory. Dôležitou súčasťou prieskumu je analýza sentimentu v písaných prejavoch, ktorá môže odhaliť posuny v postojoch, ktoré v anketách, prieskumoch verejnej mienky alebo v dotazníkoch nie sú viditeľné.

2 Analýza, aplikácie a výzvy pre analýzu sentimentu

Analýza sentimentu je spracovanie prirodzeného jazyka a extrakcia informácií. Cieľom je získať informácie o pocitoch človeka, ktorý daný text napísal. Tieto informácie môžu byť pozitívne ale aj negatívne. Všeobecne povedané, analýza sentimentu si kladie za cieľ zistiť postoj autora, pokiaľ ide o určitú tému. V posledných rokoch exponenciálny nárast využitia internetu je hnacou silou pre analýzu sentimentu. Internet je obrovské úložisko štruktúrovaných

¹ Ing. Mária Szivósová, PhD., University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemska cesta 1, 852 35 Bratislava, E-mail: maria.szivosova@euba.sk

a neštruktúrovaných dát. Analyzovať takéto dáta a extrahovať z nich užitočné informácie verejnej mienky je náročná úloha. Analýza emócií je založená na tom, že určí či daný text obsahuje negatívny alebo pozitívny sentiment. Môže sa to určovať na základe viet, pokiaľ sú tieto vety klasifikované. Tak isto aj slová v daných vetách musia byť klasifikované. Analýza sentimentu tieto slová identifikuje a určí, ktoré slovo má akú emóciu. Autor môže písať o objektívnych skutočnostiach alebo svojich subjektívnych názoroch a postojoch k danej problematike. Je nutné rozlišovať predmet, ku ktorému je sentiment smerovaný. Text môže obsahovať veľa subjektov ale je nutné nájsť entitu, ku ktorej bol sentiment smerovaný. Emócie sú klasifikované ako objektívne (fakty), pozitívne (označujú stav šťastia, blaženosti, ktoré pochádzajú zo strany spisovateľa), alebo negatívne (označujú stav smútku, sklúčenosti či sklamanie zo strany spisovateľa).

"Word of mouth" (WOM) je proces odovzdávania informácií od človeka k človeku a hrá dôležitú úlohu pri rozhodovaní zákazníkov pri nákupe. V komerčných situáciách, WOM obsahuje zdieľanie spotrebiteľských postojov, názory a reakcie o firmách, produktoch alebo službách s ostatnými zákazníkmi.² WOM komunikačné funkcie sú založené na sociálnych sieťach a na dôvere. Ľudia sa spoliehajú na názory rodín s deťmi, priateľov, známych osobností a iných ľudí na sociálnych sieťach. Výskum poukázal, že ľudia ľahko uveria zdanlivo nezaujatým názorom iných ľudí na takzvaných on-line hodnotiacich stránkach. Toto je miesto, kde prichádza na rad Analýza Sentimentu. Rastúca dostupnosť bohatých zdrojov spotrebiteľskej mienky ako napríklad blogov, statusov na sociálnych sieťach zjednodušujú "rozhodovací proces". S explóziou Web 2.0 platformy majú spotrebiteľia veľký priestor pre zdieľanie názorov. Významné firmy si uvedomujú akú silu majú v dnešnej dobe spotrebiteľské hlasy a snažia sa formovať ich názory. Analýza sentimentu tak nájde uplatnenie v oblasti spotrebiteľského trhu pre hodnotenie tovarov alebo služieb. Pomocou sociálnych sietí je možné napríklad zistiť názor na nové trendy. Tak isto je možné zistiť či nedávno zverejnený film je hit. Všeobecne klasifikujúce aplikácie je možné rozdeliť do niekoľkých kategórií.

- Applications to Review-Related Websites

(sú to napríklad filmové recenzie, recenzie výrobkov atď.)

- Applications as a Sub-Component Technology

(napríklad detekcia spamu, kontextová detekcia informácie atď.)

- Aplikácie v obchodných a vládnych spravodajstvách

(zistenie postojov a spotrebiteľských trendov)

- Aplikácie medzi rôznymi doménami

(zistenie verejnej mienky o vládnych predstaviteľoch alebo mienka o zákonoch, pravidlách a nariadeniach).³

Prístupy sentiment analýzy majú za cieľ extrahovať pozitívny a negatívny sentiment, ktorý nesie slovo v určitom texte. Musí klasifikovať text ako pozitívny, negatívny, objektívny alebo neurčitý, ak nie je schopný nájsť určité slovo vo svojom ložisku. V tomto ohľade môžeme využiť postup ktorý sa nazýva klasifikácia textu. V klasifikácii textu existuje veľa tried, ktoré zodpovedajú rôznym témam, zatiaľ čo v analýze sentimentu máme len tri široké kategórie. Zdá

² Dey, L., & Haque, S. K. (n.d.). Opinion mining from noisy text data. Retrieved August 28, 2017, from <https://www.semanticscholar.org/paper/Opinion-mining-from-noisy-text-data-Dey-Haque/5d7cfff09bfa17ec2f45b7e0731bd66ed773c4ae5/figure/15>

³ Di Bo Pang (Autore), Lillian Lee (Autore). (n.d.). Opinion Mining and Sentiment Analysis (Inglese) Copertina flessibile – 8 lug 2008. Retrieved August 28, 2017, from <https://www.amazon.it/Opinion-Mining-Sentiment-Analysis-Pang/dp/1601981503>

sa teda, že analýza sentimentu je jednoduchšie ako klasifikácia textu. Avšak nie je to celkom pravda. Všeobecné problémy možno zhrnúť do týchto tried:

Implicitný sentiment a sarkazmus

Doménová závislosť

Zmenené očakávania

Pragmatika

Svetové poznanie

Subjektívna detekcia

Entitná identifikácia

Negácia

3 Sentiment na sociálnej sieti Twitter

V porovnaní s ostatnými sociálnymi sieťami sa na Twitteri objavuje menšie percento neutrálnych príspevkov. K tomuto javu dopomáhajú médiá často vyzývajúce obecnosť ku aktivite pomocou zverejňovania sémantických značiek alebo reakcií na zadanú otázku. Slovenský Twitter má viac ako 130 tisíc používateľov a ich počet stále narastá.

Twitter je jednou z najväčších sociálnych sietí. Využíva ju viac ako 550 miliónov používateľov po celom svete. Títo užívatelia generujú obrovský počet dát. Jedná sa o mikro blogovací systém ktorý sa nazýva "databáza názorov". Užívatelia tu môžu komunikovať pomocou textových správ s maximálnou možnou dĺžkou 140 znakov (tzv. tweet). Toto obmedzenie má vplyv na výpovednú hodnotu každého príspevku. Preto sociálna sieť Twitter poskytuje špeciálny formát pre kľúčové slová a označenie používateľov. Kľúčové slová, ktoré sú nazývané sémantickými značkami (hashtags) sú zapísané vo forme #slovo a používateľov označujeme pridaním znakov "@", takže @používateľ. Jednotlivé príspevky je možno zaradiť medzi obľúbené alebo ich je možné ďalej preposielať (retweet), čoho sa dosiahne použitím špeciálnej funkcie alebo pridaním reťazca " RT " pred posielanou správou. Tento princíp preposielania je podstatou rýchleho šírenia správy naprieč celou sieťou. Každý používateľ môže sledovať (follow) príspevky iných používateľov, čím sa stáva sledujúcim (follower). Pre zhromažďovanie príspevkov určitých používateľov slúžia zoznamy (lists), ktoré odoberateľom (subscribers) dovoľujú odoberať príspevky od členov (members). Zoznamy môžu byť súkromné alebo verejné. Podľa toho, koho sledujete a kde sa práve nachádzate, vám sociálna sieť Twitter ponúka trendy - kľúčové slová vybrané pre každého používateľa na mieru. Okrem príspevkov môžeme posielat' súkromné správy (direct messages), ktoré sú viditeľné len medzi komunikujúcimi používateľmi. Twitter umožňuje okrem textu pridávať k príspevkom aj obrázky, ktoré sú najskôr nahraté na server a potom vo forme URL adresy pridané k príspevku. Pretože URL adresy môžu obsahovať veľké množstvo znakov, často sa využívajú služby skracovania URL adresy systémom Twitteru alebo tretích strán.

3.1 Funkcia ohodnotenia n-gramu

Táto funkcia je využívaná všetkými metódami pre výpočet sentimentu. Jej úlohou je vybrať hodnotenie pre konkrétny n-gram, ktoré bude ďalej používané v metódach pre vyhodnotenie sentimentu. Funkcie sa rozlišujú číselnou hodnotou sentimentu a špeciálnymi značkami. Pokiaľ je väčšinové hodnotenie n-gramu špeciálna značka, bude tomuto n-gramu priradená ako hodnota práve táto špeciálna značka. Ak je väčšinové hodnotenie číselné, potom sa vypočíta priemer a tento priemer bude použitý ako hodnota n-gramu.

3.1.1 Triviálna metóda

Jedná sa o základnú metódu pre vyhodnocovanie sentimentu príspevku. Je vypočítaná podľa označovaných unigramov. Funguje na vyhľadávaní všetkých unigramov v korpuse a získaní ich ohodnotení. Neutrálne unigramy algoritmus ignoruje a nezapočítava do hodnotenia. Pokiaľ algoritmus objaví zosilujúce značky, tak ich nasledovník bude ohodnotený dvojnásobnou mierou, naopak ak sa jedná o značku zoslabenia potom hodnota unigramu bude vydelená dvoma. Cudzozajčné a nevyžiadané slová algoritmus ignoruje, ale zároveň ich ukladá do korpusu, aby sa ďalej nezobrazovali pri značkovaní ostatných používateľov. Pri nájdení značky negácie je negovaná hodnota nasledujúceho unigramu. Výsledná hodnota sentimentu je priemer všetkých nenulových hodnôt (vrátane tých zosílených / zoslabených) pre každý príspevok. Túto metódu je tiež možné modifikovať tak, že sa robí na bigramoch alebo trigramoch. Pri tejto metóde dochádza k najlepším výsledkom pri silne negatívnych alebo silne pozitívnych príspevkoch. Pri malom počte neutrálnych n-gramov platí, že pomocou bigramov a trigramov sú výsledky presnejšie. Najväčšou slabinou tejto metódy je detekcia neutrálnych príspevkov, pokiaľ sa jedná o príspevok obsahujúci ojedinele sentimentálne zafarbené slová. Tieto väčšinou neutrálne príspevky sú často hodnotené ako pozitívne alebo negatívne.

3.1.2 Priemerová metóda

Táto metóda vychádza z predchádzajúcej triviálnej metódy a obohacuje ju o koeficient počtu neutrálnych n-gramov, ktorý prináša "zjemnenie" pre detekciu neutrálnych príspevkov. Algoritmus vypočíta podiel neutrálnych n-gramov, ktorý sa nazýva koeficient počtu núl. Týmto koeficientom sa ďalej násobí aritmetický priemer nenulových n-gramov. Toto roznásobenie berie do úvahy veľký počet neutrálnych slov a "zjemňuje" význam ojedinelých sentimentálne zafarbených slov. Tento algoritmus je taktiež modifikovaný pre prácu na bigramoch a trigramoch. Ojedinelé mierne pozitívne alebo negatívne slová automaticky nezaznamenajú príspevok s takýmto sentimentom, ale väčšinou sú vyhodnotenú ako neutrálne. Toto je hlavnou prednosťou tohto algoritmu. Avšak tento algoritmus nedokáže detekovať veľmi slabý sentiment, či je pozitívny alebo negatívny.

3.1.3 Rozdielová metóda

Táto metóda pracuje s n-gramami podľa priority, kde trigram má najväčšiu váhu a unigram najmenšiu. Algoritmus prehľadáva v korpuse najprv trigramy, ktorých sentiment je kladný alebo záporný a až potom hľadá bigramy a unigramy. Pokiaľ nenájde žiadny sentimentálne zafarbený n-gram, dosadí najmenšiu neutrálnu jednotku - unigram. Po rozdelení nasleduje výpočet pozitívneho a negatívneho koeficientu. Algoritmus spočíta podiel pozitívnych a negatívnych n-gramov a ten vynásobí ich priemernou hodnotou. Nakoniec spočíta rozdiel medzi týmito koeficientmi, ktorý sa považuje za výslednú hodnotu sentimentu. Rozdielová metóda bola navrhnutá tak, aby dokázala brať v úvahu nejednoznačné príspevky, v ktorých sa objavujú pozitívne ale aj negatívne n-gramy. Jej hlavnou nevýhodou sú príspevky, v ktorých sa nachádza rovnomerný počet pozitívnych a negatívnych n-gramov. Túto situáciu môže algoritmus chybné považovať ako neutrálnu, pretože rozdielový koeficient sa bude blížiť k nule.

Analýza sentimentu na Twitteri sa značne odlišuje od iných štúdií sentimentu. Príspevky na sociálnej sieti Twitter môžu mať maximálne 140 znakov, ale často sú oveľa kratšie a obsahujú na viac hypertextové odkazy, ktoré sú pre detekcie sentimentu irelevantné. Z dôvodu limitu znakov sú často používané skratky alebo slangový jazyk. Navyše jazyk, ktorý sa používa na sociálnych sieťach sa odlišuje od tradičných textov. Príkladom môžu byť slová ako "lol", "rofl", emotikony alebo špeciálne značky pre Twitter ako "RT", "#" či "@". Samotné slová

môžu byť hodnotené ako negatívne alebo pozitívne, ale kontext ktorý vytvárajú môže byť úplne iný.⁴

3.2 Problémy analýzy sentimentu

Väčšina nástrojov na monitorovanie sociálnych sietí poskytuje strojovú analýzu sentimentu. Problém nastáva pri overovaní presnosti jej výsledku. Bohužiaľ neexistuje žiadna množina dát, ktorá by bola autorizovaná a ku ktorej by bolo možné porovnávať konkrétne postupy. Jednou z príčin, prečo takýto súbor dát neexistuje je to, že s vyhodnocovaním sentimentu majú problém nie len počítače, ale aj ľudia. Predstavme si vetu " Vaša izba sa nachádza v prízemí, hneď vedľa recepcie". Pokiaľ autor chcel izbu na pokojnom mieste s výhľadom na more, jeho postoj k tejto vete bude negatívny. Avšak, ak sa jedná o osobu s invalidným vozíkom, bude spokojná, že má izbu na ľahko prístupnom mieste. Pre všetkých ostatných bude táto veta iba obyčajný neutrálny fakt. Bolo dokázané, že ľudia nie sú schopní zhodnúť sa na tom, čo je pozitívne, neutrálne a čo je negatívne. Ďalším faktom je to, že lepšie sa detekuje extrémna negatívita ako drobná pozitívita. Okrem toho, že vyhodnotenie sentimentu je zložité aj pre ľudí, existujú aj ďalšie problémy ako sú kontext, negácia, irónia, ktoré robia vyhodnotenie sentimentu textu extrémne zložité. Čo sa týka sociálnych sietí a Twitteru platí jednoduchá rovnica - čím kratší text, tým zložitejšia analýza.

3.2.1 Kontext a sarkazmus na Twitteri

Ďalším častým problémom je kontext, ktorý plní dôležitú úlohu a môže často meniť význam celého textu. Súčasná metóda analýzy sentimentu sú statické a neberú do úvahy kontext, ktorý je pre ľudský intelekt úplne prirodzený ale počítaču spôsobuje obrovské problémy. Ako príklad si môžeme uviesť vetu " Moji chudáci z Realu znova prehrali s Barcelou 3:5. Do toho Real!" Detekované boli dve silno negatívne slová "chudáci" a "prehrali". Tento príspevok je však používaný v inom kontexte. Autor ľutuje Real Madrid a v závere im fandí. Kontext tiež spôsobuje problémy u fráz, ktoré majú špecifický význam a súčasne synonymne popisujú nejakú vec. Príkladom je napríklad veta: "To je bomba!" Autor mohol mať na mysli, že je niečo naozaj skvelé - pozitívny sentiment. Na druhú stranu mohol tiež popisovať nejaký druh bomby, napríklad vodíkovú bombu. Čo sa týka sarkazmu, ako u každého typu spracovania prirodzeného jazyka záleží na kontexte. Analyzovanie prirodzeného jazyka je veľmi komplikovaný problém a sarkazmus a iné druhy ironického jazyka sú neodmysliteľne problematické pri detekcii počítačom, pokiaľ k nim pristupujeme pomocou izolovaných slov. Príkladom môže byť nasledujúca veta: " To je teda naozaj úžasné. " Slovo "úžasný" je veľmi pozitívne a na viac je zosílené slovom "naozaj", ale pokiaľ by sa jednalo o sarkasticky myslenú vetu, tak by mala byť táto veta myslená ako negatívna. Aby sme mohli presnejšie detekovať podobné javy, potrebovali by sme poznať profil autora a vedieť, ako často používa sarkazmus alebo iróniu. Posledným z kontextuálnych problémov sú viacznačné slová. Niektoré slová môžu mať dva a viac významov, pričom jeden význam môže znamenať pri značovaní určitú mieru sentimentu a ďalší môže označovať špeciálnu značku. Príkladom na takéto slovo môže byť slovo "strašne". Toto slovo môže byť považované pri značovaní ako zosílené alebo ako slovo s negatívnym sentimentom. Problém potom nastáva pri vyhodnocovaní algoritmov, ktoré do svojho výpočtu dosadzujú len tú hodnotu, ktorá je väčšinová.

3.2.2 Cudzozajčné slová

Ako už bolo zmienené, na "slovenskom Twitteri" užívatelia často používajú anglické názvy. Väčšinou sa jedná o klasický počítačový žargón alebo slová bežne používané na

⁴ Twitter Libraries from: <https://dev.twitter.com/docs/twitter-libraries>.

internete. Vo väčšine prípadov ich môžeme považovať za neutrálne, ale objavujú sa aj výnimky, ktoré majú veľmi silný sentiment. Príkladom môže byť najpopulárnejšia sémantická značka "#fail", ktorá je veľmi negatívna a preto ju nemôžeme ignorovať ako ostatné cudzojazyčné slová.

3.2.3 Negácia

Používanie negácie sa v slovenskom jazyku riadi určitými pravidlami a pri analýze sentimentu prináša nemalé problémy. Zápory rozlišujeme podľa toho, či chceme negovať obsah celej vety, jej časti alebo len slova. Rozlišujeme tri kategórie : 1. vetný zápor - popiera obsah celej vety " Čakal som tam, ale zase nikto neprišiel". 2. členský zápor - popiera platnosť jedného vetného člena "Nie mne, ale mame to hovor. " 3. slovný zápor - popiera význam samotného slova " Dnešný obed bol naozaj nechutný." Niektoré druhy negácií dokážeme úspešne vyhodnotiť, ale problém môže nastať pri vete: " To naozaj nie je zlé. " Veta obsahuje dve negatívne slová, ktoré sú v skutočnosti súčasťou pozitívnej vety.

4 Nástroje pre analýzu sentimentu

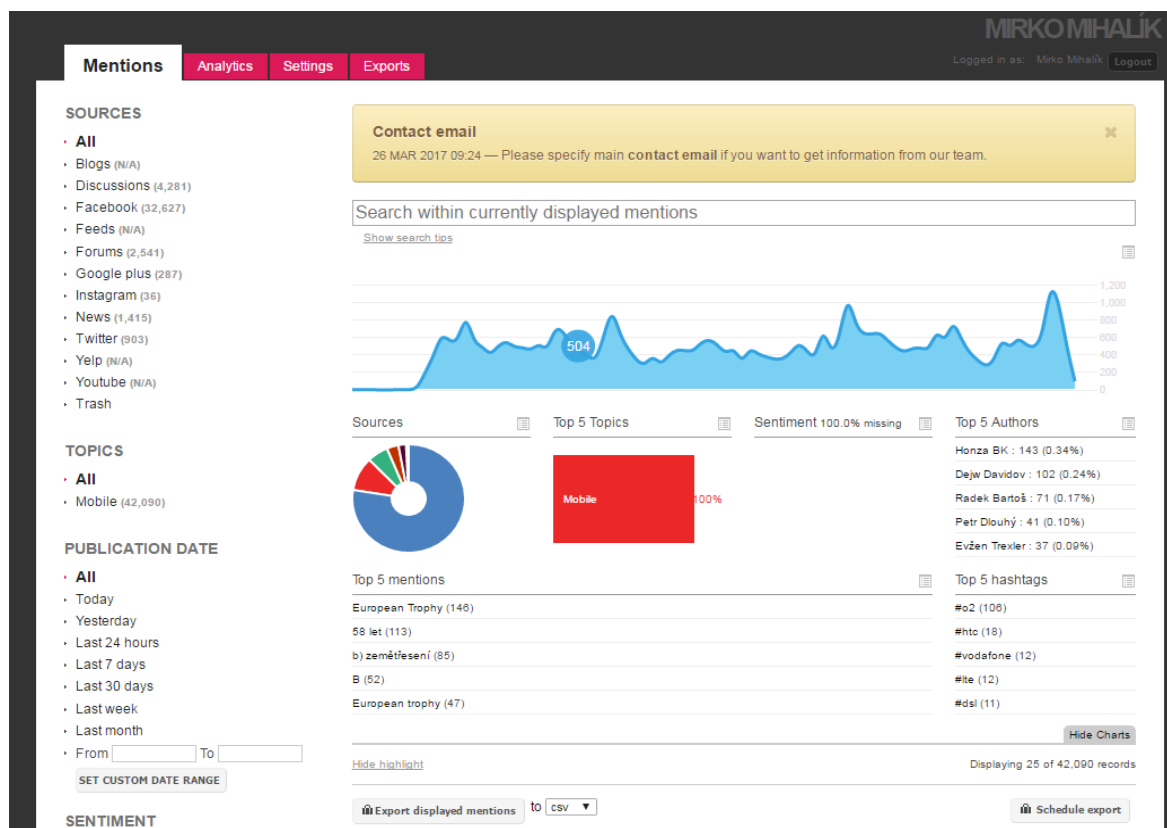
Veľmi jednoduchý systém pre vyhľadávanie sentimentu má v sebe zbudovaná sociálna sieť Twitter, konkrétnejšie jeho model vyhľadávania. V rozšírenom vyhľadávaní priamo ponúka funkcia vyhľadávania pozitívnych alebo negatívnych príspevkov. Táto funkcia je založená na tom, či príspevok obsahuje emotikon, čo sa nedá nazvať ako nástroj pre analýzu sentimentu. Jedným známym nástrojom, ktorý dokáže analyzovať sentiment v slovenčine je produkt Ataxo Social Insider.

4.1 Ataxo Social Insider

Na slovenskom webe sú dve väčšie firmy s nástrojmi pre monitoring a analýzu sociálnych sietí zameriavajúce sa na slovenské prostredie. Tým prvým je nástroj Social Insider od spoločnosti Ataxo Interactive (Social Insider, 2014), ktorý je plne lokalizovaný pre Slovenskú a Českú republiku.⁵ Ide o komerčný nástroj ale používateľom môže byť poskytnutá trial verzia na dobu 14 dní. Je potrebné sa prihlásiť pomocou Facebooku alebo Twitteru a odoslať žiadosť spoločnosti na sprístupnenie trial verzie. Ataxo Social Insider (ďalej len ASI) dokáže prehľadávať v blogoch, diskusných fórach, sociálnych sieťach a aj na spravodajských serveroch. ASI pre vyhľadávanie zadaných dotazov používa vyhľadávač, RSS aj crawler. Vyhľadávanie prebieha na základe zvolených kľúčových slov (max.500) alebo podľa zvolených tém. Ďalší porovnateľný komerčný nástroj Monitoring sociálnych sietí od spoločnosti Newton Media, a.s nemá v skúsobnej verzii sprístupnené vyhľadávanie podľa ľubovoľných kľúčových slov, ale sú len k dispozícii iba vybrané témy týkajúce sa finančného sektora (produkty, banky, poisťovne), tak sú aj v ASI vyhľadávané podobne témy podľa kľúčových slov : banka, pôžička, úver. V pokročilom vyhľadávaní je možné pri kľúčových slovách využiť logické operátory. ASI má pomerne prepracované filtrovanie a kategorizáciu výsledkov vyhľadávania. Vyhľadávané výsledky je možné členiť podľa zdroja (blogy, diskusie, Facebook, fóra, Google +, Instagram, Twitter, Yelp, Youtube a spravodajské servery), podľa kľúčových slov (aj v rámci témy), dát publikácie (v závislosti na zdroji), sentimentu (pozitívny, negatívny, neutrálny a neoznačený), a v neposlednej rade podľa "bydliska" (či je príspevok z Českej alebo Slovenskej republiky).

⁵Ataxo - Social Insider [online]. from WWW: [http:// ataxosocialinsider.cz/](http://ataxosocialinsider.cz/). *Methodological manual*

Obrázok 1: Hlavná stránka nástroja ASI



Zdroj: Ataxo - Social Insider [online]. Dostupné na WWW: <http://ataxosocialinsider.cz>

Vyhodnocovanie sentimentu nie je automatické. Používateľ musí zvolený príspevok ručne ohodnotiť. Množstvo ohodnotených príspevkov sa potom sumarizuje a podľa príslušnej kategórie a je zobrazené v prstencovom grafe.

Obrázok 2 : Vyhľadovaný príspevok v prostredí ASI

Datové jaro se blíží, T-Mobile nafoukne datové limity. Už zítra [spekulace]

27 MAR 2017 07:28 — "Data přidává i **O2**, 50 MB navíc dostane tarif Start". Tak moc?? Neblbněte!! Naši operátoři jsou hlupáci a amatéři. Z datových tarifů mohli mít zlatý důl - ale ceny nastavili tak hloupě a nekřesťansky vysoko, že se spousta lidí zalekla mobilního internetu a ve výsledku jsou 3G+4G využité sotva z 20%. Takový internetový datový balíček jsem si ještě nikdy nedokupoval - ale kdyby měli 1500MB za 99,-Kč... [read more](#)

AUTHOR: TOM BUT

SOURCE: DISCUSSIONS

DOMICILE: CS

DOMAIN: WWW.MOBILMANIA.CZ

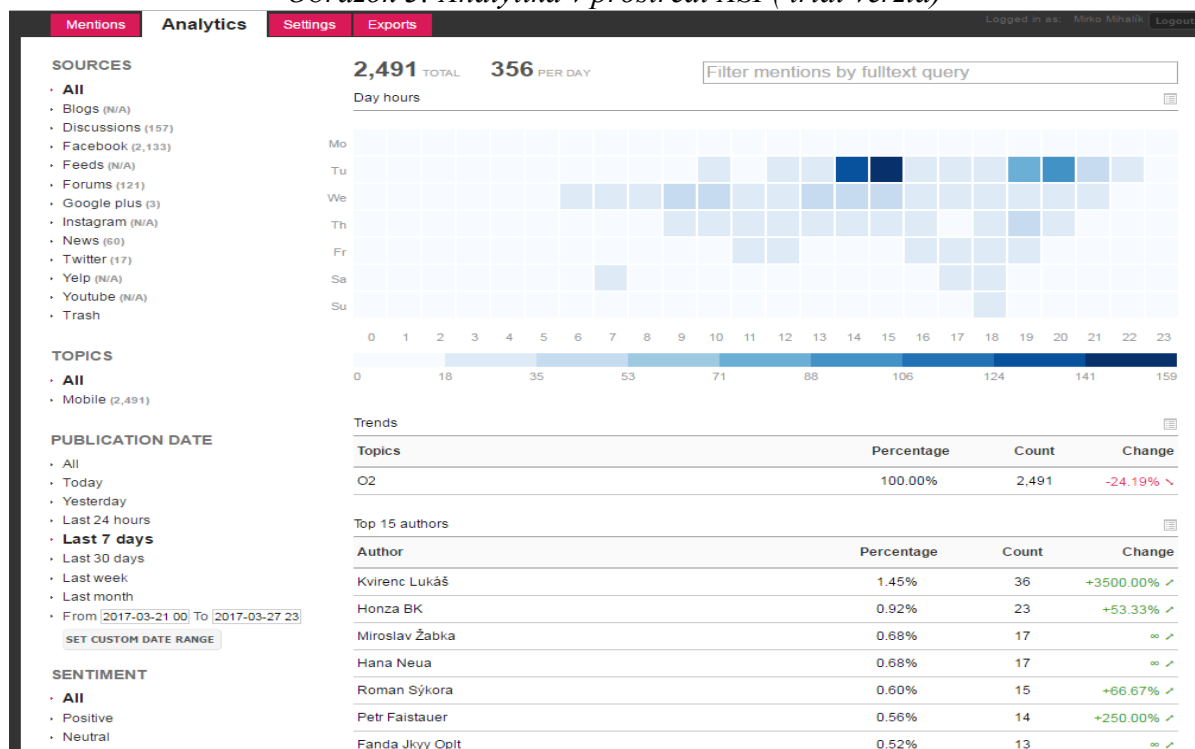


Zdroj: Ataxo - Social Insider [online]. Dostupné na WWW: <http://ataxosocialinsider.cz>

Pre ilustráciu, ako sa dá s vyhľadávaným príspevkom v aplikácii pracovať, je na obrázku 2 zobrazený príspevok vyhľadávaný podľa kľúčového slova "O2", slovo je zvýraznené. Každý príspevok obsahuje dátum, kedy bol publikovaný, vlastný text, meno autora, zdroj, bydlisko a v prípade Twitteru ešte počet "followers", teda tých ktorí používateľa sledujú a počet "friends", tých, ktorých používateľ sleduje. Ďalej je tam zobrazená ešte hodnota "klout", ktorá symbolizuje jeho vplyv na sociálnych sieťach. V pravom dolnom rohu sú tlačidlá na označenie

sentimentu, na odoslanie príspevku e-mailom, na zapísanie poznámky k príspevku a nakoniec sa tam nachádza tlačidlo na odstránenie príspevku z výberu. Export dát je možný hneď do štyroch formátov (CSV, XLXS, HTML a DOCX). Exportuje sa všetko, čo je vyhládané podľa navolených kritérií. Data potom obsahujú 12 kategórií, ako napríklad obsah príspevku, autora, dátum, odkaz na zdroj, označený sentiment, kľúčové slovo, a ďalšie údaje z Twitteru. Export je možný naplánovať denne, týždenne alebo mesačne.

Obrázok 3: Analytika v prostredí ASI (trial verzia)



Zdroj: Ataxo - Social Insider [online]. Dostupné na WWW: <http://ataxosocialinsider.cz>

V trial verzii je analytický nástroj avšak nie je nejako bohatá. Na obrázku 4 je porovnaný graf z platenej verzie, ktorá už ponúka vizuálne pútavejšie výstupy. Na hlavnej záložke Mentions je zobrazený iba vývoj počtu príspevkov v čase a dva prstencové grafy znázorňujú podiel vyhládaných zdrojov a podiel označeného sentimentu (obe v %). Ďalšia záložka Analytics už je niečo zaujímavejšia. Na obrázku 3 je zachytený počet príspevkov na Twitteri za posledných 7 dní podľa kľúčového slova "O2". Čím tmavšie políčko, tým viac príspevkov, ako je vidno na modrej škále pod grafom. Z grafu je možné vyčítať, že utorok medzi 14. a 15. hodinou bolo napísaných najviac príspevkov - 159. Pod grafom sú ďalšie prehľady ale tie už nie sú grafické. Ako prvé sú trendy, teda aký bol vývoj príspevkov s kľúčovým slovom v danom týždni oproti minulému. Ďalej sú tam jednoduché štatistiky najviac čítaných príspevkov, ako napríklad rebríček 15 autorov, ktorý najviac prispievali, 15 najčastejších "hashtagov" vo vybraných príspevkoch, najčastejšie písane emotikony a najčastejšie spomínané odkazy a domény v príspevkoch. Ako už bolo zmienené, ASI je lokalizovaný pre Českú a Slovenskú republiku. Pri hľadaní kladie dôraz aj na skloňovanie slov. Avšak so slovenskými slovami má značný problém, preto je lepšie využívať české slová.

Obrázok 4: Analytika v prostredí ASI (plná verzia)



Zdroj: Ataxo - Social Insider [online]. Dostupné na WWW: <http://ataxosocialinsider.cz>

4.2 Sentiment 140

Sentiment140 je na webe voľne prístupný nástroj na analýzu sentimentu. Ide o akademický projekt troch študentov počítačových vied zo Standfordskej univerzity z roku 2009. Pomocou nástroja Sentiment140, predtým známeho ako Twitter Sentiment, je možné objaviť sentiment značiek, produktov a tém iba na sociálnej sieti Twitter. Klasifikátor funguje na princípoch strojového učenia, konkrétne využíva metódu klasifikácie maximálnej entropie. Klasifikátor bol naučený na dátach, ktoré boli stiahnuté z Twitter API na základe emotikonov, za predpokladu, že príspevky obsahujúce :) sú pozitívne a príspevky obsahujúce :(sú naopak negatívne. Klasifikátor dosahuje viac ako 80 % úspešnosť.⁶

Nástroj klasifikuje príspevky písane iba v anglickom a španielskom jazyku, takže pre slovenčinu nie je vhodný. Avšak podnik ho môže využiť pre vlastnú analýzu, pokiaľ tweetuje v anglickom jazyku. Tento nástroj je zameraný na zahraničný trh alebo aj ako monitoring konkurencie v zahraničí.

Používateľské rozhranie webovej aplikácie je veľmi jednoduché. Na obrázku 6 je hlavná stránka, ktorá sa načíta pri autorizovaní aplikácie cez vlastný účet na Twitteri. Stačí zadať ľubovoľné slovo či slovné spojenie, vybrať jazyk (angličtina alebo španielčina) a vyhľadať.

⁶Sentiment140 [online] from WWW: <http://www.sentiment140.com>

Obrázok 5: Úvodná stránka Sentiment140



Zdroj: Sentiment140 [online] from <http://www.sentiment140.com>

Výsledky vyhľadávani sú jednoduché a prehľadné. Je otázkou, či je to skôr výhodou (rýchla a jednoduchá orientácia), alebo nevýhodou (napr. obmedzenia výsledkov a manipulácia s nimi). Sentiment140 tiež nezobrazuje historické výsledky, ktoré siahajú viac do minulosti, maximálne niekoľko dní späť. Je to zásluhou obmedzenia Twitter API. Aj napriek svojim obmedzeným výsledkom a manipuláciou s nimi, Sentiment140 je rýchly a prehľadný nástroj pre analýzu sentimentu, ktorý je bezplatný a úspešný. Autori na stránkach priznávajú, že oblasť analýzy sentimentu má ešte veľa nevyriešených problémov. Niektorými záležitosťami sa musia v budúcnosti zaoberať a nástroje tak vylepšovať. Napríklad vyriešenie problémov s negáciou, a tiež, aby si lepšie poradili s neformálnym štýlom jazyka.

5 Záver

V porovnaní so súčasnými nástrojmi, je veľmi pravdepodobné, že v budúcnosti by analýza sentimentu mohla spĺňať oveľa viacej kritérií. Ideálny nástroj by mohol poskytovať nasledujúce funkcie:

- Vyhľadávanie ľubovoľného slova, skratky, symbolu či emotikony, ktoré sa objavia v ľubovoľnom príspevku.
- Pokročilé vyhľadávanie obsahujúce vyhľadávanie podľa dát, používateľa alebo značky.
- Rozlišovanie medzi príspevkami od určitého používateľa, príspevky pre určitého používateľa a príspevky ktoré sa zmieňujú o používateľovi.
- Generovanie farebných grafov znázorňujúcich odlišné nálady
- Možnosť vyhľadávania za dlhé časové obdobie
- Trvalé grafy znázorňujúce trendy a relevantné príspevky
- Minimálne 90 percentnú presnosť vyhodnocovania príspevku
- Možnosť aktualizácie v reálnom čase

Žiadny so súčasných nástrojov nedokáže splniť všetky podmienky a dokonca sa im niektoré nástroje len z ďaleka približujú. Niektoré sú na dobrej ceste a ich výsledky sa každou novou verziou zlepšujú. Pravdepodobným smerom, ktorým sa vydajú nástroje analýzy sentimentu v ďalších rokoch, bude zlepšovanie súčasných metód a väčšie zapojenie umelej inteligencie. Nástroje by sa mohli učiť zo svojich chýb, čo bude vyžadovať, aby im niekto povedal, že urobili chybu. Tieto algoritmy budú veľmi pravdepodobne podporené štatistickými a pravdepodobnostnými algoritmami, ktoré budú využívať pri svojom následnom rozhodovaní. V neposlednej rade by mali využiť podobný princíp, ako je v projekte Sentiment140. Vedci zo Standfordskej univerzity využili ľudí, aby vytvorili tabuľky kľúčových slov, s ktorými neskôr pracujú algoritmy. Ľudský faktor bude treba využiť pre maximálnu presnosť vyhodnocovania. Mohlo by to vyzeráť tak, že by človek robil finálnu kontrolu nad vyhodnocovacím algoritmom. Kombinácia umelej inteligencie, štatistiky, pravdepodobnosti a ľudskej manuálnej kontroly sa javí ako riešenie budúcnosti

Literatúra

- [1] Ataxo - Social Insider [online]. from <http://www.ataxosocialinsider.cz/>. *Methodological manual*
- [2] Dey, L., & Haque, S. K. (n.d.). *Opinion mining from noisy text data*. Retrieved August 28, 2017, from <https://www.semanticscholar.org/paper/Opinion-mining-from-noisy-text-data-Dey-Haque/5d7cff09bfa17ec2f45b7e0731bd66ed773c4ae5/figure/15>
- [3] Di Bo Pang (Autore), Lillian Lee (Autore). (n.d.). *Opinion Mining and Sentiment Analysis* (Inglese) Copertina flessibile – 8 lug 2008. Retrieved August 28, 2017, from <https://www.amazon.it/Opinion-Mining-Sentiment-Analysis-Pang/dp/1601981503>
- [4] Sentiment140 [online] from <http://www.sentiment140.com>