

Cílem bakalářské práce je vytvořit návrh služby pro podporu vyhledávání vhodných kandidátů "product placement" reklamy na sociálních sítích.

1. Stručně popište historii sociálních sítí.
2. Popište současný stav řešení problematiky pro danou formu marketingu (product placement).
3. Navrhněte proces, který pro sociální sítě Facebook a Twitter:
  - a) získá data o uživatelích,
  - b) vybere uživatele, kteří mají velkou sledovanost
  - c) analyzuje kvalitu "následovníků" daného uživatele (např. místo bydliště, věk, zájmy),
  - d) vypočítá rating uživatele a přiřadí k němu vhodné značky (tagy),
  - e) na dotaz od zákazníka vrátí nabídku vhodných kandidátů pro reklamu produktu.
4. Navrhněte vlastní řešení a implementujte ve vhodném vývojovém prostředí pilotní verzi.



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ  
KATEDRA SOFTWAREVÉHO INŽENÝRSTVÍ



Bakalárska práca

## **Návrh služby pro podporu prodeje reklamy na sociálních sítích**

*Martin Vaňo*

Vedúci práce: Ing. Petra Pavlíčková, Ph.D.

5. mája 2015



---

## Pod'akovanie

Rád by som sa poďakoval rodičom za ich podporu počas štúdia, Ing. Petre Pavlíčkovej, Ph.D. za vedenie práce, jej odborné pripomienky a konzultácie a v neposlednom rade sa chcem poďakovať aj spolužiakom z FITu, na ktorých pomoc som sa mohol spoľahnúť počas celého štúdia.



---

# Prehlásenie

Prehlasujem, že som predloženú prácu vypracoval(a) samostatne a že som uviedol(uviedla) všetky informačné zdroje v súlade s Metodickým pokynom o etickej príprave vysokoškolských záverečných prác.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona, v znení neskorších predpisov. V súlade s ustanovením § 46 odst. 6 tohoto zákona týmto udeľujem bezvýhradné oprávnenie (licenciu) k užívaniu tejto mojej práce, a to vrátane všetkých počítačových programov ktoré sú jej súčasťou alebo prílohou a tiež všetkej ich dokumentácie (ďalej len „Dielo“), a to všetkým osobám, ktoré si prajú Dielo užívať.

Tieto osoby sú oprávnené Dielo používať akýmkoľvek spôsobom, ktorý nezníži hodnotu Diela, a za akýmkoľvek účelom (vrátane komerčného využitia). Toto oprávnenie je časovo, územne a množstevne neobmedzené. Každá osoba, ktorá využije vyššie uvedenú licenciu, sa však zaväzuje priradiť každému dielu, ktoré vznikne (čo i len čiastočne) na základe Diela, úpravou Diela, spojením Diela s iným dielom, zaradením Diela do diela súborného či zpracovaním Diela (vrátane prekladu), licenciu aspoň vo vyššie uvedenom rozsahu a zároveň sa zaväzuje sprístupniť zdrojový kód takého diela aspoň zrovnateľným spôsobom a v zrovnateľnom rozsahu ako je zprístupnený zdrojový kód Diela.

V Prahe 5. mája 2015

.....

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2015 Martin Vaňo. Všetky práva vyhradené.

*Táto práca vznikla ako školské dielo na FIT ČVUT v Prahe. Práca je chránená medzinárodnými predpismi a zmluvami o autorskom práve a právach súvisiacich s autorským právom. K jej využitiu, s výnimkou bezplatných zákonných licencií, je nutný súhlas autora.*

### **Odkaz na túto prácu**

Vaňo, Martin. *Návrh služby pro podporu prodeje reklamy na sociálních sítích*. Bakalárska práca. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2015.



---

## Abstrakt

Cieľom této práce je automatizovať proces vyhľadávania vhodných kandidátov pre reklamu na sociálnych sieťach. Podstatou procesu vyhľadávania takýchto kandidátov je analýza užívateľov sociálnych sietí na základe ich profilu, nasledovníkov a verejných konverzácií, ktoré vedú s ostatnými užívateľmi. Súčasťou programu je tiež návrh jednotlivých typov dátového úložiska pre efektívne spracovanie dát a vyhľadávanie. Výsledkom je systém, v ktorom je možné na základe požiadavky od zákazníka jednoducho vyhľadať užívateľov, ktorí majú vplyv na jeho cieľovú skupinu.

**Kľúčová slova** influence–marketing, sociálne siete, twitter, facebook, influencer, follower, java

---

## Abstract

The objective of this work is to simplify process of searching relevant candidates for advertisement purposes on various social networks. This goal is reached by creating a process for evaluating the users of social network according to number and quality of their followers and their public conversations. Part of this work is a proper design of data storage with concern to effective data

processing and querying. The result of this work is simply a system which returns relevant candidates that are suitable for marketing purposes according to customer requirements based on his target customer group.

**Keywords** influence–marketing, social networks, twitter, facebook, influencer, follower, java

---

# Obsah

Úvod	1
<b>1 Ciele práce a metodika</b>	<b>3</b>
1.1 Ciele . . . . .	3
1.2 Metodika . . . . .	3
<b>2 Rešeršná časť</b>	<b>5</b>
2.1 Sociálne siete . . . . .	5
2.2 Marketingové techniky . . . . .	9
2.3 Existujúce riešenia . . . . .	10
2.4 Porovnanie súčasných riešení . . . . .	15
2.5 Zhrnutie kapitoly . . . . .	17
<b>3 Analýza a návrh</b>	<b>19</b>
3.1 Čo je nevyhovujúce v súčasných riešeniach . . . . .	19
3.2 Novovznikajúce problémy . . . . .	21
3.3 Skutočné faktory pôsobiace na vplyv užívateľa . . . . .	21
3.4 Proces výpočtu . . . . .	22
3.5 Stiahnutie dát zo sociálnych sietí . . . . .	22
3.6 Uloženie dát . . . . .	23
3.7 Analýza dát . . . . .	24
3.8 Prípady použitia . . . . .	28
3.9 Zhrnutie kapitoly . . . . .	42
<b>4 Implementácia</b>	<b>43</b>
4.1 Technológie . . . . .	43
4.2 Implementácia prípadov použitia . . . . .	43
4.3 Komponenty programu . . . . .	43
4.4 Zhrnutie kapitoly . . . . .	45

<b>5 Overenie riešenia</b>	<b>47</b>
5.1 Proces overenia . . . . .	47
5.2 Zhrnutie kapitoly . . . . .	49
<b>6 Doporučenie ďalšieho postupu</b>	<b>51</b>
6.1 Finalizácia implementácie . . . . .	51
6.2 Doplnenie ďalších analýz . . . . .	51
<b>Záver</b>	<b>53</b>
<b>Literatúra</b>	<b>55</b>
<b>A Zoznam použitých skratiek</b>	<b>59</b>
<b>B Slovník a vymedzenie pojmov</b>	<b>61</b>
<b>C Použité programy</b>	<b>63</b>
<b>D Obsah priloženého CD</b>	<b>65</b>

---

## Zoznam obrázkov

3.1	Zmena Klout skóre . . . . .	20
3.2	Prehľad analýzy dát . . . . .	25
4.1	Užívateľské rozhranie implementovaného riešenia . . . . .	44
4.2	Náhľad na architektúru systému . . . . .	46
5.1	Profil ČVUT na sociálnej sieti Twitter a jeho vybrané vzťahy . . .	48



---

## Zoznam tabuliek

2.1	Rozdiely medzi sociálnymi sieťami . . . . .	8
2.2	Zjednotenie pojmov . . . . .	8
2.3	Prehľad existujúcich riešení . . . . .	11
2.4	Porovnanie existujúcich riešení . . . . .	16





---

# Úvod

Sociálne siete zažívajú v posledných desaťročiach nevídaný rozmach. Facebook, Twitter a ďalšie sociálne siete považujú mnohí ľudia za neoddeliteľnú súčasť svojich životov. Ulahčujú im komunikovanie s ostatnými užívateľmi, zdieľanie svojich názorov a v neposlednom rade im poskytujú rýchly prístup k novým informáciám, čím často zastávajú úlohu tradičných spravodajských portálov. Tento rozmach sociálnych sietí však položil predpoklady pre vyprofilovanie úzkej, veľmi vplyvnej skupiny ľudí, ktorých aktivitu sleduje veľký počet ostatných užívateľov, čím táto úzka skupina ľudí získava moc nad obsahom, ktorý budú ostatní užívatelia sociálnej siete konzumovať.

Začiatok nápadu analyzovať dáta na sociálnych sieťach za účelom identifikovania vplyvných ľudí sa datuje do obdobia leta 2014, kedy som na webovom portáli Zoznam.sk zhliadol video *Príbeh Brita: Za sekundu zarobí neuveriteľných 2-tisíc libier*[1]. Po zhliadnutí tohto videa mi ihneď napadla myšlienka, ako sú marketéri spoločností schopní vyhľadávať ľudí, ktorí by mohli využiť svoj potenciál na sociálnej sieti k rozšíreniu povedomia alebo priamo zvýšeniu predaja danej značky. Ako programátorovi mi samozrejme napadla myšlienka, že by tento proces bolo možné automatizovať.

Obsahom a cieľom tejto práce je porozumenie problematike analyzovania dát zo sociálnych sietí za účelom identifikácie vplyvných užívateľov. Neoddeliteľnou súčasťou tejto úlohy je aj porozumenie fungovania už existujúcich platforiem ohodnocujúcich užívateľov a tiež zoznámenie sa s rôznymi marketingovými technikami súvisiacimi s touto úlohou. Konkrétne hovoríme o podpore product–placement marketingu, ktorý sa na sociálnych sieťach prezentuje formou nového marketingového postupu influence–marketing.



---

# Ciele práce a metodika

## 1.1 Ciele

Cielom tejto práce je v prvom rade získať dostatočný prehľad o existujúcich platformách pre podporu product–placement reklamy na sociálnych sieťach. Analyzovať ich výhody, nevýhody, pochopiť hlbšie podstatu ich použitia pre influence–marketing a takto získané znalosti následne využiť pri *analýze a návrhu* vlastného riešenia.

Druhým cieľom práce je nadobudnúť skúsenosti s efektívnym získavaním, ukladaním a spracovávaním dát zo sociálnych sietí, čo zahŕňa naštudovanie nových technológií, vyhľadanie a využitie externých programových knižníc a používanie rôznych druhov databázových riešení pre rôzne typy spracovávaných dát.

Tretím a posledným cieľom je všetky tieto získané skúsenosti zúročiť pri implementácii pilotnej verzie systému, v ktorej bude spracovaných niekoľko prípadov použitia z analýzy a návrhu a ktorá bude slúžiť pre overenie validity navrhovaného riešenia.

## 1.2 Metodika

Metodika bakalárskej práce je založená na dôkladnej analýze existujúcich platforiem venujúcich sa vyhodnocovaniu skóre jednotlivých užívateľov na základe ich vplyvu na sociálnych sieťach. Táto analýza je podporená štúdiom knižných a online zdrojov odbornej literatúry venujúcich sa danej problematike. Na základe znalostí získaných štúdiom vyššie spomenutých zdrojov a tiež existujúcich platforiem budú identifikované nedostatky, na základe ktorých bude navrhnutý vhodný postup výpočtu influence–skóre. Tento postup bude detailne rozpracovaný v prípadoch použitia aplikácie. Vybrané prípady použitia aplikácie budú pre praktické overenie validity riešenia implementované do pilotnej verzie aplikácie. Na základe výsledkov overenia riešenia bude doporučený ďalší postup a prípadné návrhy na úpravu riešenia.



---

## Rešeršná časť

### 2.1 Sociálne siete

Článok „Users of the world, unite! The challenges and opportunities of Social Media“ pomerne stručne a výstižne definuje pojem sociálne siete:

*„Social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content“ [2, str.61].*

Pre detailnejšie pochopenie kontextu vytvorenia a súčasného masívneho používania sociálnych sietí je vhodné porozumieť podstate vzniku internetu a prvých portálov združujúcich ľudí online.

#### 2.1.1 História sociálnych sietí

##### 2.1.1.1 Začiatky

V roku 1969 sa podarilo Americkou vládou financovanému projektu počítačovej siete ARPANET preniesť prvú správu medzi dvoma počítačmi. Po tomto roku vzniklo niekoľko projektov, ktoré umožňovali komunikáciu užívateľov medzi viacerými počítačmi (napríklad BBS), no stále sa jednalo o veľmi obmedzené služby, ktoré umožňovali maximálne sťahovanie súborov a software z centrálného úložiska alebo posielanie krátkych správ ostatným užívateľom.

Spoločnosť CompuServe v sedemdesiatych rokoch 20. storočia prišla s revolučnou myšlienkou zriadiť diskusné fóra (mimo iného poskytovala aj služby zdieľania súborov a zobrazovania aktuálnych správ), kam užívatelia mohli prispievať vlastné príspevky k aktuálnym témam a mohli priamo komunikovať s ostatnými užívateľmi. Tieto fóra boli najstaršími predchodcami dnešných svetovo populárnych diskusných portálov, ako napríklad reddit.

Prvý predchodca sociálnych sietí je však americká spoločnosť AOL, ktorá začiatkom 90 rokov vytvorila platformu, v ktorej bolo možné vytvárať užívateľské profily a následne tieto profily aj vyhľadávať (mimo iného samozrejme

aj komunikovať s ostatnými užívateľmi), čo je v súčasnej dobe hlavný rys sociálnych sietí.

Tieto služby boli však stále značne obmedzené v porovnaní s dnešnými modernými sociálnymi sieťami, a to hlavne menšou užívateľskou základňou (súvisiacou s nízkou vybavenosťou domácností výpočtovou technikou) a tiež veľmi obmedzenou rýchlosťou vtedajšieho pripojenia.

Vďaka novovytvoreným možnostiam v polovici 90-tých rokoch 20. storočia, kedy vznikali prvé online nákupné platformy ako Amazon a Yahoo a sa dramaticky zvýšila dostupnosť domácich počítačov sa vytvorili priaznivé podmienky pre vznik sociálnych sietí, sietí, ako ich poznáme v dnešnej dobe [3].

### 2.1.1.2 Prvé moderné sociálne siete

Rok 2002 sa dá označiť za rok, kedy začali vznikať sociálne siete, ako ich poznáme dnes. Prvou takouto službou bol *Friendster*, ktorý bol postavený na myšlienke, že bohatá online komunita môže existovať iba medzi ľuďmi, ktorí majú medzi sebou nejaké väzby. V prvom roku služba *Friendster* získala viac ako 3 milióny užívateľov a taktiež mnoho investícií, no po technických ťažkostiach sa *Friendster* vzdal segmentu sociálnych sietí a v súčasnosti existuje iba ako herná platforma. O rok neskôr, v roku 2003 vznikli dve pomerne známe a stále fungujúce sociálne siete LinkedIn a MySpace. Zatiaľ čo LinkedIn si vzal za cieľ združovať profesné kontakty (väzby medzi užívateľmi na LinkedIn sa volajú *spojenia*, radšej než *priateľstvá*, ako na iných sociálnych sieťach), MySpace sa rozhodol cieľiť na komunitu mladých ľudí formou zdieľania hudby, videí a príjemného rozhrania, čo malo za následok prechod užívateľov služby *Friendster* k MySpace a teda definitívneho úpadku *Friendster*–u ako sociálnej siete. Ani MySpace sa však nedokázal vysporiadať s príchodom Facebooku (2004), kedy tejto službe začali klesať čísla návštevnosti, a tak v súčasnosti pôsobí hlavne ako sociálna sieť zameraná na hudobníkov a hudobné skupiny.

### 2.1.1.3 Úspech súčasných sociálnych sietí

Faktorov vplývajúcich na úspech súčasných sociálnych sietí je niekoľko, no ten najhlavnejší je potrebné hľadať v ľudskej psychike. Ľudia majú prirodzenú potrebu kontrolovať iných ľudí, udalosti dejúce sa okolo nich a v čo najväčšej miere mať kontrolu nad dojmom, ktorý vzbudzujú v ostatných ľuďoch. Táto myšlienka je podporená aj v nasledujúcej citácii:

*With respect to the social dimension of Social Media, the concept of self-presentation states that in any type of social interaction people have the desire to control the impressions other people form of them (Goffman, 1959). On the one hand, this is done with the objective of influencing others to gain rewards (e.g., make a positive impression on your future in-laws); on the other hand, it is driven by a wish to create an image that is consistent with one's personal*

*identity (e.g., wearing a fashionable outfit in order to be perceived as young and trendy) [2, str.62].*

Na základe tejto myšlienky sa v súčasnej dobe zaraďujú medzi najúspešnejšie sociálne siete tie, ktoré dokážu poskytnúť svojim užívateľom funkcie, ktorými môžu vytvárať, manažovať a kontrolovať dojem (mienku) o svojej osobe u iných užívateľov.

Je taktiež potrebné si uvedomiť druhý fakt, že aj keď sociálne siete modernej doby sú postavené na myšlienke zdieľania obsahu, reálne väčšinu obsahu vytvára iba malá skupinka užívateľov, zatiaľ čo 99% užívateľov tento obsah iba pasívne konzumuje (títo užívatelia sú ďalej v texte referovaní ako macro-influenceri).

### 2.1.2 Facebook

Facebook (pôvodne theFacebook) je sociálna sieť založená v roku 2004 na Harvardovej univerzite Markom Zuckerbergom. Pôvodne sieť slúžila výlučne študentom tejto univerzity, po pár mesiacoch sa však rozšírila aj na ostatné univerzity (hlavne školy Ivy League), až sa nakoniec v roku 2006 otvorila pre verejnosť.

O úspechu Facebooku sa vedú mnohé polemiky. Mnoho ľudí sa ale zhodne, že počítačová exkluzivita registrovania do služby, jednoduchosť používania, stabilita a v neposlednom rade aj chytľavé meno prispeli k masívnemu odlivu užívateľov iných sociálnych sietí práve k Facebooku. Facebook je tiež autorom niekoľkých veľmi inovatívnych myšlienok poslednej doby, ako napríklad možnosť využívať sociálnu sieť, aj keď sa užívateľ nenachádza priamo na jej webovej stránke (napríklad formou „Like“ tlačidla alebo pridania komentáru na väčšine webových stránok).

V súčasnosti Facebook využíva 1.39 miliardy užívateľov mesačne [4] a jedná sa teda o najpoužívatejšiu sociálnu sieť na svete. Facebook vstúpil na burzu v máji 2012, pričom o rok neskôr jeho kapitalizácia prekročila hranicu 100 miliárd dolárov.

### 2.1.3 Twitter

Twitter je mikro-blogovacia sociálna sieť, ktorá svojim užívateľom umožňuje pridávať príspevky do veľkosti 140 znakov. Založená bola v roku 2006 Jackom Dorseyom. V súčasnej dobe má 288 miliónov aktívnych mesačných užívateľov [5].

Myšlienka a úspech sociálnej siete Twitter spočíva vo fakte, že dnešná generácia ľudí mladších ako 35 rokov dostáva denne tak veľa podnetov (reklamy, správy, mobilné aplikácie a rôzne iné sociálne siete), že sú omnoho náchyľnejší k rozptýleniu a teda neschopní dlhodobo sa koncentrovať ako staršie generácie. Z tohto dôvodu ich veľmi jednoducho odradí čítanie dlhých

Tabuľka 2.1: Rozdiely medzi sociálnymi sieťami

Oblasť	Facebook	Twitter
Sledovanie trendov	✗	✓
Pridávať priateľov	✓	✗
Dĺžka správy	10 000 znakov	140 znakov
Editácia správ	✓	✗

Tabuľka 2.2: Rozdielne pomenovania funkcií na Facebooku a Twitteri

Funkcia	Facebook	Twitter	Táto práca
A	Post	Tweet	Správa/Príspevok
B	Like	Star	Páči sa
C	Comment	Reply	Komentár
D	Share	Retweet	Zdieľanie
E	Mention	Mention	Zmienka

správ alebo pozeranie videí, ktoré majú viac ako 3 minúty. Twitter teda vsadil na možnosť písať krátke správy, takzvané mikro-blogy, ktoré je možné veľmi rýchlo prečítať a doplniť o hyperlinky alebo obrázky. Správy navyše využívajú špecifické prefixy pre užívateľov, ktorých prefixuje znakom (napríklad *#vanomart*), alebo témy (presnejšie tagy - používané ďalej v texte) prefixuje znakom # (napríklad *#volby2014*).

### 2.1.4 Facebook a Twitter - porovnanie

Spoločným znakom uvedených (a iných moderných sociálnych sietí) je jednoduchá integrovateľnosť s inými webovými alebo mobilnými aplikáciami pomocou ich verejne prístupného API. Užívateľia môžu štandardne prispievať správami, nahrávať na sieť obrázky a videá, prípadne zdieľať obsah od iného užívateľa. Medzi Facebookom a Twitterom však existujú isté špecifické rozdiely, ktoré sú uvedené v tabuľke 2.1.

#### 2.1.4.1 Vymedzenie pojmov

Facebook a Twitter používajú pre podobné funkcie rôzne názvy, preto je v tabuľke 2.2 uvedené jednotné názvoslovie, ktoré bude ďalej v texte používané. Jednotlivé skratky použité v tabuľke sú popísané nižšie.

- A - Užívateľ pridá na sociálnu sieť nový obsah.
- B - Užívateľovi sa páči aktivita iného užívateľa.
- C - Užívateľ pridá svoj názor na príspevok.



- D - Užívateľ vezme príspevok iného užívateľa a uverejní ho pod svojim profilom.
- E - Užívateľ spomenie v správe iného užívateľa.

## 2.2 Marketingové techniky

V súčasnosti existuje niekoľko techník využívaných pre vloženie reklamy produktu alebo služby do nereklamného prostredia. Všetky tieto marketingové techniky je možné zastrešiť pod pojem product-placement, ktorý je vysvetlený v 2.2.1.

Product-placement na sociálnych sieťach je pomerne nová marketingová technika, ktorá už ale má svoje vlastné, špecifické pomenovanie, a to „influence-marketing“. Influence-marketing čerpá zo základov product-placementu, ale jeho veľmi výrazným črtom je **schopnosť šíriť správu ústnym podaním** (bližšie špecifikované v 2.2.2), ktoré úspešne v ľudskej spoločnosti funguje už od nepamäti. V konečnom výsledku je influence-marketing spojením dvoch existujúcich marketingových techník s využitím sociálnych sietí.

### 2.2.1 Product-placement

Product-placement (doslovne „umiestnenie produktu“) je technika zasadenia komerčného výrobku do umeleckých diel. Jedná sa o formu skrytej reklamy, teda takú formu reklamy, ktorú je možné veľmi ťažko odlišiť od očakávaného obsahu. Product-placement sa typicky vyskytuje vo filmoch a seriáloch, kde hlavní hrdinovia používajú výrobky istej firmy, ktorá si za to zaplatí (napríklad nápoje Coca Cola alebo autá Ford). Môže sa vyskytovať ale aj v agresívnejšej forme, a to vo forme vkladania krátkych sekvencií zobrazujúcich logo firmy do filmu, ktoré majú podvedome evokovať prijímateľovi potrebu kúpiť si výrobky tejto firmy.

### 2.2.2 Word-of-mouth marketing

Reklama prenášajúca sa ústnym podaním je súčasťou ľudskej interakcie od nepamäti. Ľudia si touto formou medzi sebou odporúčajú výrobky a služby, ktoré oni osobne využívajú a sú s nimi spokojní. Nejedná sa teda o platenú reklamu (do tej miery, že by za ňu dostal odporúčiteľ zaplatené), ale spontánnu reakciu na kvalitný produkt. Tento druh reklamy môže mať samozrejme aj opačný efekt, ak sa jedná o nekvalitný produkt alebo službu.

V nedávnej dobe vzniklo celé odvetvie zaoberajúce sa word-of-mouth marketingom (WOMM), ktoré na rozdiel od jeho prirodzenej podoby podporuje ľudí, ktorí aktívne odporúčajú produkty a služby svojim známym. Môže sa jednať o celé spoločnosti založené na multi-level marketingu alebo jednotlivcov

fungujúcich na vizitkovom systéme (zákazník, ktorý privedie nového zákazníka získa zľavu alebo finančnú odmenu). Aj keď je WOMM pomerne nová technika, zaraďuje sa medzi prístupy, ktoré majú rastúcu tendenciu, a teda je na ňu každým rokom vyčlenených viac a viac finančných prostriedkov [6].

### 2.2.3 In-game advertising

Súčasťou dnešných moderných hier sú často výrobky úspešných spoločností produkujúcich napríklad automobily alebo športové výrobky. Reklama tohto formátu zaznamenala v posledných rokoch rapídny vzostup, aj keď jej počiatky sa datujú už do roku 1978. Jedná sa o špecifický druh product-placement reklamy.

Forma zobrazenia reklamy je obmedzená iba fantáziou, a preto môže byť reklama do hry integrovaná napríklad na pozadí počas načítavania hry či ako billboard v hráčskom prostredí, no môže byť taktiež neoddeliteľnou súčasťou hry [7] (napríklad autá v hernej sérii Need for Speed [8]).

### 2.2.4 Influence-marketing

Influence-marketing, ako už bolo v skratke spomenuté v úvode tejto sekcie, je pomerne nová marketingová technika, ktorá čerpá zo základov product-placementu, word-of-mouth marketingu a funguje na sociálnych sieťach. Práve sociálne siete poskytujú obrovský potenciál pre zber a analýzu dát o správaní užívateľov, čo následne môžu využiť firmy pre zvýšenie efektivity zobrazovanej reklamy.

Influence-marketing je charakteristický vyhľadávaním významných ľudí na sociálnych sieťach, ktorí majú vplyv na ostatných užívateľov sociálnych sietí (teda potenciálnych kupcov). Vplyvní užívatelia sú motivovaní propagovať za nejakú protihodnotu produkt, službu alebo priamo značku zákazníka, preto sa jedná o win-win scenár, kedy nikto zo zúčastnených strán neprichádza k strate. Je ale užitočné spomenúť, že sa jedná o istú formu skrytej reklamy, keďže nasledovníci daného vplyvného užívateľa nie sú oboznámení s faktom, že vplyvný užívateľ do svojich správ alebo videí vkladá takúto reklamu.

## 2.3 Existujúce riešenia

### 2.3.1 Vymedzenie konkurencie

V oblasti influence-marketingu v súčasne dobe existuje niekoľko široko používaných riešení aj napriek tomu, že toto odvetvie nie je tak vyvinuté ako napríklad podobne zamerané odvetvie SEO optimalizácie. Za **priame konkurenčné riešenie** je teda považovaný každý systém, ktorý je:

- schopný analyzovať užívateľov sociálnych sietí za účelom vyhodnotenia ich influence-skóre,

Tabuľka 2.3: Prehľad konkurencie

Priama konkurencia	Nepriama konkurencia
eCairn Conversation	trendspottr
oneQube	Klout
Onalytica	Kred
Tellagence	Appinions
Traackr	Empire Avenue
Peerindex	

- zameriavajúci sa na **priame kontaktovanie** vhodných užívateľov s ponukou spolupráce.

Medzi **nepriame konkurenčné riešenia** patria platformy, ktoré taktiež vyhodnocujú influence-skóre užívateľov, no za iným účelom než sú dva vyššie spomenuté body. Niektoré spoločnosti sa napríklad sústreďujú na sledovanie trendov vznikajúcich na sociálnych sieťach, merajú dopady internetových kampaní alebo zefektívňujú spravovanie firemných profilov naprieč viacerými sociálnymi sieťami, a influence-skóre je iba časťou ich procesu výpočtu.

### 2.3.2 Zoznam existujúcich riešení

Tabuľka 2.3 obsahuje prehľadný výpis existujúcich riešení podľa delenia na priame a nepriame konkurenčné riešenia (2.3.1).

V nasledujúcich sekciách sú podrobnejšie popísané vybrané priame a nepriame konkurenčné riešenia.

### 2.3.3 Klout

Klout je momentálne najrozšírenejšia [9] služba analyzujúca influence-skóre užívateľov za použitia dát z najrozšírenejších sociálnych sietí. Joe Fernandez a Binh Tran ju založili v roku 2008. Vďaka veľkému počtu užívateľov má stabilné miesto na trhu a v roku 2014 bola spoločnosť predaná Lithium Technologies za 200 miliónov dolárov. Donedávna bol Klout pomerne pravidelne kritizovaný za časté zmeny algoritmu vyhodnocovania a neadekvátny výpočet influence-skóre.

Klout svojim algoritmom prideluje každému užívateľovi **Klout Score**, ktorým je číslo na stupnici 1..100. Klout motivuje užívateľov ku generovaniu obsahu a zvyšovaniu svojho influence-skóre službou **Klout Perks**, ktorá užívateľom s vysokým skóre ponúka rôzne výhody. Medzi tieto výhody patria napríklad bezplatné služby alebo produkty od spoločností, ktoré si od poskytnutia služby alebo produktu sľubujú to, že daný užívateľ s vysokým skóre tento produkt alebo službu spomenie vo svojom blogu alebo príspevku na sociálnu sieť, čím zvýši povedomie o tejto službe alebo produkte.

Proces výpočtu influence-skóre síce nie je verejný, avšak podľa svojej webovej stránky [10] berie do úvahy faktory ako: pomer vytvoreného obsahu a reakcie ostatných užívateľov na tento obsah, influence-skóre užívateľov, s ktorými je analyzovaný užívateľ prepojený, časovú zložku (s časom nadobudnuté influence prirodzene klesá) a celkový počet sociálnych sietí užívateľa.

### 2.3.4 Kred

Kred bol založený v roku 2011 spoločnosťou PeopleBrowsr a je považovaná za prirodzeného konkurenta pre Klout. Obe spoločnosti vypočítavajú influence-skóre užívateľov sociálnych sietí a tiež disponujú vlastnými spôsobmi odmeňovania užívateľov s vysokým skóre (pre Kred je to Kred Rewards). Na rozdiel však od služby Klout, Kred má verejný proces výpočtu influence-skóre, pre ktorý využíva Twitter a Facebook. Výsledkom jeho analýzy sú dve čísla:

- *Influence* - the ability to inspire action. It is scored on a 1,000 point scale.
- *Outreach* - reflects generosity in engaging with others and helping them spread their message [11].

*Influence* podľa vyššie uvedenej interpretácie odráža schopnosť inšpirovať akciu u ostatných užívateľov, *Outreach* na druhej strane vyjadruje mieru s akou influencer komunikuje so svojim publikom (zdieľanie správ a komentovanie).

### 2.3.5 OneQube

OneQube je spoločnosť založená v roku 2010 so sídlom v New Yorku, ktorej cieľom je uľahčenie práce so správou účtu na Twitteri. Súčasťou riešenia je niekoľko funkcionalít, napríklad manažovanie followerov, vytváranie Twitter zoznamov, identifikovanie relevantných diskusií a užívateľov (influencer analýza) a pochopenie demografického členenia followerov. OneQube taktiež na základe analýzy profilu na Twitteri zisťuje, či za profilom stojí jedinec alebo spoločnosť a aká je proporcia jednotlivcov a spoločností naprieč všetkými followermi užívateľa. Ako doplnok tejto analýzy poskytuje svojim klientom ďalšie údaje o ich followeroch, ako napríklad emailové adresy, odkazy na ich privátne weby (ktoré o sebe uviedli v profile) a prípadne aj lokáciu užívateľa. Takto spracovaní followeri môžu byť ďalej podrobení filtrovaniu na základe niekoľkých kritérií, ako napríklad kľúčové slová, pohlavie, vek, lokalita alebo počet followerov.

Je dôležité na záver uviesť, že OneQube sa nesústreďí primárne na vyhľadávanie nových influencerov, ale analýzu už existujúcich followerov zákazníka a poskytovanie rôznych štatistík s prípadnými odporúčaniami, ktorí followeri sú vplyvní v jeho už existujúcej sieti.

### 2.3.6 eCairn Conversation™

Conversation je služba materskej spoločnosti eCairn Inc. založenej v roku 2006, ktorá sa sústreďuje na identifikovanie rovnako zmýšľajúcich jedincov na sociálnych sieťach a tiež skúma, ako sú títo jedinci navzájom prepojení. Kľúčové funkcie eCairn Conversation™:

- Identifikovanie komunít - pre identifikovanie komunít s pevnými väzbami (cez nejakého influencera) a kto je tento influencer.
- Automatické vyhodnocovanie influencer-skóre pre rôzne typy influencerov v závislosti na ich marketingovom potenciáli (long/short-tail).
- Výskum konverzácií - zoznam správ z viacerých zdrojov sociálnych sietí.
- Filtrovanie výsledkov na základe typu zdroja, jeho lokality a témy.
- Conversation Explorer - získavanie a zobrazovanie správ z viacerých sociálnych sietí na jednom mieste.

### 2.3.7 Tellagence

Tellagence sa sústreďuje hlavne na zisťovanie, ako sa informácie presúvajú medzi jednotlivými komunitami. Zisťuje, kto sú kľúčoví influenceri majúci podiel na týchto presunoch a ktoré vzťahy medzi užívateľmi sociálnych sietí majú vysokú pravdepodobnosť zvýšenia povedomia o značke alebo produkte.

Tellagence berie pri vyhodnocovaní takýchto vzťahov nasledujúce údaje:

- záujmy,
- počet iných užívateľov zúčastňujúcich sa v konverzáciách užívateľa,
- hĺbka vzťahu (počet spoločných známych, počet spoločných záujmov),
- úroveň užívateľa vzhľadom k jeho komunite.

Po získaní uvedených dát vie Tellagence prediktívnou analýzou odhadnúť, aký dopad bude mať marketingová kampaň na komunitu influencera. Môže teda významným spôsobom prispieť k zefektívneniu online marketingových kampaní, keďže sa nesústreďuje na vyhľadanie čo najkvalitnejšieho influencera, ale na kvalitu komunít, relevanciu týchto komunít pre danú značku a vyhľadávanie influencerov, ktorí tieto komunity spájajú a majú v nich pevné väzby.

**Tellagence je založené hlavne na identifikovaní vzťahov medzi užívateľmi sociálnych sietí.**

### 2.3.8 Onalytica

Spoločnosť Onalytica sídliaca v Londýne bola založená v roku 2009. V roku 2013 spustila svoju platformu „Influencer Relationship Management“, ktorá umožňuje ich zákazníkom manažovať vzťahy s kľúčovými influencermi zákazníkovho odvetvia.

Súčasťou procesu vyhľadávania influencerov je podľa [12] identifikovanie veľkosti užívateľovej aktivity, kontext jeho konverzácií a veľkosť influencerovej komunity až do jeho kontaktov tretieho stupňa. Vedľajším produktom takejto analýzy je tiež graf influencerovej komunity.

### 2.3.9 Traackr

Traackr influence–marketing platforma založená v roku 2009, v súčasnej dobe čítajúca približne 5000 klientov. Traackr analyzuje dáta užívateľov sociálnych sietí na základe troch kritérií:

- Dosah - veľkosť publika, ktoré sleduje influencera.
- Rezonancia - ako je influencer schopný podnietiť diskusie.
- Relevancia - ako je influencerovo publikum prepojené so značkou zákazníka.

Na základe uvedených kritérií a cieľov spoločnosti zákazníka spoločnosť Traackr identifikuje užívateľov v cieľovej skupine zákazníka, ich vzájomné prepojenia a následne identifikuje kľúčových influencerov v daných sieťach. Traackr taktiež využíva analýzu sentimentu správ. **Traackr si zakladá na identifikovaní cieľových zákazníkov klienta a ich prepojenia.**

### 2.3.10 PeerIndex

PeerIndex je spoločnosť založená v roku 2007 v Londýne, ktorá analyzuje dáta užívateľov sociálnych sietí za účelom identifikovania influencerov na základe troch kritérií:

- Aktivita - aktivita užívateľa v porovnaní s inými užívateľmi sociálnej siete.
- Publikum - veľkosť publika (aj naprieč viacerými sociálnymi sieťami).
- Autorita - relevancia influencera pre jeho komunitu.

PeerIndex (podobne ako Kred a Klout) poskytuje užívateľom službu PeerPerks, ktorá odmeňuje influencerov *exclusive or early access to products and events before the general public* [13]. Pre spoločnosti, ktoré si nemôžu dovoliť takúto propagáciu (menšie lokálne firmy alebo startupy) je cieleňá služba PerkLets.

### 2.3.11 Appinions

Appinions je spoločnosť založená v roku 2008 v New Yorku. Jej proces vyhodnocovania influence-skóre je založený na viac ako 10 ročnom výskume Cornell-ovej univerzity. Appinions sa nespolieha primárne na dáta zo sociálnych sietí, ako to robia iné platformy, ale zbiera informácie aj z blogov, webových stránok a tiež z printových médií, ako sú napríklad tlačené noviny. Appinions následne analyzuje tieto zdroje a vytvorí kontext ku každej konverzácii, článku alebo blogu a identifikuje kľúčových ľudí, ktorí sa podieľajú na vytváraní tohto kontextu. Tieto správy sú navyše podrobnejšie analyzované analýze sentimentu, čo túto platformu odlišuje od väčšiny súčasne existujúcich riešení [14].

**Appinions si zakladá na získavaní a identifikovaní názorov na dané služby a veci.**

## 2.4 Porovnanie súčasných riešení

Súčasná platformy využívajúce influence-skóre a zaoberajúce sa influence marketingom sa zameriavajú na riešenie rôznych typov problémov. Niektoré využívajú influence-skóre ako nevyhnutnú časť svojho biznisu, iné iba ako doplnok k už existujúcim službám.

Služby ako Klout alebo Kred poskytujú meranie influence-skóre bez bližšie definovaného cieľa, takže ich výsledky môžu byť použité pre marketingové procesy, ale napríklad aj pre proces prijímania nových zamestnancov. Táto na prvý pohľad silná stránka je však paradoxne ich slabina, keďže v prípade špecifickejších vyhľadávacích kritérií neposkytujú najrelevantnejšie výsledky.

Iné spoločnosti (a platformy stojace za nimi) naopak stavajú na všeobecnosti vyššie spomenutých riešení a zameriavajú sa teda na riešenia špecifickejších problémov, ako napríklad nájsť vhodných užívateľov pre zvýšenie predaja produktu alebo služby (Tellagence, Traackr), alebo spravovať účty na sociálnych sieťach s pokročilou analýzou dát o užívateľoch (OneQube, Onalytica). Z toho vyplýva neprístupnosť služby pre verejnosť a teda orientácia služby výlučne na biznis zákazníkov.

Tabuľka 2.4 ponúka prehľad súčasných riešení a porovnania ich funkcionality v závislosti k navrhovanej službe.

Tabuľka 2.4: Porovnanie existujúcich riešení na základe šiest kritérií

Názov služby	PRK	ORV	ORI	GAM	OPI	VER
Klout			✓			✓
Kred			✓		✓	✓
eCairn Conversation	✓					
trendspottr						
oneQube		✓	✓			
Onalytica	✓	✓	✓			
Tellagence	✓	✓				
Appinions						
Traackr	✓	✓	✓			
Empire Avenue				✓		✓
Peerindex	✓					✓

- PRK - Priama konkurencia,
- ORV - Orientácia na vzťahy v komunitách,
- ORI - Orientácia na influencera,
- GAM - Gamifikácia procesu,
- OPI - Otvorenosť procesu vyhodnocovania skóre,
- VER - Verejnosť má prístup do služby.



### 2.4.1 Umiestnenie služby do existujúceho konkurenčného prostredia

Navrhované riešenie sa zaraďuje na trh medzi konkurentov, ktorí majú svoju platformu prístupnú iba pre klientov (nie Klout) a poskytujú pokročilejšie metodiky spracovania dát (napríklad eCairn Conversation). Jeho priamými konkurentmi by mohli byť už existujúce služby ako napríklad Tellagence, no tieto služby pôsobia prevažne v zahraničí (USA, UK) pričom v čase písania tejto práce neexistuje služba zaoberajúca sa influence–marketingom, ktorá by pôsobila v pevninskej časti Európy a ďalej na východ. Preto je vhodné zasadiť službu do vyššie spomenutého prostredia a sústrediť sa na spoločnosti operujúce na danom trhu (veľkosť spoločnosti zákazníka nehrá rolu).

### 2.4.2 Odlíšenie od konkurencie

Navrhované riešenie je inovatívne z pohľadu prístupu k micro a macro–influencerom. Zatiaľ čo ostatné služby túto otázku buď vôbec neriešia, alebo sa výlučne sústreďujú na identifikovanie iba jedného typu užívateľov, navrhované riešenie si kladie za úlohu prispôbiť sa potrebám zákazníka na základe skutočnosti, či ho motivuje zvýšenie povedomia o danej značke, produkte alebo službe (cielené vyhľadávanie macro–influencerov), alebo cieľ primárne na zvýšený predaj (vyhľadávanie micro–influencerov). Aj keď sa tieto dve skutočnosti nevylučujú, micro–influenceri majú oveľa väčší vplyv vo svojich komunitách a teda prispievajú k zvýšeniu predaja väčšou mierou, ako macro–influenceri. Tí naopak majú omnoho väčší vplyv na zvýšenie povedomia a rozprúdenie diskusií o danom produkte, službe alebo značke.

Druhá inovácia spočíva v identifikovaní miery vplyvu pre dané oblasti (tagy) užívateľa. Žiadna v súčasnosti existujúca platforma nerieši celkom zrejmy fakt, že väčšina užívateľov je vplyvná vo viacerých oblastiach, no v niektorých oblastiach sú títo užívatelia omnoho vplyvnejší ako v ostatných. Navrhované riešenie preto uvádza percentuálne zastúpenie vplyvu danej témy (oblasti) medzi ostatnými oblasťami, v ktorých je užívateľ na sociálnej sieti vplyvný.

## 2.5 Zhrnutie kapitoly

Moderné sociálne siete ako ich poznáme v súčasnosti začali vznikáť približne v roku 2002. S rapídny nárastom užívateľov existujúcich sietí a príchodom nových, špecificky zameraných sociálnych sietí (Pinterest, Twitter, Instagram) sa začali profilovať určití užívatelia, ktorých aktivitu a príspevky sledovalo veľa iných užívateľov, a nie iba úzky okruh ich priateľov a známych. Prvé platformy identifikujúce takýchto užívateľov (a vypočítavajúcich ich influence–skóre) sa začali formovať niekedy v roku 2007, čo neskôr prerástlo do vytvorenia úplne nového odvetvia marketingu - influence–marketingu. Jedná sa o techniku,

ktorá čerpá zo základov product-placementu a word-of-mouth marketingu, pričom využíva dáta pre analýzu zo sociálnych sietí.

Aj keď platformy pre vyhodnocovanie influence skóre užívateľov na sociálnych sieťach fungujú už približne 5 rokov, toto odvetvie je stále možné považovať iba za rozvíjajúce sa. Metódy získavania, ukladania a spracovania dát zo sociálnych sietí, ako aj metódy pre spracovanie prirodzeného jazyka výpočtovou technikou sa neustále zdokonaľujú, pričom firmy investujú stále väčší podiel finančných prostriedkov do marketingu na sociálnych sieťach. Tieto faktory značnou mierou prispievajú k vytváraniu tlaku na už existujúce influence-marketing platformy (pre poskytovanie relevantných údajov) a tiež otvárajú nové možnosti pre rozširovanie influence-marketing platforiem (existencia odvetví, ktoré zatiaľ nevyužívajú influence-scoring platformy).

---

## Analýza a návrh

### 3.1 Čo je nevyhovujúce v súčasných riešeniach

#### 3.1.1 Súčasné riešenia sú manipulovateľné

Čím viac je influence-scoring platforma otvorená bežnými užívateľom sociálnej siete a čím viac ju bežní užívatelia aj využívajú, tým je táto platforma náchylnejšia k ovplyvňovaniu výsledkov influence-skóre analýzy. Schopnosť užívateľa vidieť svoje influence-skóre ho na jednej strane motivuje k tomu, aby stále generoval obsah, ktorý bude rezonovať medzi ostatnými užívateľmi (a tým zvyšoval, alebo aspoň neznižoval dosiahnuté skóre), no taktiež to nábáda ovplyvňovať výsledky, ak skóre nie je dostatočne vysoké. Užívateľ tak môže napríklad vytvárať falošné profily, pomocou ktorých bude umelo zvyšovať svoje influence-skóre. Z tohto dôvodu je do istej miery pochopiteľný fakt, že väčšina influence-scoring platforiem je skôr uzavretá pre bežných užívateľov sociálnych sietí a influence-skóre používajú iba pre vytváranie zoznamov relevantných kandidátov na reklamu pre svojich klientov.

#### 3.1.2 Súčasné riešenia neodrážajú realitu

Na službe Klout sa dá demonštrovať ďalší fakt a to že skóre užívateľa niekedy nemusí odrážať jeho reálny vplyv medzi ostatnými užívateľmi sociálnej siete. Ukázané je to na nasledujúcom prípade:

Na Facebook bol dňa 19.2.2015 vložený obrázok, ktorý počas 24 hodín nazbieral viac ako 70 „páči sa“. Influence-skóre užívateľa sa zmenilo 34.42 na 41.84 (obrázok 3.1). Je možné, aby jeden nič nehovoriaci obrázok tak razantne zmenil influence-skóre užívateľa? Vypovedá toto skóre o tom, aký potenciál má užívateľ pre šírenie reklamy a motivovanie ostatných užívateľov k nákupu nejakého produktu alebo služby?

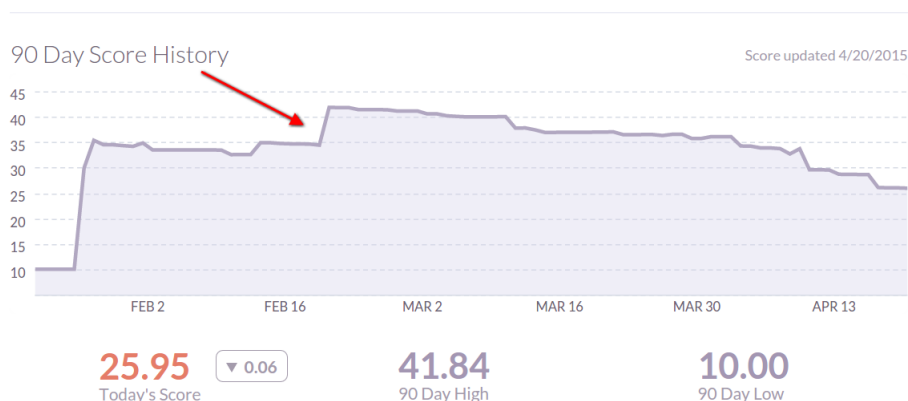
Ďalším prípadom je test [15], ktorý si položil jednoduchú otázku:

*Can you become influential on Twitter, and get a high Klout Score, merely by Tweeting a lot?*

### 3. ANALÝZA A NÁVRH

---

#### Measure and track your impact



Obr. 3.1: Zmena Klout skóre

Na základe výsledkov tohto testu vyplýva, že v prvom rade čím viac užívateľ pridáva na Twitter obsah, tým viac followerov si dokáže získať. Druhým, omnoho dôležitejším zistením však je, že automatický bot pridávajúci náhodný obsah na Twitter každú minútu bol schopný získať vyššie influence-skóre, ako bežný užívateľ sociálnej siete (50 bodov).

#### 3.1.2.1 Vplyv alebo veľkosť publika?

*„Influenced by David Ogilvy, the man many call The Father of Advertising, we believe that brand amplification does not provide the measurement and proof that a business’s executive demands, nor the results clients expect. Ogilvy built his reputation and success on the principle that the function of advertising is to sell and that successful advertising for any product is based on information about its consumer. The biggest problem with current influence measurement is that several of the better-known platforms purely measure influence by amplification and popularity metrics; if someone is active enough online and has a large following, then that person is influential. The flaw in this approach is it ignores key factors such as relevance, context, trust, and how a message resonates with that “influencer’s” audience“[14, lokácia 2778].*

Na základe vyššie uvedenej myšlienky niektoré súčasné riešenia (prevažne tie otvorené verejnosti) neodrážajú skutočný vplyv užívateľa na ostatných užívateľov sociálnej siete, ale z veľkej časti sa spoliehajú na počet followerov a frekvenciu generovania obsahu ako na relevantnú metriku bez nejakej hlbšej analýzy kontextu vzťahov a obsahu pridávaných správ. Je teda nanajvýš pochybné sa domnievať, že sú schopné identifikovať vhodných ľudí pre influence-marketingové kampane a následné zvýšenie predajov inzerovaných produktov

alebo služieb.

## 3.2 Novovznikajúce problémy

S rastom odvetvia influence–marketingu a teda príchodom nových platforiem pre vyhodnocovanie skóre užívateľov sociálnych sietí vznikajú nové výzvy, ktorým musia takéto platformy čeliť.

### 3.2.1 Optimalizovanie profilu

Influence Engine Optimization - optimalizovanie profilu na sociálnej sieti je technika, ktorou je možné zvýšiť svoje influence–skóre. Táto optimalizácia funguje na princípe sledovania a vyhodnocovania praktík, ktoré prispievajú k zvyšovaniu skóre a následné využívanie uvedených znalostí k jeho umelému zvýšeniu. Táto technika umelo generuje influence a hlavne *potláča prirodzené chovanie na sociálnej sieti*. Jedná sa teda o nežiadúci jav a služby vyhodnocujúce influence–skóre by mali byť schopné túto optimalizáciu profilu detekovať.

Paralelu s touto technikou je možné nájsť pri optimalizovaní webových stránok za účelom lepšieho miesta vo vyhľadávaní pri behu algoritmu Google Page Rank (SEO).

## 3.3 Skutočné faktory pôsobiace na vplyv užívateľa

Pri identifikovaní správnych užívateľov, ktorí majú vplyv na ostatných užívateľov sociálnej siete je potrebné uvážiť fakt, či je kladený dôraz na *zvýšenie povedomia o inzerovanom produkte*, alebo *zvýšenie jeho predaja*. Zo záverov z knihy *Influence marketing: how to create, manage, and measure brand influencers in social media marketing* [14] vyplýva, že micro–influenceri, teda užívatelia, ktorí majú menšie publikum (a tým pádom ho lepšie poznajú a majú viac času sa mu venovať) omnoho viac vplývajú na rozhodovanie svojich followerov ako takzvaní macro–influenceri, teda užívatelia, ktorí majú rádovo väčšie publikum ako micro–influenceri, no väzby medzi týmto publikom a macro–influencerom sú omnoho menšie než pri micro–influenceri.

Ktoré faktory teda vplývajú na to, či je užívateľ potenciálne prospešný pre predaj reklamy na sociálnych sieťach? Medzi tie najhlavnejšie patrí:

- Kontext a sentiment generovaného obsahu.
- Kvalita vzťahov medzi užívateľmi sociálnej siete.
- Veľkosť publika a kvalita followerov užívateľa.
- Úroveň, s akou followeri reagujú na túto aktivitu influencera odráža jeho vplyv, nie samotná aktivita.

#### 3.3.1 Ďalšie požiadavky kladené na proces

Mimo analýzy vyššie spomenutých atribútov užívateľov je potrebné brať do úvahy aj iné faktory vedúce k spoľahlivému identifikovaniu vhodných kandidátov pre influence–marketing. Medzi ne patria napríklad:

- Identifikovanie, v akých oblastiach je užívateľ vplyvný a do akej miery.
- Označenie nevalidných účtov - falošné alebo neaktívne účty.
- Identifikácia typu užívateľa - micro alebo macro–influencer.
- Odolnosť voči optimalizácii profilu za účelom zvýšenia influence–skóre.

#### 3.4 Proces výpočtu

Proces výpočtu influence–skóre je možné rozčleniť do troch častí v závislosti na tom, čo sa deje s dátami o užívateľovi a jeho konverzáciách.

1. *Vytváranie fronty dotazov* na server, ktoré majú za úlohu stiahnuť dáta o užívateľoch, ktoré sú potrebné pre ďalšiu analýzu. Jedná sa o ich konverzácie, preferencie a bio (základné informácie o užívateľoch).
2. Vhodné uloženie nespracovaných dát zo sociálnych sietí (teda všetko, čo sa o profile užívateľa zo sociálnej siete dá zistiť) pre ďalšie spracovanie.
3. *Spracovanie dát* z kroku 2:
  - Vyhodnotenie tagov užívateľa (kľúčové slová, ktoré sú extrahované z užívateľových konverzácií, preferencií alebo priamo jeho profilu).
  - Identifikovanie neaktívnych alebo falošných účtov.
  - Vyhodnotenie jednotlivých konverzácií užívateľa.
  - Vyhodnotenie skóre pre jednotlivé tagy užívateľa a určenie typu užívateľa (micro, macro).
  - Uloženie zistených informácií v takom formáte, aby k nim bol efektívny prístup na základe dotazu od koncového užívateľa systému.

#### 3.5 Stiahnutie dát zo sociálnych sietí

Dáta o jednotlivých užívateľoch a ich aktivitách sa zo sociálnych sietí sťahujú pomocou ich verejne prístupných rozhraní. Facebook a Twitter používajú rozhranie typu REST, pomocou ktorého je možné vytvárať, zobrazovať, editovať a prípadne aj mazať dáta na sociálnych sieťach. Pre navrhované riešenie je podstatné mať čo najrozsiahlejší prístup k dátam o užívateľoch.

### 3.5.1 Obmedzenia

Facebook a Twitter vyžadujú autorizované API dotazy. V skratke to znamená, že pre prístup k dátam sa využívajú pri každom dotaze rôzne API kľúče (súkromné aj verejné), ktoré jednoznačne identifikujú aplikáciu, ktorá o tieto dáta žiada, jej práva a limity pre odosielanie dotazov.

Facebook je v tejto oblasti striktnejší a nepovolí aplikáciám tretích strán stiahnuť takmer žiadne dáta o užívateľoch, ktorí používanie tejto externej aplikácie explicitne nepovolili. Twitter využíva autorizáciu skôr na obmedzenie počtu dotazov na svoje servery a neuzatvára základné informácie o užívateľoch pod autorizačné kľúče, no prístup k citlivejším dátam (napríklad súkromné správy) musí už byť povolený užívateľom (rovnako ako v prípade Facebooku, potvrdením o používaní externej aplikácie).

Z vyššie uvedeného dôvodu a tiež pre možnosť implementovať časť navrhovaných procesov bol zvolený ako jediný zdroj dát Twitter. Navrhované riešenie však počíta aj s pripojením k sociálnej sieti Facebook.

Je tiež potrebné podotknúť, že podľa platnej legislatívy nie je možné bez súhlasu užívateľa sociálnej siete získavať a spracovávať jeho osobné údaje, aj keď sa jedná o údaje z jeho verejne prístupného profilu. Z tohoto dôvodu je pre finálnu verziu aplikácie potrebné získavať dáta o jednotlivých užívateľoch iba s ich explicitným súhlasom (užívateľ potvrdí prístup k jeho osobným údajom aplikáciám tretích strán). Druhým (vhodnejším) riešením je presunúť poskytovanie navrhovanej služby mimo platnosť európskej legislatívy, kde nenastáva problém so získavaním a spracovaním osobných údajov (napríklad USA).

## 3.6 Uloženie dát

Aplikácia vzhľadom na bežne zavedené limity API dotazov na servery sociálnych sietí potrebuje efektívne ukladať niekoľko druhov dát. Typ úložiska vybraného pre rôzne fázy behu procesu sa teda líši hlavne v závislosti od štruktúry dát a ich vzájomnými vzťahmi.

### Fronta dotazov

Prvou fázou behu procesu je sťahovanie dát zo sociálnych sietí. Tieto dáta sa však kvôli API limitom nemôžu sťahovať naraz, preto je nutné vytvárať fronty úloh, ktoré budú tieto dáta sťahovať, akonáhle sa uvoľní API limit. Nie je nutné dbať na veľkosť úložiska, nakoľko dáta sa z tohto úložiska po spracovaní mažú. Na túto úlohu je teda vhodná obyčajná bežne používaná relačná databáza, napríklad MySQL alebo NoSQL key–value riešenie ako napríklad Redis [16].

### Ukladanie nespracovaných dát

Dáta stiahnuté zo sociálnych sietí poskytujú zvyčajne oveľa viac informácií, než je potrebné pre výpočet influence–skóre užívateľa. Tieto nadbytočné dáta

sú však pomerne cenné, nakoľko sa v budúcnosti môže zmeniť politika spoločnosti stojacej za danou sociálnou sieťou a nebude tieto dáta poskytovať vôbec alebo iba v obmedzenom množstve (napríklad pod špeciálnymi autorizačnými tokenmi, ako to v nedávnej minulosti spravili Facebook a Twitter). Do úvahy tiež treba brať fakt, že výpočet skóre užívateľa sa môže v čase meniť a bolo by neefektívne tieto dáta sťahovať znovu. Keďže sú klasifikované ako pomerne cenné, je nutné ich dlhodobo štruktúrovane ukladať. Pre takéto ukladanie je vhodné zvoliť škálovateľnú databázu vytvorenú pre ukladanie veľkého počtu dát, napríklad Apache Cassandra [17].

#### Ukladanie spracovaných dát

Spracované dáta musia byť ľahko dostupné pre vyhľadávanie (dotazy od používateľa služby) a tiež jednoducho updatovateľné. Do úvahy je potrebné brať aj fakt, že medzi dátami existujú rôzne vzťahy (napríklad užívateľ  $x$  je priateľom užívateľa  $y$ , užívateľ  $x$  komentoval príspevok užívateľa  $y$  a podobne). Preto bola na základe knihy *Graph Databases* [18] zvolená ako vhodná databáza s grafovou štruktúrou, konkrétne Neo4J [19].

## 3.7 Analýza dát

Pre každého užívateľa sociálnej siete sú *identifikované kľúčové slová*, ktoré charakterizujú jeho aktivitu na sociálnej sieti. Kľúčové slová sú extrahované z jeho verejného profilu a verejných konverzácií, ktoré vedie s ostatnými užívateľmi. Pre každé kľúčové slovo je následne vypočítané číslo, ktoré popisuje percentuálne zastúpenie tohto slova (tagu) v aktivite užívateľa.

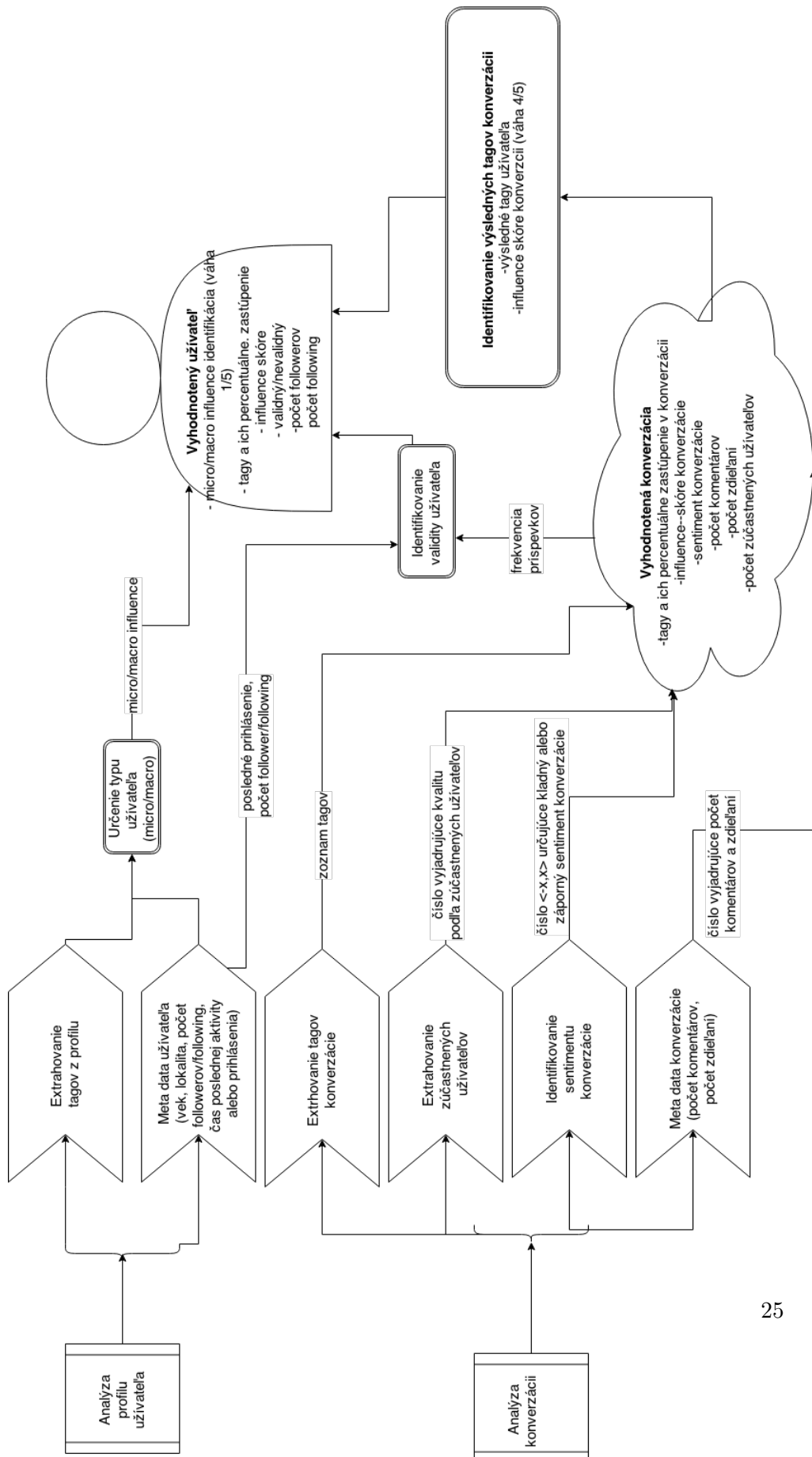
Následne sa na základe informácií o aktivite užívateľa a počtu jeho nasledovníkov dopočíta číslo, ktoré vyjadruje pomer v akom užívateľ inklinuje skôr ku micro alebo macro influencingu. Toto číslo je ďalej označované ako micro/macro influencer váha. Po tejto analýze tiež nasleduje validovanie užívateľa, na základe ktorého sa identifikujú falošné alebo neaktívne účty. Celý proces analyzovania dát zo sociálnej siete je znázornený na obrázku 3.2.

### 3.7.1 Analyzovanie užívateľa

#### 3.7.1.1 Extrahovanie tagov z profilu užívateľa

Pre analýzu kľúčových slov užívateľa sú relevantné informácie z jeho profilu, ktoré o sebe uvádza on sám a jeho konverzácie, ktoré vedie s ostatnými užívateľmi. Kľúčové slová z profilu je možné získať priamo z tagov (slová začínajúce na #), ktoré o sebe užívateľ uviedol vo svojom profile sám. Ak takéto slová nie sú k dispozícii, na profilovom texte sa prevedie lematizácia a identifikovanie slovných druhov (tzv. POS - Part of speech tagging). Za kľúčové slová sa pova-





Obr. 3.2: Zjednodušený globálny pohľad na proces analyzovania dát

žuje prvých päť podstatných mien v nominatíve. Ak užívateľ o sebe neuviedol žiadne informácie v profile, tak sa tagy doplnia z analýzy konverzácií.

#### 3.7.1.2 Určenie typu užívateľa (micro/macro influencer váha)

Celkový počet užívateľov, ktorí sledujú aktivitu aktuálne analyzovaného užívateľa logicky tiež vplýva na vypočítané skóre. Do úvahy je však potrebné brať iba validné účty (teda nie falošné a neaktívne účty).

#### 3.7.1.3 Vyhodnotený užívateľ

Pod pojmom vyhodnotený užívateľ sa rozumie užívateľ, ktorý má záznam vo finálnom úložisku (grafovej databáze). Podľa úrovne spracovania môže vyhodnotený užívateľ obsahovať influence-skóre, tagy, informáciu o validite a ďalšie atribúty (počet followerov, following, posledná aktivita). Jediný povinný atribút je meno na sociálnej sieti a identifikátor v rámci aplikácie.

### 3.7.2 Analyzovanie konverzácií užívateľa

Pod pojmom konverzácia sa rozumie interakcia medzi viacerými užívateľmi sociálnej siete. Konverzácia v kontexte tejto práce je skupina komentárov vytvorená pod príspevkom užívateľa. Do úvahy pre ďalšiu analýzu sa berú iba konverzácie a komentáre vzniknuté validnými užívateľmi. V texte sa ďalej bude k takejto konverzácii referovať ako o „užívateľovej konverzácii“.

#### 3.7.2.1 Extrahovanie tagov konverzácie

Z každej konverzácie sa vezmú tagy poskytnuté priamo užívateľmi (začínajúce na #). Ak je takýchto tagov menej ako 5, tak sa na zvyšku slov prevedie lematizácia a identifikovanie slovných druhov. Za kľúčové slová sa považuje prvých päť slov, ktorých počet výskytov je väčší ako aritmetický priemer výskytov všetkých slov.

#### 3.7.2.2 Extrahovanie influence-skóre konverzácie

Kvalita užívateľov zúčastnených na konverzácii prispieva významou mierou k identifikovaniu vplyvu užívateľa. Čím kvalitnejší ostatní užívatelia sa zúčastňujú na jeho konverzáciách, tým je daný užívateľ váženejší vo svojej komunite a tým si zasluhuje mať vyššie influence-skóre. Výhodou tejto analýzy je tiež fakt, že informáciu o tom, kto pridal komentár k príspevku nie je potrebné pracne extrahovať z textu, ale sociálne siete poskytujú priamo identifikátor užívateľa, ktorý takúto akciu vykonal. Influence-skóre konverzácie sa určí ako stredná hodnota influence-skóre užívateľov zúčastnených na konverzácii.

### 3.7.2.3 Identifikovanie sentimentu konverzácie

Sentiment konverzácii a príspevkov žiadnym spôsobom neovplyvňuje influence-skóre užívateľa, je ním však možné detekovať prípadné negatívne postoje k určitému tagu. Táto informácia je užitočná pre užívateľa navrhovaného riešenia, keďže je takto možné jednoducho identifikovať nespokojných (alebo naopak veľmi spokojných) zákazníkov určitej služby alebo produktu. Sentiment je číslo na stupnici  $\langle -x, x \rangle$ , na získanie ktorého sa použije buď externá webová služba, alebo knižnica s inicializačnými dátami.

### 3.7.2.4 Meta data konverzácie

Je potrebné brať do úvahy meta dáta o konverzácii

- počet zúčastnených užívateľov,
- celkový počet komentárov,
- počet zdieľaní príspevku.

Informácie o celkovom počte komentárov a zúčastnených užívateľov sú využité k spresneniu identifikácie užívateľa ako micro-influencera, na základe nasledujúcej úvahy:

Užívateľ, ktorý pravidelne vzbudzuje stredne veľké diskusie (diskusie obsahujúce 10–50 príspevkov), v ktorých sa často angažuje priamo on sám (vhodná proporcia vlastných odpovedí vzhľadom k príspevkom od ostatných užívateľov) je vhodným kandidátom na micro-influencera.

### 3.7.2.5 Vyhodnotená konverzácia užívateľa

Pod pojmom vyhodnotená konverzácia sa rozumie konverzácia, na ktorej prebehli všetky potrebné analýzy (tagov, sentimentu, influence-skóre konverzácie).

### 3.7.2.6 Vyhodnotenie výsledných tagov všetkých konverzácií užívateľa

Zo všetkých konverzácií vzniknutých pod užívateľovými príspevkami sa získajú výsledné informácie o užívateľovi, ktoré sú doplnené do jeho analyzovaného profilu.

- Tagy užívateľa - zo všetkých konverzácií sa vezmú všetky tagy, zoradia sa podľa počtu ich výskytov a prvých 5 takto získaných tagov sa pridá k tagom užívateľa.

- Influence-skóre - 4/5 influence-skóre užívateľa tvorí analýza jeho konverzácií. Preto sa z každej užívateľovej konverzácie vezme jej celkové influence-skóre a stredná hodnota tohto skóre sa pridá k celkovému influence-skóre užívateľa s váhou 4/5.

Fakt, že väčšina influence-skóre užívateľa vznikne za základe jeho konverzácií, pričom influence-skóre konverzácie vznikne na základe influence-skóre zúčastnených užívateľov značnou mierou obmedzuje možnosť ovplyvňovať výsledky analýzy, ako bolo spomenuté v 3.1.1.

#### 3.7.3 Ďalšie požiadavky na proces identifikovania influence-skóre

##### 3.7.3.1 Vyradenie falošných účtov

Jedná sa o účty, za ktoré publikuje nové správy umelo vytvorený program (script). Identifikovanie takýchto účtov môže byť vymedzené na základe jednoduchých pravidiel:

- *Pomer aktivity generovanej užívateľom a reakcie na túto aktivitu od ostatných užívateľov* - užívateľ, ktorý veľmi často prispieva na sociálnu sieť, ale iba vo výnimočných prípadoch dostáva nejaký odozvu od ostatných užívateľov je buď robot, ktorý iba generuje príspevky alebo je to reálny užívateľ, ktorý aj tak nie je relevantný pre ďalšie spracovanie.
- *Nepriaznivý pomer following/followers* - užívateľ má veľkú disproporciu medzi počtom sledovaných ľudí a ľudí, ktorí ho sledujú. Vyplývajúci efekt je popísaný v predchádzajúcom bode.
- *Periodické pridávanie príspevkov* - vzniknú len v prípade strojovo generovaných príspevkov. V niektorých špeciálnych prípadoch (spravodajské služby a podobne) sa nejedná o nechcený jav, no tieto účty nie sú aj tak vhodné pre ďalšiu analýzu za účelom predaja.

##### 3.7.3.2 Vyradenie neaktívnych účtov

Jedná sa o účty, ktoré už nevykazujú žiadnu aktivitu. Takéto dáta sú následne irelevantné pre ďalšiu analýzu. Je ich možné identifikovať na základe poslednej zverejnenej správy alebo posledného prihlásenia do danej sociálnej siete. Za neaktívneho je považovaný užívateľ, ktorého posledná aktivita bola vykonaná pred viac ako troma mesiacmi.

## 3.8 Prípady použitia

K detailnejšiemu popisu chovania navrhovaného systému sú použité prípady použitia (anglicky *use case*). V princípe sa jedná o analýzu funkcií, ktoré sys-

tém pre analýzu influence-skóre musí minimálne obsahovať. Zahŕňajú aktérov a ich záujmy, predpoklady, rozsah prípadu použitia a scenáre s prípadnými výnimkami. Nižšie uvedené prípady použitia tvoria základ pre implementačnú časť práce. Štruktúra prípadov použitia je inšpirovaná knihou *Writing effective use cases* [20] a je vhodná pre grafické zobrazenie v UML notácii [21].

### 3.8.1 Užívateľ

#### 3.8.1.1 Sťahovanie dát o užívateľoch zo sociálnej siete

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“)

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

#### Zúčastnené strany a záujmy:

- Systém - úspešne stiahnuť dáta o užívateľovi sociálnej siete.
- API rozhranie sociálnej siete (ďalej len API) - poskytnúť dáta, nie však nad povolený limit.
- Úložisko reprezentujúce frontu dotazov - ďalej len DBQ.
- Úložisko pre ukladanie nespracovaných dát - ďalej len DBL.

#### Predpoklady

- Systém je schopný autorizovať API dotazy.
- Neprekročený limit API dotazov pre sťahovanie dát o užívateľovi.

**Minimálne plnenie:** Systém získa profil užívateľa s údajmi uvedenými nižšie označenými ako  $(p)$ .

**Úspešné plnenie:** Stiahnutie špecifických dát o užívateľovi zo sociálnej siete:

- meno  $(p)$ ,
- vek,
- pohlavie,
- lokalita,
- bio,

### 3. ANALÝZA A NÁVRH

---

- počet following (koho sleduje) a followerov (kto sleduje užívateľa) ( $p$ ),
- čas poslednej aktivity (prihlásenie alebo pridanie príspevku) ( $p$ ).

**Spúšťač** : Vyzdvihnutie dotazu z fronty dotazov (DBQ)

#### **Scenár úspešného plnenia:**

1. Systém vyzdvihne dotaz pre stiahnutie dát o užívateľovi z DBQ.
2. Systém overí neprekročenie API limitu pre sťahovanie dát o užívateľovi.
3. Systém stiahne dáta o užívateľovi.
4. Systém uloží dáta do DBL.
5. Systém vymaže požiadavku pre sťahovanie užívateľa z DBQ.
6. Systém pokračuje v spracovávaní ďalšieho požiadavku.

#### **Výnimky**

- **2 Prekročený limit pre API dotazy** - systém sa pokúsi nájsť autentifikačný token s voľným API limitom, v prípade neúspechu sa uspí na dobu potrebnú pre obnovenie API limitu.
- **3 Neúspešné stiahnutie dát** - môže nastať v prípade:
  - nedostupnosti služby,
  - neúspešnej autorizácie dotazu na server,
  - prekročení API limitu.

V tomto prípade sa sťahovanie dát o užívateľovi ukončí, no fronta dotazov na server sa nemodifikuje, takže tento dotaz sa vykoná neskôr.

**Frekvencia použitia** bez obmedzenia (v závislosti od API limitu a počtu autentifikačných tokenov).

#### **3.8.1.2 Identifikovanie micro/macro influencer váhy**

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

**Zúčastnené strany a záujmy:**

- Systém - rozhodnúť o micro/macro influencer váhe.
- Úložisko pre ukladanie nespracovaných dát- ďalej len databáza DBL.
- Úložisko reprezentujúce spracované dáta - DBG.

**Predpoklady**

- Stiahnuté dáta o užívateľovi.
  - Počet nasledovníkov (followerov),
  - Počet nasledujúcich (following).
- Užívateľ nie je klasifikovaný ako nevalidný.

**Úspešné plnenie:** Identifikovanie micro/macro influencer váhy.

**Scenár úspešného plnenia:**

1. Systém získa dáta o užívateľovi z DBG.
  - Alternatívny 1 - Systém získa dáta o užívateľovi z DBL.
2. Systém overí či je užívateľ validný.
3. Systém vypočíta micro/macro influencer podľa nasledujúceho vzorca:

$$\begin{aligned}
 flw &= \text{počet followerov užívateľa} \\
 macro\_weight &= \log_{5000} flw \\
 &\quad if(flw < 5000) \\
 micro\_weight &= \frac{flw}{5000} \\
 &\quad else \\
 micro\_weight &= \frac{1}{\log_{5000} flw}
 \end{aligned} \tag{3.1}$$

4. Systém aktualizuje micro/macro influencer váhy užívateľa v DBG.

**Alternatívny scenár**

1. **Užívateľ nie je vyhodnotený**
  - Systém získa dáta o užívateľovi z DBL.
  - Systém ďalej neoveruje validitu užívateľa.

#### Výnimky

- **2 Užívateľ nie je validný** - Podľa prípadu použitia Identifikovanie nevalidných užívateľov (3.8.1.5) systém identifikoval užívateľa ako nevalidného. Spracovanie sa preruší.

#### 3.8.1.3 Extrahovanie tagov z profilu

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

#### Zúčastnené strany a záujmy:

- Systém - získať tagy definujúce užívateľa.
- Úložisko pre ukladanie nespracovaných dát- ďalej len DBL.
- Úložisko reprezentujúce spracované dáta - ďalej len DBG.

#### Predpoklady

- Stiahnuté dáta o užívateľovi.
  - Bio užívateľa (profilové údaje, ktoré o sebe užívateľ uviedol sám).

**Minimálne plnenie:** Systém extrahuje aspoň 1 tag.

**Úspešné plnenie:** Extrahovanie tagov z bio užívateľa.

#### Scenár úspešného plnenia:

1. Systém získa dáta o užívateľovi z DBL.
2. Systém extrahuje všetky slová začínajúce na #. Tieto slová predstavujú tagy užívateľa.
3. Systém aktualizuje tagy užívateľa do DBG.

#### Výnimky

- **2 Extrahovaný menej ako 1 tag** - Systém lematizuje bio užívateľa a identifikuje všetky podstatné mená v prvom páde ako tagy.



#### **3.8.1.4 Identifikácia tagov užívateľa na základe jeho konverzácií a profilu**

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

##### **Zúčastnené strany a záujmy:**

- Systém - určenie výsledných tagov užívateľa.
- Úložisko pre ukladanie nespracovaných dát- ďalej len DBL.
- Úložisko reprezentujúce spracované dáta - ďalej len DBG.

##### **Predpoklady**

- Spracované konverzácie užívateľa v DBG.
- Spracované dáta o užívateľovi v DBG.

**Úspešné plnenie:** Identifikovanie výsledných tagov užívateľa.

##### **Scenár úspešného plnenia:**

1. Systém nahrá dáta o konverzáciách a užívateľovi do pamäte.
2. Z každej konverzácie vyberie všetky tagy identifikované v 3.8.2.3 a zoradí ich podľa počtu. Prvých 5, ktoré majú počet výskytu väčší ako aritmetický priemer všetkých výskytov sú identifikované ako tagy užívateľa.
3. Tagy užívateľa identifikované v 3.8.1.4 sú porovnané s množinou v predchádzajúcom kroku a tie tagy užívateľa, ktoré sa vyskytujú v tejto množine sú tiež klasifikované ako tagy užívateľa.

#### **3.8.1.5 Identifikovanie nevalidných užívateľov**

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

#### Zúčastnené strany a záujmy:

- Systém - identifikovať nevalidné účty.
- Úložisko pre ukladanie nespracovaných dát - ďalej len DBL.
- Úložisko reprezentujúce spracované dáta - ďalej len DBG.

#### Predpoklady

- Stiahnuté dáta o užívateľovi.
  - Čas poslednej aktivity užívateľa,
  - Počet nasledovníkov (followerov),
  - Počet nasledujúcich (following).
- Stiahnuté dáta o príspevkoch užívateľa.

**Úspešné plnenie:** Systém verifikuje validitu užívateľa.

#### Scenár úspešného plnenia:

1. Systém získa dáta o užívateľovi z DBL.
2. Systém získa dáta o užívateľových príspevkoch a ich konverzáciách z DBL.
3. Systém vyhodnotí nasledujúce kritériá:
  - Pomer aktivity generovanej užívateľom a reakcie na túto aktivitu od ostatných užívateľov - určí sa podľa vzorca:

$$\begin{aligned} prisp &= \text{Počet príspevkov užívateľa} \\ akt &= \text{Aktivita generovaná ostatnými užívateľmi} \\ boolean\ valid &= \frac{prisp}{akt} > 2 \end{aligned} \quad (3.2)$$

Aktivita generovaná ostatnými užívateľmi je určená a má váhu:

- Páči sa: 0.5,
- Zdieľanie: 0.75,
- Komentár: 1.
- Nepriaznivý pomer followerov/following - určí sa podľa nasledujúceho vzorca:

$$boolean\ valid = \frac{following}{followers} > 10$$

- užívateľ sleduje 10x viac užívateľov ako sledujú jeho.

- Periodické pridávanie príspevkov - rozdiel medzi pridaním  $x$ -tého príspevku a  $x+1$ -tého príspevku je rovnaký (až na odchýlku  $< 10$  sekúnd) ako rozdiel medzi pridaním príspevku  $x+1$  a príspevku  $x+2$ . V takomto prípade je užívateľ označený za nevalidného.
- Vyradenie neaktívnych účtov - čas poslednej aktivity (login alebo pridanie príspevku) je starší ako 3 mesiace.

4. Systém aktualizuje boolean informáciu o užívateľovej validite do DBG.

### Výnimky

- **2 Extrahovaný menej ako 1 tag** - Systém lematizuje bio užívateľa a identifikuje všetky podstatné mená v prvom páde ako tagy.

## 3.8.2 Konverzácie užívateľa

### 3.8.2.1 Stiahnutie dát o príspevkoch na sociálnej sieti

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

#### Zúčastnené strany a záujmy:

- Systém - získať príspevky vytvorené užívateľom.
- Úložisko pre ukladanie nespracovaných dát- ďalej len DBQ.
- Úložisko reprezentujúce spracované dáta - ďalej len DBL.

#### Predpoklady

- Systém je schopný autorizovať API dotazy.
- Neprekročený limit pre API dotazy pre sťahovanie dát o užívateľovi.

**Úspešné plnenie:** Systém získa dáta o príspevku na sociálnej sieti v rozsahu:

- čas príspevku,
- text príspevku,
- počet páči sa a zdieľaní.

**Spúšťač** : Vyzdvihnutie dotazu z fronty dotazov.

**Scenár úspešného plnenia:**

1. Systém vyzdvihne dotaz pre stiahnutie dát o príspevku z DBQ.
2. Systém overí neprekročenie API limitu pre sťahovanie dát o príspevkoch.
3. Systém stiahne dáta o príspevku.
4. Systém uloží dáta do DBL.
5. Systém vymaže požiadavku pre sťahovanie príspevku s DBQ.
6. Systém pokračuje v spracovávaní ďalšieho požiadavku.

**Výnimky**

- **2 Prekročený limit pre API dotazy** - systém sa pokúsi nájsť autentifikačný token s voľným API limitom, v prípade neúspechu sa uspí na dobu potrebnú pre obnovenie API limitu.
- **3 Neúspešné stiahnutie dát** - môže nastať v prípade:
  - nedostupnosti služby,
  - neúspešnej autorizácie dotazu na server,
  - prekročení API limitu.

V tomto prípade sa sťahovanie dát o príspevku ukončí, no fronta dotazov na server sa nemodifikuje, takže tento dotaz sa vykoná neskôr.

#### 3.8.2.2 Stiahnutie dát o konverzáciách na sociálnej sieti

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“).

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

**Zúčastnené strany a záujmy:**

- Systém - získať dáta o komentároch viazaných na príspevky užívateľa.
- Úložisko pre ukladanie nespracovaných dát - ďalej len DBQ.
- Úložisko reprezentujúce spracované dáta - ďalej len DBL.

### Predpoklady

- Systém je schopný autorizovať API dotazy.
- Neprekročený limit pre API dotazy pre sťahovanie dát o komentároch.

**Úspešné plnenie:** Systém získa dáta o všetkých komentároch k danému príspevku, a to v rozsahu:

- čas komentára príspevku,
- autor komentára,
- text komentára.

**Spúšťač** : Vyzdvihnutie dotazu z fronty dotazov

### Scenár úspešného plnenia:

1. Systém vyzdvihne dotaz pre stiahnutie dát o komentároch pre príspevok z DBQ.
2. Systém overí neprekročenie API limitu pre sťahovanie komentárov o príspevkoch.
3. Systém stiahne komentáre príspevku.
4. Systém uloží dáta o komentároch do DBL.
5. Systém vymaže požiadavku pre sťahovanie komentárov z DBQ.
6. Systém pokračuje v spracovávaní ďalšieho požiadavku.

### Výnimky

- **2 Prekročený limit pre API dotazy** - systém sa pokúsi nájsť autentifikačný token s voľným API limitom, v prípade neúspechu sa uspí na dobu potrebnú pre obnovenie API limitu.
- **3 Neúspešné stiahnutie dát** - môže nastať v prípade:
  - nedostupnosti služby,
  - neúspešnej autorizácie dotazu na server,
  - prekročení API limitu.

V tomto prípade sa sťahovanie dát o komentároch ukončí, no fronta dotazov na server sa nemodifikuje, takže tento dotaz sa vykoná neskôr.

#### 3.8.2.3 Extrahovanie tagov a sentimentu z konverzácie

**Hlavný účastník:** Navrhovaný systém (ďalej len „Systém“)

**Rozsah:** Komponent

**Úroveň:** Podfunkcia

**Zúčastnené strany a záujmy:**

- Systém - identifikovanie kontextu konverzácie.
- Úložisko pre ukladanie nespracovaných dát - ďalej len DBL.
- Úložisko reprezentujúce spracované dáta - ďalej len DBG.

**Predpoklady**

- Konverzácia je uložená v DBL.

**Úspešné plnenie:** Systém extrahuje minimálne 5 tagov o konverzácii.

**Scenár úspešného plnenia:**

1. Systém nahrá príspevok a jeho komentáre z DBL do pamäte.
2. Systém extrahuje tagy označené užívateľmi (začínajúce na #) a priradí ich k tagom konverzácie.
3. Systém lematizuje texty konverzácie a zoradí všetky podstatné mená prvom páde podľa počtu výskytov. Prvých päť takýchto slov, ktorých výskyt je väčší ako aritmetický priemer výskytov všetkých slov je identifikovaných ako tagy.
4. Systém získa z textov konverzácie (použitie externej knižnice) sentiment konverzácie a upraví ho na stupnicu <-4,4>.
5. Systém uloží extrahované tagy a sentiment konverzácie do DBG k spracovanej konverzácii.

#### 3.8.3 Užívateľ aplikácie

##### 3.8.3.1 Definovanie parametrov vyhľadávania

**Hlavný účastník:** Užívateľ vyhľadávacej aplikácie (zamestnanec alebo priamo zákazník spoločnosti, ďalej len „užívateľ“).

**Rozsah:** Analytický systém

**Úroveň:** Uživatelský cieľ

**Zúčastnené strany a záujmy:**

- Užívateľ - obdržať zoznam ľudí vyhovujúci vyhľadávaniu.
- Systém - vrátiť relevantné výsledky v čo najkratšom čase.

**Predpoklady** Systém musí mať dostatok dát pre analýzu, vstupné dáta sú validné.

**Minimálne plnenie:** Užívateľovi sa zobrazí jedna z nasledujúcich možností:

- zoznam vhodných kandidátov pre inzerovanie reklamy,
- informáciu, že nikto relevantný nebol nájdený,
- chybovú hlášku informujúcu, že nie je k dispozícii dostatok dát pre analýzu.

**Úspešné plnenie:** Systém vráti zoznam vhodných kandidátov pre inzerovanie reklamy.

**Spúšťač:** Užívateľ na obrazovke klikne na spustenie vyhľadávania

**Scenár úspešného plnenia:**

1. Užívateľ si otvorí v systéme vyhľadávanie a zadá potrebné dáta (v zátvorke uvedené obmedzenia pre validáciu dát):
  - pohlavie (muž/žena),
  - vekový rozsah (dve celé čísla od 0-150, druhé musí byť väčšie ako prvé),
  - lokalitu cieľovej skupiny (súradnice *lon,lat* a radius),
  - tagy identifikujúce cieľovú skupinu (slová oddelené čiarkou, maximum päť slov),
  - počet *n* hľadaných výsledkov (nezáporné číslo, maximálne 10 000).
2. Systém zvaliduje vstupné parametre vyhľadávania podľa obmedzení uvedených v predchádzajúcom kroku.

### 3. ANALÝZA A NÁVRH

---

3. Užívateľovi sa zobrazí informácia o začatí vyhľadávania: „Prebieha analýza požiadavky. Počet nájdených výsledkov  $x$ “. Nad týmto textom je zobrazený spinner,  $x$  určuje aktuálny počet nájdených výsledkov.
4. Systém začne vyhľadávanie a notifikuje užívateľa o postupe vyhľadávania (zmenou hodnoty premennej  $x$ ). Po nájdení požadovaného počtu výsledkov  $n$  alebo analyzovaní celého dátového úložiska systém zobrazí dáta o jednotlivých užívateľoch vo forme zoznamu.

#### Výnimky

- **1 Validácia dát zlyhá** - užívateľ je notifikovaný na obrazovke o nesprávne zadaných údajoch a pri každom nesprávne zadanom údají je uvedené, aký vstup sa očakáva na základe obmedzení v 1. bode hlavného scenára.
- **4 Žiadne relevantné výsledky nenájdené** - užívateľovi sa namiesto zoznamu nájdených užívateľov zobrazí v strede obrazovky správa „Žiadni relevantní užívatelia nenájdení. Prajete si redefinovať vyhľadávanie?“. Pod touto správou je vycentrované tlačítko, ktoré po kliknutí presmeruje na obrazovku so zadávaním parametrov vyhľadávania.

**Frekvencia použitia** Menej ako 10 denne.

#### 3.8.3.2 Filtrovanie nájdených výsledkov

**Hlavný účastník:** Užívateľ vyhľadávacej aplikácie (zamestnanec alebo priamo zákazník spoločnosti, ďalej len užívateľ).

**Rozsah:** Analytický systém

**Úroveň:** Užívateľský cieľ

#### Zúčastnené strany a záujmy:

- Užívateľ - zúžiť výber z veľkého množstva výsledkov v čo najkratšom čase.
- Systém - filtrovať a zobrazíť filtrované výsledky v čo najkratšom čase.

**Predpoklady** Systém z analýzy vrátil aspoň 1 užívateľa, ktorý sa dá filtrovať. Formulár pre filtrovanie neobsahuje žiadne dáta.

**Úspešné plnenie:** Systém zobrazí užívateľov podľa vyplnených údajov vo filtri.



**Spúšťač:** Užívateľ na obrazovke klikne na spustenie filtrovania výsledkov.

**Scenár úspešného plnenia:**

1. Systém má zobrazené dáta o užívateľoch z analýzy vo forme zoznamu.
2. Systém zobrazí na obrazovke filtrovací formulár obsahujúci nasledovné položky (v zátvorke sú uvedené validačné pravidlá)
  - pohlavie (muž/žena),
  - vekový rozsah (dve celé čísla od  $x,y$ , kde  $x$  je vek najmladšieho užívateľa vráteneho z analýzy a  $y$  je vek najstaršieho užívateľa z analýzy.  $x \geq y$ ),
  - lokalitu cieľovej skupiny - súradnice  $lon,lat$  a radius (bez obmedzenia),
  - tagy (iba tagy užívateľov vrátených z analýzy).
3. Užívateľ zadá do formulára pre filtrovanie vstupné dáta a potvrdí kliknutím na spustenie filtrovania výsledkov.
4. Systém zvaliduje zadané dáta od užívateľa.
5. Systém prejde zoznam užívateľov v pamäti a zobrazí užívateľovi nový zoznam vyfiltrovaný podľa ním zadaných kritérií.

**Výnimky**

- **4 Validácia dát zlyhá** - užívateľ je notifikovaný na obrazovke o nesprávne zadaných údajoch a pri každom nesprávne zadanom údají je uvedené, aký vstup sa očakáva na základe obmedzení. v 4. bode hlavného scenára.
- **5 Žiadni užívatelia nevyhovujú filtru** - užívateľovi sa namiesto zoznamu nájdených užívateľov zobrazí v strede obrazovky správa „Žiadni užívatelia nevyhovujú filtru. Redefinujte parametre filtra alebo filter vymažte“. Pod touto správou je vycentrované tlačítko ktoré po kliknutí vyčistí všetky položky filtra a zobrazí všetkých užívateľov bez filtrovania.

**Frekvencia použitia** Menej ako 10 pre každé vyhľadávanie podľa 3.8.3.1.

## 3.9 Zhrnutie kapitoly

### Nedostatky

Na základe analýzy súčasných verejne dostupných platforiem boli identifikované nedostatky, ktoré tieto platformy obsahujú a sú nežiadúce pre navrhované riešenie. Boli taktiež popísané faktory, ktoré majú najväčší podiel na správnom vyhodnocovaní influence-skóre užívateľov sociálnych sietí. Na základe uvedených faktorov bol navrhnutý proces identifikácie skóre.

### Etapy procesu

Navrhované riešenie je na základe skúmania domény rozdelené do troch etáp: získanie dát zo sociálnych sietí, ich uloženie a následná analýza. Pre každú etapu je vhodné použiť iné dátové úložisko na základe toho, aké operácie budú následne s dátami vykonávané. Pre vytváranie fronty dotazov pre sťahovanie dát je odporúčaná obyčajná SQL databáza (v tomto prípade MySQL), pre ukladanie nespracovaných stiahnutých dát je vhodné použiť NoSQL logovaciu databázu (v tomto prípade Apache Cassandra) a na záver pre uloženie spracovaných údajov, efektívne vyhľadávanie, interpretáciu vzťahov medzi jednotlivými entitami a škálovateľnosť riešenia bola zvolená grafová databáza (v tomto prípade Neo4J).

### Entity a analytické operácie

Pre analýzu boli identifikované dva typy relevantných entít - užívateľ a konverzácie. Užívateľom sa rozumie profil na sociálnej sieti, pričom konverzácia je súbor komentárov pod príspevkom vytvoreným užívateľom. Nad týmito entitami sa prevedú rôzne analýzy (napríklad extrahovanie tagov, identifikácia micro/macro influencer), ktorých výsledkom je súbor tagov užívateľa s percentuálnou mierou určujúcou, v akej oblasti je užívateľ vplyvný a tiež jeho výsledné influence-skóre.

### Prípady použitia

Pre jednoznačnú implementáciu jednotlivých krokov analýzy boli vytvorené prípady použitia, kde je presne špecifikované, ako naprogramovať jednotlivé kroky analýzy influence-skóre užívateľa.

### Prínosy riešenia

Hlavným prínosom navrhovaného riešenia je presun výpočtu influence-skóre z aktivity generovanej užívateľom na aktivitu generovanú inými užívateľmi, čo má za následok zníženie náchylnosti navrhovaného riešenia k optimalizácii profilu analyzovaného užívateľa. Medzi ďalšie prínosy je možné zaradiť rozdelenie druhu užívateľa na základe veľkosti jeho publika a typu konverzácií, ktoré je schopný podnecovať a priradovanie váh ku jednotlivým tagom užívateľa.

---

# Implementácia

## 4.1 Technológie

Pre implementáciu navrhovaného riešenia bol vybraný programovací jazyk Java. Tento jazyk bol zvolený s ohľadom na jeho rozšírenosť a dostupnosť širokého spektra knižníc pre rôzne operácie vyžadované programom (sieťové volania, serializácia/deserializácia dát, databázové pripojenia a spracovanie prirodzeného jazyka).

## 4.2 Implementácia prípadov použitia

Pre praktickú demonštráciu navrhovaného riešenia bolo zvolené implementovať prípady použitia 3.8.1.1, 3.8.1.2 a 3.8.1.4. Implementačná časť teda pokrýva stiahnutie dát o užívateľovi, analýzu micro/macro influencer váhy a extrahovanie tagov z profilu užívateľa. Všetky tieto dáta sa následne ukladajú do vhodnej databázi podľa charakteru dát.

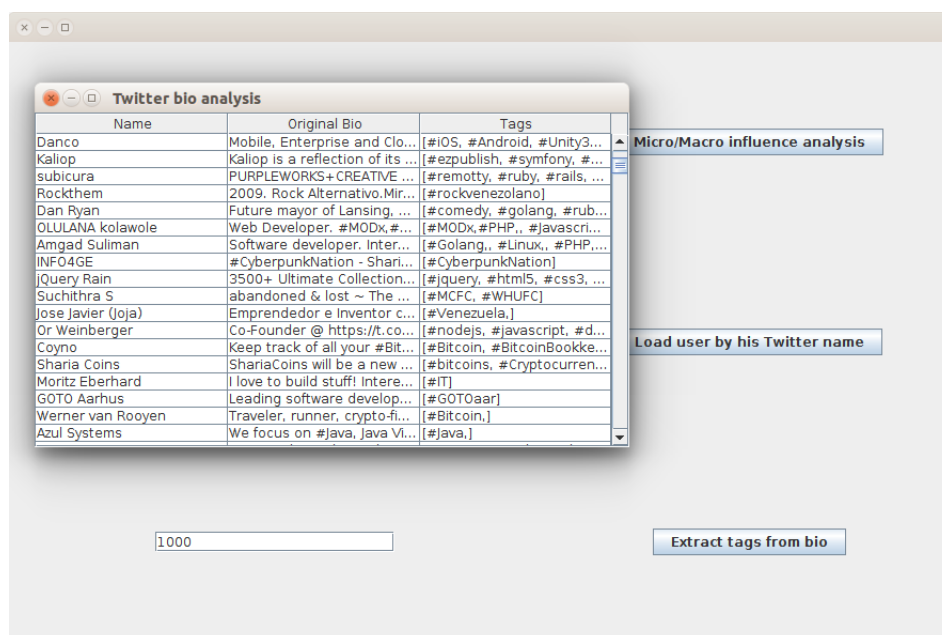
Demonštrovanie fungovania vyššie spomenutých prípadov použitia je umožnené pomocou jednoduchého rozhrania vytvoreného použitím základných Java Swing a AWT grafických komponentov. Toto rozhranie podporuje všetky tri vyššie spomenuté operácie, pričom dáta z micro/macro-influencer analýzy a extrahovania tagov zobrazí v prehľadnej tabuľke (obrázok 4.1).

## 4.3 Komponenty programu

Jednotlivé časti programu sú rozdelené do balíčkov (Java package) podľa spôsobu využitia ich tried. Pre tieto balíčky bola zvolená nasledovná štruktúra:

- database - komunikácia s databázami Neo4J, MySQL a Apache Cassandra,
- client - balíček slúžiaci pre komunikáciu so sociálnymi sieťami,

## 4. IMPLEMENTÁCIA



Obr. 4.1: Testovacie užívateľské prostredie

- ui - obsahuje užívateľské prostredie,
- analysis - triedy obsahujúce logiku spracovania dát.

Zloženie a kompiláciu programu má na starosti moderný nástroj pre automatizáciu vývoja softwaru *gradle*, ktorý stavia na svojich predchodcoch Apache Maven a Ant.

### 4.3.1 Použité knižnice

#### 4.3.1.1 Lematizér pre identifikovanie kľúčových slov

Pre lematizáciu a POS tagovanie bol využitý projekt Stanfordskej univerzity Stanford CoreNLP [22]. Táto knižnica je licencovaná pod licenciou GPL.

#### 4.3.1.2 Sťahovanie dát zo sociálnych sietí

Dotazy na servery sociálnych sietí sa musia autentifikovať, preto je potrebné vygenerovať na oboch sociálnych sieťach autentifikačné kľúče a tokeny, ktoré sa následne použijú pri inicializácii nižšie popísaných knižníc Twitter4J a Facebook4J.

### Twitter

Pre získanie dát zo sociálnej siete Twitter je použitá knižnica Twitter4J, ktorá je neoficiálnou implementáciou pre využívanie API rozhrania Twitteru

pre Javu. Poskytuje triedy pre všetky operácie, ktoré poskytuje Twitter API. Jej autorom je Yusuke Yamamoto a licencovaná je licenciou Apache License 2.0.

### Facebook

Získavanie dát zo siete Facebook je riešené pomocou knižnice Facebook4J, ktorej autorom je Ryuji Yamashita a je licencovaná pod Apache License 2.0.

#### 4.3.1.3 Databázové pripojenie

##### Fronta dotazov

Fronta dotazov je implementovaná formou MySQL databázy. Pre zjednodušenie obsluhy databázy je využitá knižnica ORM Lite, ktorá podporuje objektové mapovanie do databázových tabuliek a využíva návrhový vzor DAO pre prístup k dátam.

##### Logovanie dát

Pre ukladanie dát ostrej verzie programu je navrhnutá databáza Apache Cassandra, ktorá však v pilotnej verzii nie je implementovaná. Ukladanie nespracovaných dát zo sociálnych sietí sa stará databáza MySQL.

##### Ukladanie spracovaných údajov

Ukladanie spracovaných údajov je riešené formou grafovej databázy Neo4J, pre ktorú existuje integračná knižnica do Javy.

## 4.4 Zhrnutie kapitoly

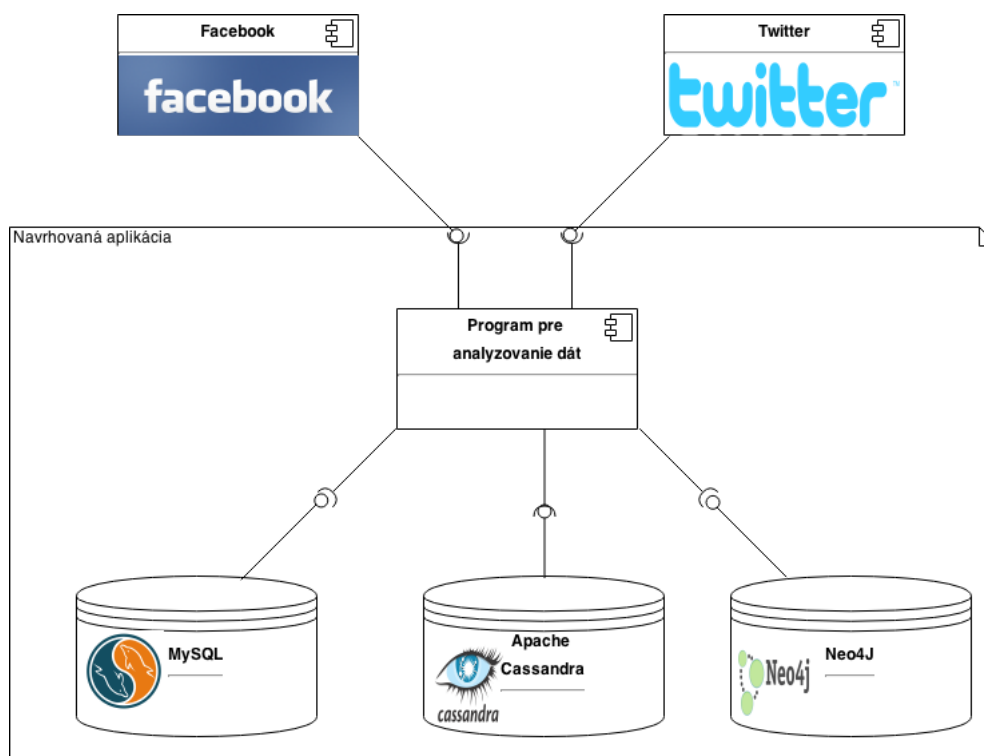
Pre praktickú ukážku pilotnej verzie navrhovaného riešenia bola zvolená implementácia troch prípadov použitia aplikácie týkajúce sa analyzovania profilu užívateľa sociálnej siete. Na splnenie tejto úlohy bolo potrebné vytvoriť projekt vo zvolenom vývojovom prostredí (IntelliJ Idea) a s pripojením k príslušným databázam (MySQL a Neo4J) a inicializovaním knižnice NLP pre proces lematizácie.

Po úspešnej inicializácii všetkých obslužných knižníc bolo implementované sťahovanie dát zo sociálnej siete Twitter. Pri tomto sťahovaní bol identifikovaný problém, kedy často došlo k prečerpaniu limitu pre sťahovanie dát. Tento problém bol odstránený pridaním ďalších dvoch autentifikačných tokenov a kontrolou prečerpania limitu. V prípade vyčerpania limitu na všetkých troch tokenoch sa proces sťahovania dát automatiky uspeje do vypršania limitu.

Pre zjednodušenie obsluhy troch spracovaných procesov bolo vytvorené užívateľské rozhranie s tabuľkovým prehľadom výsledkov uvedených procesov. Cez toto rozhranie je možné pohodlne sťahovať dáta o užívateľovi sociálnej siete, ktoré sú následne uložené do logovacej databázy. Nad týmito dátami

#### 4. IMPLEMENTÁCIA

---



Obr. 4.2: Architektúra systému

je možné následne vyvolať analýzu micro/macro influence a extrahovania tagov, pričom výsledky týchto analýz sú následne uložené do grafovej databázy. Výsledok spracovania dát je možné vidieť na obrázku 4.2.

Implementáciou bola demonštrovaná životaschopnosť navrhovaného riešenia. V nasledujúcej kapitole je detailnejšie analyzovaná validita takéhoto riešenia.

---

## Overenie riešenia

Implementáciou vybraných prípadov použitia v predchádzajúcej kapitole boli položené základy potrebné pre overenie, či sú vyhodnotené dáta validné. Pre overovanie riešenia sa nepoužívali žiadne automatické testy, nakoľko sa validita výpočtu dá odhadnúť prostým pohľadom na vyhodnotené dáta.

### 5.1 Proces overenia

#### 5.1.1 Stiahnutie dát

Stiahnutie dát zo sociálnych sietí a následné definovanie vzťahov medzi jednotlivými užívateľmi (obrázok 5.1) prebieha bez problémov. Dobrá úprava logiky (spomenutá v 4.4) bola potrebná z dôvodu prekročenia API limotov na sieti Twitter.

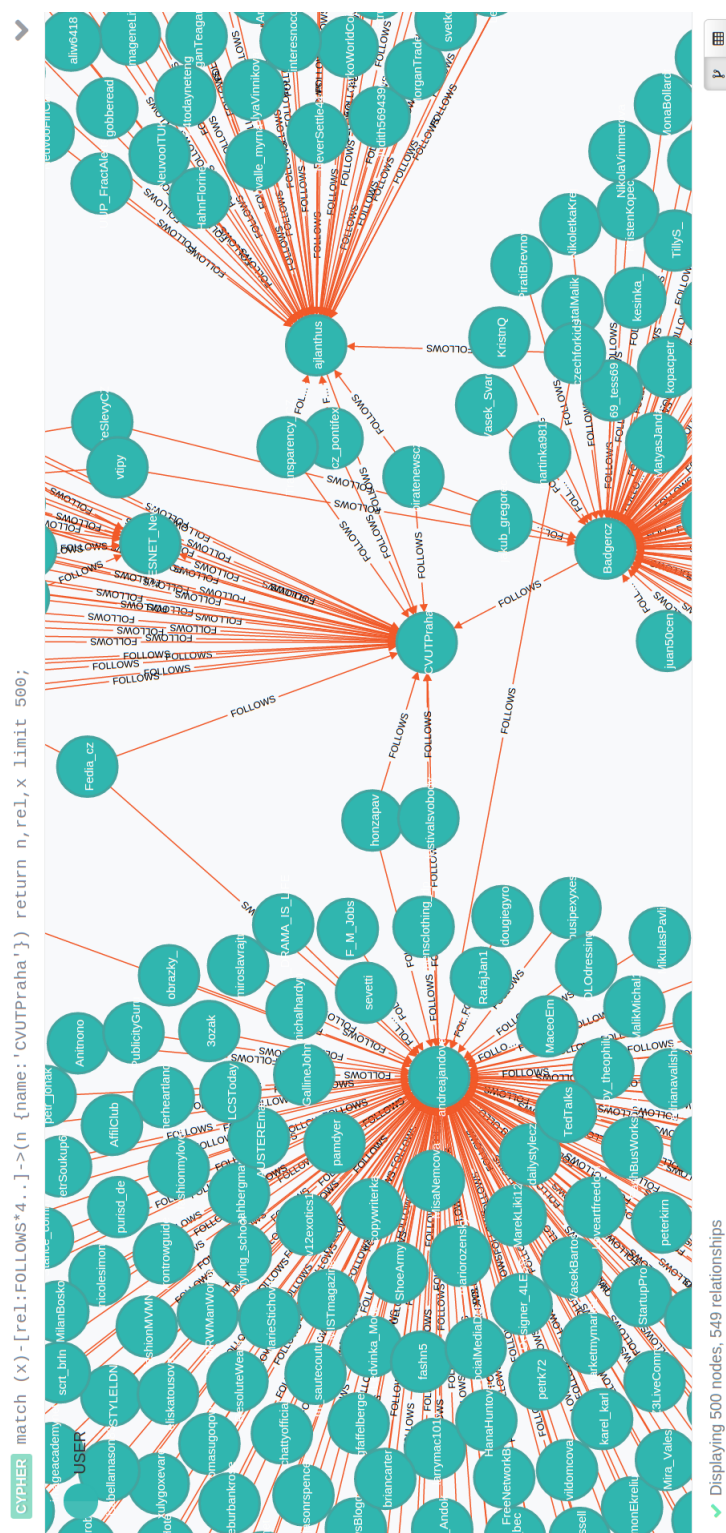
#### 5.1.2 Výpočet micro/macro influence

Vo výpočte micro/macro influence-skóre na základe počtu followerov bolo nutné zohľadniť veľké rozdiely vo vypočítanom skóre medzi užívateľmi s počtom followerov do 5000 a užívateľmi s počtom followerov nad 5000. Pre korektnjšie vyhodnocovanie dát bolo teda nutné pridať matematickú funkciu log.

#### 5.1.3 Extrahovanie tagov

Extrahovanie tagov z profilu užívateľa prebieha bez problémov. Ak však nie je dostupný žiadny priamy tag (slovo začínajúce na #), je tagy potrebné získať pomocou POS tagovania textu. Pre beh tejto analýzy sa dá použiť štandardný tagger pribalovaný do balíčka Stanford NLP knižnice, no na extrahovanie tagov z Twitteru je navrhnutý špeciálny, efektívnejší tagger schopný dosiahnuť až 91% presnosť pri identifikovaní slovných druhov [23]. Preto bol po nedosta-

## 5. OVERENIE RIEŠENIA



Obr. 5.1: Profil ČVUT na sociálnej sieti Twitter



točných výsledkoch štandardný tagger vymenený za špeciálnejší, určený pre Twitter.

## 5.2 Zhrnutie kapitoly

Implementácia a následné overovanie riešenia so sebou priniesli zistenie, že do každého navrhovaného prípadu použitia bolo potrebné v určitej miere vykonať drobné úpravy. Jednalo sa však o pomerne malé zmeny jednotlivých komponentov programu, pričom sa logika celej aplikácie nemusela nijakým spôsobom upravovať. Z tejto skutočnosti vyplývajú dve zistenia:

- Existuje pravdepodobnosť, že v každom ďalšom neimplementovanom prípade použitia existuje nedokonalosť, ktorú bude nutné doplniť až pri samotnej implementácii.
- Logika spracovania dát je navrhnutá dostatočne správne a navyše je dostatočne flexibilná na to, aby sa takéto drobné doplnenia nijakým spôsobom negatívne nepremietli vo validite navrhovaného riešenia.



---

## Doporučenie ďalšieho postupu

Kedže implementovaný stav navrhovaného riešenia nie je konečný a existuje niekoľko konceptov, ktoré v navrhovanom riešení nie sú rozpracované, je vhodné uviesť niekoľko návrhov na zlepšenie procesu vyhľadávania vhodných kandidátov pre reklamu na sociálnych sieťach. Tieto návrhy sa veľkou mierou opierajú o zistenia z analytickej časti práce a vlastné nápady, ktoré vznikli počas písania práce a neboli začlenené do textu.

### 6.1 Finalizácia implementácie

Ako prvé odporúčanie ďalšieho postupu určite navrhujem doprogramovať prípady použitia, ktoré neboli spracované v implementačnej časti. Overí sa tak ich validita a bude možné vytvoriť produkt nasaditeľný na vhodnom aplikačnom serveri.

### 6.2 Doplnenie ďalších analýz

Na základe skutočností zistených z dostupnej literatúry a osobného uváženia je do procesu vyhodnocovania užívateľov vhodné začleniť nižšie uvedené elementy:

- Schopnosť identifikovať fázu nákupného cyklu, v ktorej sa užívateľ momentálne nachádza.
- Schopnosť identifikovať emočný stav užívateľa (buď na základe hlbšej analýzy sentimentu alebo prípadne inou formou - napríklad s využitím merania dynamiky klikania na klávesnicu).
- Rozšíriť alebo vymeniť knižnicu pre spracovanie prirodzeného jazyka - pridanie možnosti spracovať aj iné jazyky ako angličtinu.

## 6. DOPORUČENIE ĎALŠIEHO POSTUPU

---

- Spracovať identifikovanie komunit a vzťahov medzi užívateľmi týchto komunit.

---

## Záver

Cieľom práce bolo analyzovať a navrhnúť postup pre vyhľadávanie užívateľov sociálnych sietí vhodných pre využitie v influence–marketingu.

V rešeršnej časti boli na úvod analyzované marketingové techniky súvisiace s influence–marketingom, popísané jednotlivé existujúce platformy zaoberajúce sa vyhodnocovaním užívateľov, ich porovnanie a určenie miesta, ktoré zastáva navrhované riešenie medzi existujúcou konkurenciou.

V analytickej časti boli identifikované problémy, ktorým čelia existujúce riešenia a bol navrhnutý proces analyzovania dát užívateľa za účelom vyhodnocovania jeho influence–skóre.

Všetky poznatky získané v predchádzajúcich častiach práce boli následne využité pri implementácii troch prípadov použitia do úvodnej verzie programu. Táto implementácia overila validitu riešenia a je súčasťou práce ako príloha (dodatok D).

Na záver práce bolo identifikovaných niekoľko bodov pre zlepšenie výpočtu influence–skóre užívateľa sociálnej siete a smer akým je vhodné ďalej aplikáciu rozvíjať.

Stanovené ciele práce boli úspešne splnené.



---

## Literatúra

- [1] Príbeh Brita: Za sekundu zarobí neuveriteľných 2-tisíc libier. Karierainfo.sk: Zoznam.sk [online]. 2014 [cit. 2015-03-26]. Dostupné z: <http://karierainfo.zoznam.sk/cl/1000164/1422117/Pribeh-Brita-Za-sekundu-zarobi-neuveritelnych-2-tisic-libier>.
- [2] M. KAPLAN, Andreas a Michael HAENLEIN. Users of the world, unite! The challenges and opportunities of Social Media [online]. 2010, s. 59-68 [cit. 2015-04-14]. Dostupné z: <http://openmediart.com/log/pics/sdarticle.pdf>.
- [3] Digital Trends. . THE HISTORY OF SOCIAL NETWORKING [online]. 2014 [cit. 2015-04-16]. Dostupné z: <http://www.digitaltrends.com/features/the-history-of-social-networking/>.
- [4] Newsroom. FACEBOOK, Inc. Our Mission: Stats [online]. 2015. vyd. 2015 [cit. 2015-04-14]. Dostupné z: <https://newsroom.fb.com/company-info/>.
- [5] About Twitter, Inc: Company. TWITTER INC. <https://about.twitter.com/> [online]. 2015 [cit. 2015-04-14]. Dostupné z: <https://about.twitter.com/company>.
- [6] Womma.org. In: The State of Word of Mouth Marketing: A Survey of Marketers [online]. 2014 [cit. 2015-04-16]. Dostupné z: <http://members.womma.org/d/do/778>.
- [7] Emerging Issues in In-Game Advertising. In: BOYD, Greg a Vejay LALLA. <http://www.gamasutra.com/> [online]. 2009 [cit. 2015-04-14]. Dostupné z: [http://www.gamasutra.com/view/feature/3927/emerging\\_iss-ues\\_in\\_ingame\\_.php](http://www.gamasutra.com/view/feature/3927/emerging_iss-ues_in_ingame_.php).

- [8] ELECTRONIC ARTS INC. Need for Speed [online]. 2015 [cit. 2015-04-14]. Dostupné z: <http://www.needforspeed.com/>.
- [9] Harvey, K.: *Encyclopedia of social media and politics*. SAGE Publications, první vydání, 2014, ISBN 1452290261, 1640 s.
- [10] THE KLOUT SCORE. KLOUT, Inc. Klout.com [online]. [cit. 2015-03-26]. Dostupné z: <https://klout.com/corp/score>.
- [11] Kred Scoring Guide. PEOPLEBROWSR. [Http://kred.com/](http://kred.com/) [online]. 2015 [cit. 2015-03-26]. Dostupné z: <http://kred.com/rules>.
- [12] HILHORST, ARTHUR. Influencer Identification: Two Methods To Find The Perfect Influencers. In: Onalytica.com [online]. 2014, 30 Júla 2014 [cit. 2015-04-17]. Dostupné z: <http://www.onalytica.com/blog/posts/influencer-identification-two-methods-to-find-the-perfect-influencers/>.
- [13] Brandwatch PeerIndex. PEERINDEX. PeerPerks: We value you.: Have you heard of our PeerPerks programme? [online]. 2012 [cit. 2015-04-17]. Dostupné z: <http://blog.peerindex.com/peerperks-we-value-you/>.
- [14] BROWN, D. a. S. F.: *Influence marketing: how to create, manage, and measure brand influencers in social media marketing*. Que Publishing; 1 edition, [Kindle DX version]. Retrieved from Amazon.com, první vydání, 2013, ISBN 978-078-9751-041.
- [15] OUR BLOG. PELZER, Adriaan. Klout is broken [online]. 2010 [cit. 2015-04-21]. Dostupné z: <http://wewillraakyou.com/2010/12/klout-is-broken/>.
- [16] Redis.io [online]. 2015 [cit. 2015-03-27]. Dostupné z: <http://redis.io/>.
- [17] KOVACS, Kristof. Cassandra vs MongoDB vs CouchDB vs Redis vs Riak vs HBase vs Couchbase vs OrientDB vs Aerospike vs Neo4j vs HyperTable vs ElasticSearch vs Accumulo vs VoltDB vs Scalaris comparison [online]. 2015 [cit. 2015-03-27]. Dostupné z: <http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis>.
- [18] ROBINSON, I. a. J. W. a. E. E.: *Graph Databases*. O'Reilly Media, první vydání, 2013, ISBN 978-1-4493-5621-7, 224 s.
- [19] NEO TECHNOLOGY, Inc. Neo4J [online]. 2015 [cit. 2015-03-27]. Dostupné z: <http://neo4j.com/>.
- [20] COCKBURN, A.: *Writing effective use cases*. Boston: Addison-Wesley, [Kindle DX version]. Retrieved from Amazon.com, první vydání, 2001, ISBN 02-017-0225-8. Dostupné z: <http://alistair.cockburn.us/get/2465>.



- [21] UML - Tutorial. In: Tutorialspoint [online]. 2015 [cit. 2015-05-01]. Dostupné z: <http://www.tutorialspoint.com/uml/index.htm>.
- [22] Stanford University. STANFORD UNIVERSITY. Stanford CoreNLP: A Suite of Core NLP Tools [online]. 2010, 2015-04-20 [cit. 2015-04-22]. Dostupné z: <http://nlp.stanford.edu/software/corenlp.shtml>.
- [23] L. Derczynski, A. Ritter, S. Clarke, and K. Bontcheva. 2013. Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data. In Proceedings of the International Conference on Recent Advances in Natural Language Processing, ACL.
- [24] MALÝ, Martin. REST: architektura pro webové API. In: Zdroják, o tvorbě webových stránek a aplikací. [online]. 2009 [cit. 2015-05-01]. Dostupné z: <http://www.zdrojak.cz/clanky/rest-architektura-pro-webove-api/>.



---

## Zoznam použitých skratiek

- API** Application programming interface - rozhranie pre komunikáciu so systémom.
- CRUD** Create, Read, Update, Delete - vytvoriť, čítať, meniť a mazať - základné operácie, ktoré zvyčajne poskytujú rôzne webové API.
- DBQ** Database Query - úložisko reprezentujúce frontu dotazov.
- DBL** Database Logging - úložisko pre ukladanie nespracovaných dát.
- DBG** Database Graph - úložisko reprezentujúce spracované dáta.
- REST** Representational state transfer - dátovo orientovaná architektúra rozhrania ktorá umožňuje pristupovať k dátam pomocou protokolu HTTP [24].



---

## Slovník a vymedzenie pojmov

**Užívateľ** - užívateľ sociálnej siete.

**Influence** - vlastnosť užívateľa ovplyvňovať správanie iných užívateľov na sociálnej sieti.

**Zákazník** - človek, ktorý pracuje so systémom, ktorý je predmetom tejto bakalárskej práce.

**Bio** - základné dáta o užívateľovi vo forme plain textu, ktoré o sebe sám užívateľ uvedie. Môžu obsahovať #hash tagy.

**Sledovanie aktivity** - pod pojmom sa rozumie čítanie správ vytvorených užívateľom.

**Tag** - kľúčové slová, ktoré sú extrahované z užívateľových konverzácií, preferencii alebo priamo jeho profilu.

**Influence engine optimization** - technika, ktorou je možné optimalizovať profil užívateľa za účelom zvýšenia skóre v analýze.

**Social scoring platforms** - platformy určené pre vyhodnocovanie influence skóre užívateľov.

**Follower** - nasledovník - užívateľ, ktorý sleduje aktivitu iného užívateľa.

**Influencer** užívateľ, ktorého sledujú followeri.

**Micro influencer** - užívateľ, ktorého správanie na sociálnej sieti má priamy dopad na správanie iných užívateľov - „*Individuals within a consumer's social graph, whose commentary, based on the personal nature of their relationship and communications, has a direct impact on the behavior of that consumer*“.

**Macro influencer** - užívateľ, kt. správanie na sociálnej sieti má veľmi malý alebo žiadny dopad na správanie ostatných užívateľov, slúži však spoľahlivo na zvyšovanie povedomia o danej službe, spoločnosti alebo výrobku - „*Individuals, businesses, or media, with a large, active social following comprised of people with whom they have a loosely defined or unknown relationship*“ [14].

**Use Case** - zoznam krokov, ktoré obvykle definujú interakciu medzi systémom a užívateľom systému. Využíva sa pri vývoji software.

---

## Pužité programy

**TexMaker** - písanie práce v  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  a následné generovanie .pdf verzie bakalárskej práce.

**IntelliJ Idea** - použité na programovanie implementačnej časti bakalárskej práce.

**Neo4J Browser** - webová aplikácia na prezeranie grafovej databázy a testovanie cypher dotazov.

**Draw.io** - webová aplikácia pre vytváranie a úpravu grafov a diagramov.

**Gimp** - editovanie obrázkov.

**Git** - záloha a verzovanie bakalárskej práce.

**Dropbox** - záloha bakalárskej práce.





---

## Obsah priloženého CD

readme.txt.....	stručný popis obsahu CD
impl	
├─ build.....	buildovací gradle script
├─ src.....	zdrojové kódy pilotnej verzie
text .....	text práce
├─ thesis.pdf .....	text práce vo formáte PDF
├─ thesis.tex.....	zdrojová forma práce vo formáte L <sup>A</sup> T <sub>E</sub> X