MORAVSKÁ VYSOKÁ ŠKOLA
Ústav informatiky a aplikované matematiky

Tomáš Daniel

Využití dat Twitteru pro prostorové analýzy

Usage of Twitter data for spatial analysis

Bakalářská práce

Mgr. Vít Pászto, Ph.D.

Olomouc 2018

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a použil jen uvedené informační zdroje. Prohlašuji, že odevzdaná tištěná verze bakalářské práce se shoduje s elektronickou verzí vloženou do IS/STAG.

Olomouc 29.3.2018

Děkuji vedoucímu mé bakalářské práce Mgr. Vítu Pásztovi, Ph.D., za vstřícnost, odborné vedení a čas, který mi věnoval. Děkuji také své rodině a přátelům, kteří mě podporovali během studia i při tvorbě této bakalářské práce.

OBSAH

ÚVOD

Informace o veškerých reálných věcech lze nalézt na internetu. Z ekonomického, politického a sociálního hlediska jsou cenným zdrojem informací sociální sítě, které přenáší vztahy, vazby a komunikaci mezi lidmi do online světa. Sociální sítě jsou dnes nepostradatelným a běžným nástrojem pro firmy při budování jejich vztahu se zákazníkem a poskytují také cenná data pro vědecký výzkum. Umožnují za nízkých nákladů komunikovat se zákazníky či jim předávat velké množství informací. Zajímávým prvkem v této komunikaci je soubor informací o aktivitách lidí, které mohou nejen obchodní firmy o svých zákaznících na sociálních sítích zjistit a využít. Zvoleným zástupcem sociálních sítí je Twitter, který v uživatelských zprávách ukládá geolokační souřadnice s možností využití v prostorové analýze.

Cílem této bakalářské práce je najít postup, jakým lze data ze sociální sítě Twitter získat, dále zpracovat v užitečné informace pro jejich následnou analýzu vzhledem k lokalitě, z které tyto informace pochází. Bakalářská práce je rozdělena na část teoretickou a část praktickou. První teoretická část se zaměřuje na popis základních pojmů, které souvisí se zpracovávanou tématikou využití dat ze sociální sítě Twitter pro prostorovou analýzu. Druhá praktická část obsahuje postup a praktické ukázky prostorové analýzy formou vizualizace prostorových dat. K tématu bakalářské práce jsem použil informace, které jsem získal prostudováním tištěných knih, časopisů a dalších dostupných článků a informací na internetu. Literární rešerše byla provedena v elektronických zdrojích Google v databázích Google Scholar, EBSCO, Scopus. Všechny použité zdroje jsem uvedl v příslušné části své bakalářské práce.

1 TEORETICKÁ ČÁST

1.1 Sociální sítě

Sociální síť patří mezi sociální média, přes které se dá online komunikovat a sdílet obsah nejen osobních pocitů a postojů, ale i chování a činnosti mezi jednotlivci i skupinami lidí. Sociální síť jako druh média nabízí tedy především interakci mezi jejími uživateli a představuje významný zdroj informací. Je možné vytvářet uživatelské profily, navazovat v sociálních sítích vztahy a komunikovat s dopadem na jednotlivce, ale i různě velké skupiny lidí. V souvislosti s dynamickým rozvojem sociálních sítí a rozvojem inteligentních mobilních zařízení vznikají stále rozsáhlejší komunity uživatelů.[1] Zástupcem sociálních sítí, který budeme v této práci zkoumat je sociální síť Twitter.

1.2 Twitter

Sociální síť Twitter jako jedna z mnoha online komunikací umožňuje svým uživatelům rychle reagovat na různé otázky anebo zprávy.

1.2.1 Hlavní charakteristiky Twitteru

Mezi hlavní charakteristiky sociální sítě Twiter patří:

- krátká zpráva – tweet má 280 znaků (140 znaků do listopadu 2017)
- zpřístupnění dat prostřednictvím rozhraní API
- mobilní použití

Tweet – zpráva s možností vložení multimediálního obsahu, publikována a zobrazena veřejně ve službě Twitter. Data, zdrojové kódy a další dokumentace jsou součást Twitter API (Twitter aplikační programovací rozhraní) a jsou poskytované a aktualizované Twitterem prostřednictvím stránky pro vývojáře (softwarová vývojářská

[1] Srov. KOCICH, D., a HORÁK, J., Twitter as a source of big spatial data, *Proceedings of the International Multidisciplinary Scientific GeoConference SGEM*, 2016, sv. 2, č. 1, s. 921, <https://sgemworld.at/sgemlib/spip.php?article8546>.

sada – SDK). Pro každý tweet je generované unikátní identifikační číslo „Tweet ID“. Důležitým nástrojem služby Twitter, který pomáhá sledovat událost nebo téma je Hashtag a je ideální pro rychlé navedení na související informace nebo aktuální dění.[2]

1.2.2 Struktura tweetu

Každý tweet obsahuje mnoho informací, které při jeho stažení můžeme dále filtrovat a ukládat pouze data, která nás zajímají. Pokud pošleme Twitter API požadavek na určitý tweet prostřednictím Rest API, vrátí se nám celá jeho struktura. Tweet obsahuje údaje o uživateli a obsahu zprávy. Je-li povoleno, klient aplikace Twitter přidá do tweetů geotagovaný údaj ve formátu GPS souřadnic. Toto je využito u mobilních zařízení s operačním systémem Android a iOS. Na webových a desktopových klientech používaných u stolních počítačů nebo notebooků již tento parametr zpravidla chybí nebo jeho souřadnice odpovídá přístupovému bodu internetové sítě poskytovatele internetu daného uživatele. Existují nástroje a rozšíření v operačních systémech, které sdílení těchto informací umí blokovat zcela nebo je úspěšně falšovat.[3]

Konkrétní tweet může obsahovat geolokační údaj v koordinátu zeměpisné šířky a délky určené pomocí GPS, případně polohováním získaným z WIFI sítě nebo triangulací ze sítě mobilního operátora. V případě, že uživatel na svém zařízní nemá povolené sdílení informací o své poloze, lze jeho možnou polohu odvodit dle místa, které má uživatel nastavené ve svém uživatelském profilu.[4]

Data stažená ze struktury tweetu prostřednictvím streaming API, která tweet obsahuje, zahrnují unikátní identifikační číslo, datum zveřejnění, uživatelské jméno, název místa, zeměpisnou šířku a délku (volitelný vstup), polygon ze souřadnic dle názvu místa (volitelný parametr) a text zprávy. Tweet, stažený jako objekt, obsahuje mnoho atributů, jejich ukázku znázorňuje mapa struktury tweetu v příloze 1 (str. 50).

---

[2] Srov. YUE, L., QINGHUA, L., a JIE, S., Discover Patterns and Mobility of Twitter Users-A Study of Four US College Cities, *ISPRS International Journal Of Geo-Information*, 2017, roč. 6, č. 2, s. 1-3, <http://www.mdpi.com/2220-9964/6/2/42>.

[3] Srov. KOCICH, D., a HORÁK, J., Twitter as a source of big spatial data, *Proceedings of the International Multidisciplinary Scientific GeoConference SGEM*, 2016, sv. 2, č. 1, s. 921-923, <https://sgemworld.at/sgemlib/spip.php?article8546>.

[4] Srov. WESTERHOLT, R., STEIGER, E., RESCH, B., aj., Abundant Topological Outliers in Social Media Data and Their Effect on Spatial Analysis, *PLoS One*, 2016, roč. 11, č. 9, s. 5, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162360>.

1.2.3 Využití dat z tweetu

Prostřednictvím API rozhraní jsou často stahována a filtrovaná data se spektrem atributů, které můžeme dále zpracovávat podle tématického zaměření. Kvalita a množství získaných dat je rozdílná s ohledem na prostor a čas i další zpracovávání v reálném čase za použití nových aplikací – algoritmů, které geoprostorové údaje analyzují na základě extrakce a fúze textových, časových a prostorových dat a nabízejí jejich využití ve vybraných oblastech konkrétních lidských činností (ekonomika, marketing, turistika, politika…).[5]

1.3 Prostorová analýza

Analýza dat v prostoru se zabývá stanovením otázek či problému v dané oblasti s předpokládanými výsledky. K této oblasti zajistíme požadavovaná data, která se upraví s ohledem na zastoupení prostorových informací. Grafickým zpracováním prostorových informací ověříme předpokládané otázky a možné odpovědi, které sběrem, úpravou a analýzou prostorových dat získáme.

1.3.1 Prostorová data

Prostorová data jsou všechny řetězce s názvy lokací nebo přímých souřadnic, které mohou být v tweetu uloženy. Jednou z důležitých vlastností Twitteru je jeho dostupnost na chytrých telefonech, které mohou automaticky přidávat prostorová data do zpráv především díky GPS modulům. Uživatelé také mohou ručně přidávat svou pozici fomou lokality (město, stát atp.) ve svém profilu. V průměru 2 % tweetů tak obsahují prostorová data, což dělá ze sociální sítě Twitter nepostradatelný zdroj prostorových dat.[6]

[5] Srov. KOCICH, D., a HORÁK, J., Twitter as a source of big spatial data, *Proceedings of the International Multidisciplinary Scientific GeoConference SGEM*, 2016, sv. 2, č. 1, s. 921, <https://sgemworld.at/sgemlib/spip.php?article8546>.

[6] Srov. YUE, L., QINGHUA, L., a JIE, S., Discover Patterns and Mobility of Twitter Users-A Study of Four US College Cities, *ISPRS International Journal Of Geo-Information*, 2017, roč. 6, č. 2, s. 2-3, <http://www.mdpi.com/2220-9964/6/2/42>.

1.3.2 Vizualiace prostorových dat

Souřadnice a lokace vydolované ze sociální sítě Twitter umožňují jejich vizualizaci mnohými vizualizačními nástroji. Geoinformační systémy nám umožňují získaná data zpracovávat v cenné informace, které pak lze použít pro další analýzu. Vizualizací prostorových dat pak rozumíme zejména získané mapy, které jsou výsledkem vizualizace těchto dat.[7]

Pro mapovou vizualizaci dat využíváme mapové služby jako například OpenStreeMap, Google Maps, Google Earth aj. Tyto služby nám umožní spojit získaná data s mapovými podklady a takto prostorová data vizualizovat.[8]

Kromě map s geodaty můžeme vizualizaci provést také pomocí různých grafů a diagramů.

1.4 Nástroje pro vizualizaci a prostorovou analýzu

Moderní www služby umožňují přistupovat k velkému množství informací. Proto byly ve webovém prostředí vyvinuty technologie, postupy a nástroje pro zpřístupnění geodat. Takové rozšíření www služeb lze nazvat vystihujícím označením geoweb. Podle konsorcia OGC lze geowebové služby rozdělit do několika kategorií:

1. Služby pro zpřístupnění geodat
2. Služby pro zpracování geodat
3. Katalogové služby

Služby pro zpřístupnění dat můžeme nazvat mapové servery nebo mapové služby. Nástroje s analytickými funkcemi a například pro transformaci souřadnic nebo změny formátů dat jsou služby pro zpracování geodat. Různé datové sady spolu se základními metadaty s možností sestavování seznamů a vyhledávání v těchto datových sadách umožňují katalogové služby.[9]

[7] Srov. NOVOTNÁ, M., ČECHUROVÁ, M., a BOUDA, J., *Geografické informační systémy ve školách*, s. 9-13.

[8] Srov. tamtéž, s. 44-45.

[9] Srov. RAPANT, P., *Geoinformační technologie*, s. 81-83.

Pro prostorové zpracování dat, jejich vizualizaci a analýzu byly vyhledány tyto webové a desktopové nástroje:

1 ARCGIS ONLINE
2 MS Excel 2016 s pluginem 3D Maps
3 Google Fusion Tables (dynamické tabulky Google)
4 Mapbox
5 QGIS

Uvedené nástroje nabízí různou škálu možností a funkcí, jak lze data prostorově analyzovat a vizualizovat. Předmětem této práce není srovnávat výhody a nevýhody jednotlivých nástrojů. Kritériem pro výběr vhodných nástrojů pro případové studie této práce je uživatelská přívětivost bez hlubší znalosti těchto nástrojů. Na základě tohoto kritéria byly vybrány nástroje z ekosytému firmy Google a Microsoft, které paří mezi globálně nejrozšířenější a nezjnámější sady nástrojů, systémů a služeb.

Pászto uvádí: „Geografické informační systémy byly původně obsluhovány pouze úzkou skupinou odborníků, ale zejména s rozvojem internetu se dnes může stát uživatelem téměr každý.“[10] S tímto tvrzením můžeme jenom souhlasit.

1.5 Způsoby stahování dat z Twitteru

Data lze ze sociální sítě Twitter oficiálně stahovat pouze prostřednictvím API služby, kterou nabízí přímo Twitter. Ostatní způsoby stahování dat, jako například web-scraping, kdy se data stahují ze člověkem čitelného zdroje, jako je webová stránka Twitteru, nejsou dle licenčních podmínek Twitteru přístupné.

1.5.1 Twitter API

Twitter v dnešní době nabízí vývojářům několik verzí API přístupových rozhraní. Ty se mezi sebou liší ve způsobu užití, hodí se pro různé aplikace a mají taktéž různá omezení co do množství volání a parametrů. Při použití více IP adres pro přístup k API je možné částečně tyto omezení zredukovat.[11]

[10] PÁSZTO, V., *Prostorová informace a vybrané metody geocomputation pro její hodnocení*, s. 7.

[11] KWAK, H., aj., What is Twitter, a social network or a news media, s. 592.

1.5.2 Rest API

Umožňuje přístup ke čtení a zapisování Twitter dat. Například vytvořit nový tweet, přečíst data z uživatelského profilu atd. Používá se pomocí ověření OAuth a jeho výstupy jsou ve formátu JSON. Na konkrétní dotazy vrací Rest API zmíněné výstupy. Každé okno aplikace může během 15 minut odeslat 15 dotazů. Využívají se zde GET a POST metody. Je možné se vyhnout vyčerpání limitu na dotazy pomocí cachovacích metod a omezením četnosti pravidelně se opakujících ekvivalentních dotazů.[12]

1.5.3 Search API

Součástí Rest API je tzv. Search API, které má definované parametry, které dotazy mohou obsahovat a jaký formát dat vrací v odpovědi. Search API však dokáže z Twitteru vrátit data maximálně 7 dní stará. Tato metoda zpracování dat je vhodná pro jednotlivé a konkrétní vyhledávání v Twitteru, čtení údajů z uživatelských profilů nebo odesílání tweetů.[13]

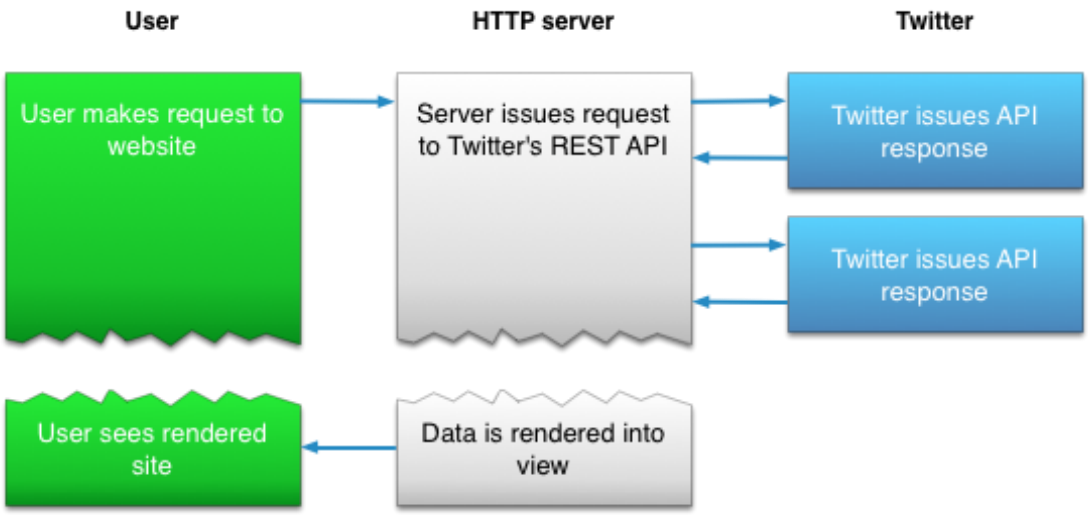
1.5.4 Streaming API

Pokud je pro účel dané aplikace Rest API nevyhovující, mají vývojáři k dispozici Streaming API. Klient využívající ve své aplikaci Streaming API může pomocí trvalého připojení kontinuálně stahovat, ukládat a dále zpracovávat v podstatě neomezené množství tweetů, aniž by přetěžoval koncový bod REST API nesčetným množstvím dotazů. Twitter nabízí tři druhy základních streaming koncových bodů, každý pro specifické využití. Public streams streamuje veřejná data skrze Twitter, což je vhodné pro sledování konkrétního uživatele nebo témat, případně pro dolování dat. User streams zahrnuje hrubá data korespondující s konkrétním uživatelským náhledem na Twitter. Site streams nabízí stejný přístup ke zpracování dat, avšak umožňuje takto přistupovat k datům více uživatelů ve stejném okamžiku. Site streams koncový bod je však ve verzi uzavřeného testovaní a nyní není možné k tomuto koncovému bodu registrovat nové aplikace.[14] Rozdíly mezi Streaming a REST API znázorňují následující obrázky 1 a 2 (str. 13).

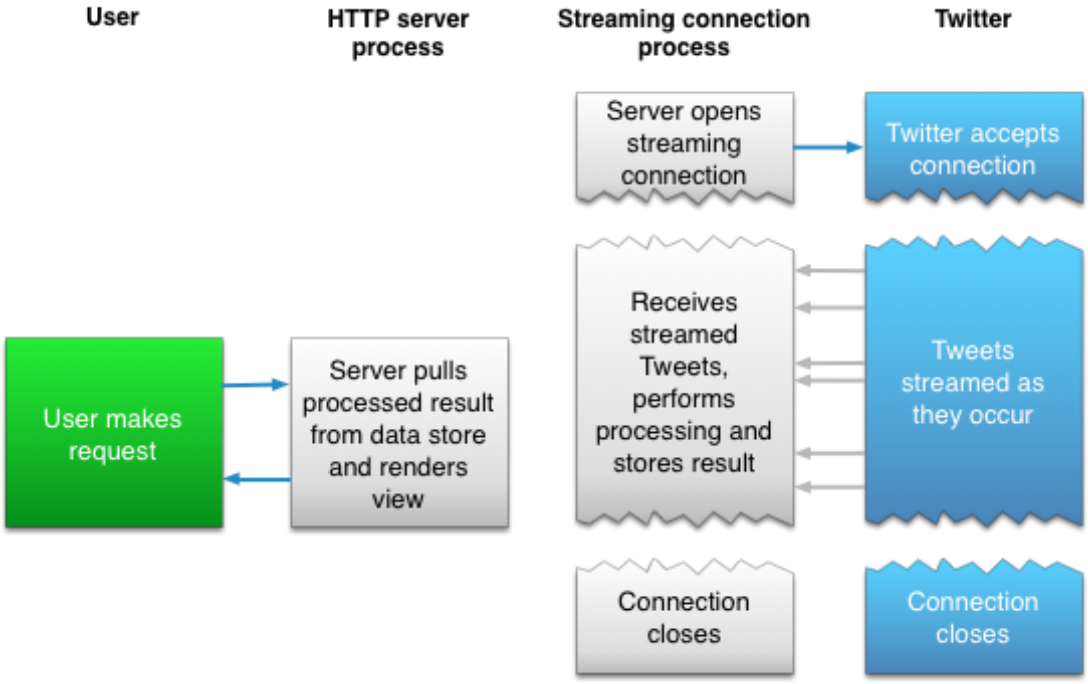
[12] Srov. TWITTER, *API documentation*, <https://dev.twitter.com/>.

[13] Srov. tamtéž.

[14] Srov. tamtéž.



Obr. 1 – Rest API[15]



Obr. 2 – Streaming API[16]

1.5.5 ADs API

Novým API, které Twitter nabízí je ADs API, které umožňuje partnerům Twitteru implementovat vlastní marketingové řešení k automatizovanému publikovaní reklamy

[15] Srov. TWITTER, *Twitter Developers – Docs*, <https://developer.twitter.com/en/docs>.

[16] Srov. tamtéž.

v rámci sociální sítě Twitter. Šíře využití tohoto API je založena na smluvní úrovni, jakou partner s Twitterem uzavře. Developer úroveň slouží pro vývoj a implementaci API do partnerova systému. Součástí testovacího provozu nesmí být žádní externí inzerenti. V průběhu základní verze musí partner během 90 dní prokázat v ostrém nasazení funkčnost a spolehlivost svého designu. Je umožněno mít až 10 privátních inzerentů. Standard verze nabízí plné využití ADs API v potřebném rozsahu reklamní kampaně, který partner Twitteru pro své inzerenty potřebuje.[17]

1.5.6 PowerTrack API

Nová verze API sloužící zejména pro firemní nasazení. Je třeba uzavřít kontrakt s Twitter prodejním týmem. Podstatné především je, jaké množství dat a s jakým omezením hodlá zákazník této služby využít. Tato služba je placená a pro neprofesionální užití poměrně drahá. Základní nabídka nabízí až 1 milion tweetů za období až 40 dní s cenou přesahující 1 000 dolarů. Cenu ovlivňuje zejména množství požadovaných tweetů a rozsah časového rámce, ze kterého mají tweety pocházet. Pro filtrování těchto dat se užívá PowerTrack filtrační jazyk.[18]

1.5.7 Twitter API dokumentace pro vývojáře

Webová dokumentace poskytuje informace pro vývojáře aplikací pracujících s Twitter API. V této dokumentaci jsou popisy všech důležitých parametrů, podmínek provozu a užití Twitter API, vzory zdrojových kódů, metod a hotových knihoven pro různé jazyky a další dostupné nástroje.[19]

[17] Srov. TWITTER, *Twitter Developers – Docs*, <https://developer.twitter.com/en/docs>.

[18] Srov. tamtéž.

[19] Srov. tamtéž.

1.5.8 Autentizační metody

Komunikace prostřednictvím API je podmíněna registrací uživatelského účtu vývojáře a zadáním popisu aplikace, kterou bude API využívat. Komunikace s koncovými body Twitter API běží na zabezpečeném protokolu s autorizovanými požadavky – dotazy. Ověřit pravost lze prostřednictvím uživatelských údajů nebo přímo pomocí údajů aplikace, kdy aplikace sama vytváří požadavky na API bez uživatelského kontextu.[20]

1.5.9 Vytvoření přístupu k Twitter API

Pro vytvoření přístupu k Twitter API je nutné být zaregistrovaný uživatel služby Twitter. Je potřeba získat autorizační údaje pro přístup k API serverům Twitteru. Tyto autorizační údaje lze získat na adrese aplikačního managementu https://apps.twitter.com/. Na této stránce se vytváří nové aplikace, pro které vývojář získává autorizační údaje. Vývojář při vytváření aplikace definuje následující detaily aplikace.

- Název – 32 znakový název aplikace
- Popis – 10 až 200 znakový popis aplikace
- Webová stránka – webová adresa aplikace. Například ke stažení aplikace nebo získání více informací.
- URL zpětného volání – návratová URL adresa po úspěšné autentizaci aplikace

Po vytvoření aplikace poskytne Twitter uživateli tyto klíče a přístupové tokeny (keys and access tokens):

- Consumer Key (API Key)
- Consumer Secret (API Secret)
- Access Token
- Access Token Secret

[20] Srov. TWITTER, *API documentation*, <https://dev.twitter.com/>.

Lze nastavit úroveň přístupu, jakou bude aplikace mít na tyto úrovně:

- čtení
- čtení a zápis
- čtení a zápis přístupem ke přímým zprávám

1.5.10 Vývojářská dohoda

Vývojářská dohoda je součástí tvorby přístupových údajů. Specifikuje podmínky a definuje pravidla pro užití autorizovaného přístupu k Twitter API. Před potvrzením vytvoření přístupových údajů musí vývojář souhlasit s licenčními podmínkami této dohody. Součástí této vývojářske dohody a podmínek užití jsou také pravidla týkající se omezení užití Twitter API. Uživatel této služby by neměl Twitter data záměrně ukládat a odvozovat následující potenciálně citlivé informace o uživatelích:[21]

- Zdraví (včetně těhotenství)
- Negativní finanční stav nebo podmínka
- Politická příslušnost nebo přesvědčení
- Rasový nebo etnický původ
- Náboženské nebo filozofické vztahy nebo víry
- Sexuální život nebo sexuální orientace
- Členství v odborových organizacích
- Údajné nebo skutečné spáchání trestného činu

[21] TWITTER, *Developer terms,* <https://developer.twitter.com/en/developer-terms>.

1.6 Nástroj pro stahování dat

Pro stažení dat prostřednictvím Twitter API lze použít skript zpracovávající metody a volání dle dokumentace Twitter API. K dispozici jsou taktéž hotové skripty, které jsou s patřičnými licencemi k dispozici ve vývojářské službě GitHub. Právě na této službě jsou k dispozici taktéž knihovny pro různé programovací jakyzy, které metody využivané v Twitter API obsahují a zahrnují taktéž proces předání autorizačních údajů potřebných pro proces autorizace dotazů či spojení k Twitter API. Možností je též využít placených i neplacených webových služeb, mezi které patří například Apigee nebo Gnip.

Tyto služby umožňují přistupovat k bezplatnému nebo podnikovému API Twitteru. Data lze stahovat prostřednictvím API nebo konzole pro vyvolání přímých dotazů, kde lze vyvolávat dotazy pro konkrétní Twitter metody.[22]

Jednoduchým jazykem přehledným i pro neprogramátory či programátory začátečníky je jazyk webových aplikací PHP. Pro potřeby této práce využiji skript „Twitter Search using the Twitter API v1.1 & PHP“ dostupným na webové stránce techiella.x0.com.[23]

Tento skript využívá twitteroauth PHP knihovny pro OAuth REST API napsané Abrahamem Williamsem, a její zdrojové kódy jsou včetně licence dostupné na GitHubu.[24]

Vlastní skript umožňuje prohledávat Twitter, upravovat vyhledávací dotazy („querry“) prostřednictvím Twitter API ve verzi 1.1.

[22] APIGEE, *Twitter API console,* <https://apigee.com/console/twitter>.

[23] TECHIELLA, *Twitter Search using the Twitter API v1.1 & PHP,* <http://techiella.x0.com/twitter-search-using-the-twitter-api-php/>.

[24] GITHUB, *abraham/twitteroauth,* <https://github.com/abraham/twitteroauth>.

1.6.1 Zdrojový kód pro metodu Search API

Původní zdrojový kód pro stahování tweetů prostřednictvím Search:

```
<?php
require_once 'lib/twitteroauth.php';
define('CONSUMER_KEY', 'your_consumer_key');
define('CONSUMER_SECRET', 'your_consumer_secret');
define('ACCESS_TOKEN', 'your_access_token');
define('ACCESS_TOKEN_SECRET', 'your_access_token_secret');
function search(array $query)
{
$toa = new TwitterOAuth(CONSUMER_KEY, CONSUMER_SECRET,
ACCESS_TOKEN, ACCESS_TOKEN_SECRET);
 return $toa->get('search/tweets', $query);
}
$query = array(
 "q" => "happy birthday",
);
$results = search($query);
foreach ($results->statuses as $result) {
 echo $result->user->screen_name . ": " . $result->text . "\n";
}
```

K spuštění skriptu je možné využít terminál/shell v operačním systému GNU/Linux pomocí nainstalovaného PHP frameworku. Budeme používat operační systém Fedora 23. V prostředí Windows 10 je díky nedávno představené integraci linuxového terminálu (Ubuntu) možno provést následující postup obdobným způsobem. Pro spuštění tohoto skriptu vytvoříme v uživatelském adresáři přihlášeného uživatele (/home/username) složku např. twitter-search (příkaz mkdir) a do ní umístíme potřebné zdrojové kody:

- Adresář lib se soubory PHP knihovny OAuth.php a twitteroauth.php
- Vlastní PHP skript – nazveme např. search.php

Soubory s potřebným skriptem a knihovnami jsou v elektronické podobě uloženy na DVD-ROM jako příloha 6 „program twitter-search“. Abychom tento skript mohli spustit, musíme mít v operačním systému linux nainstalovanou podporu PHP, tedy frameworku, který nám zmíněný skript umožní spouštět v terminálu/shellu. Instalaci provedeme ve Fedoře pomocí terminálu následovně:

```
sudo dnf install php
```

Přepneme do adresáře twitter-search následujícím příkazem:

```
cd ~/twitter-serach
```

Skript zavoláme následovně:

```
php search.php
```

Aby se skript úspěšně připojil k Twitter API, je nutné, abychom vyplnili autentizační klíče a tokeny, které jsme získali při registraci vlastní aplikace na vývojářské stránce Twitteru. Po připojení k Twitter API dojde k zadání dotazu z pole „querry", které je třeba předvyplnit podle požadovaných klíčových slov a dalších specifikací.

1.6.2 Úprava zdrojového PHP kódu

Úpravu zdrojového kódu provedeme na několika místech kódu. Zavolání knihovny twitteroauth.php a předání autorizačních klíčů a tokenů:

```
<?php
require_once 'lib/twitteroauth.php';
define('CONSUMER_KEY', '*********zzUvJo');
define('CONSUMER_SECRET', '*********j6YmnQFn');
define('ACCESS_TOKEN', '*********7VHykyi9');
define('ACCESS_TOKEN_SECRET', '*********99KdA6Ui');
```

Definice funkce, volací metody a odeslání pole s dotazem:

```
function search(array $query)
{
 $toa = new TwitterOAuth(CONSUMER_KEY, CONSUMER_SECRET,
ACCESS_TOKEN, ACCESS_TOKEN_SECRET);
 return $toa->get('search/tweets', $query);
}
```

Pole s definicí dotazu. Změna typu, rozšíření „geocode“ – lokalitu s kruhovou vzdáleností a jazyk hledaných tweetů:

```
$query = array(
 "q" => "klicove+slova",
 "count" => 100,
 "result_type" => "mixed",
 "geocode" => "49.1950600,16.6068370,100km",
 "lang" => "en",
);
```

Filtr vyhledávající definované parametry z datové struktury stažených tweetů a jejich vrácení výpisem v terminálu:

```
$results = search($query);
foreach ($results->statuses as $result) {
 echo $result->user->screen_name . ": " . $result->text . "\n";
}
```

Úpravu zdrojového kódu pro ukládání vyhledaných dat do souboru nebudeme řešit, z důvodů že tento způsob má ve vyhledávání definovaných dat velice omezené možnosti a pro kvantitativní vyhodnocování v prostorové analýze by získaná data neměla vypovídající váhu. Pro získání dostatečného množství požadovaných dat je výhodnější zvolit jiný program pracující s kontinuálním tokem dat z Twitteru prostřednictvím Streaming API.

```
○ → php search.php
tmj_cze_jobs: Can you recommend anyone for this #job in #Moravany, South Moravian Reg
ion? https://t.co/zHHvIPuvNo #IT #Hiring #CareerArc
tmj_cze_jobs: This #job might be a great fit for you: Sr. Associate, Network Services
 Maintenance - https://t.co/1V9SdlQrXZ… https://t.co/Bn2NuXULBN
attCAREERS: Want to work in #Moravany, South Moravian Region? View our latest opening
: https://t.co/b9jP7H1Cur #Engineering #Job #Jobs #Hiring
```

Obr. 3 – Výstup stahovaných dat na terminálu[25]

Na obrázku 3 jsou v terminálu zobrazena získaná data prostřednictvím Search API z PHP skriptu „twitter-search“. Zpracování dat by při této verzi aplikace spočívalo v manuálním kopírování a ukládání do textového nebo CSV souboru, kde by bylo třeba dále data filtrovat a zpracovat. Je možné skript upravit tak, aby tyto elementární kroky prováděl automaticky, ale bylo zjištěno, že pro praktickou část této práce existuje vhodnější řešení. Metoda stahování dat pomocí Search API byla s uvedenými omezeními vyzkoušena jednoduchými vyhledávacími dotazy.

Pro praktické potřeby této práce byla využita aplikace v pogramovém jazyce Python se zdrojovým kódem, který je pod licencí MIT dostupný na GitHubu.[26]

[25] Zdroj vlastní

[26] GITHUB, *twitter-streamer*, <https://github.com/pschiffe/twitter-streamer>.

2 PRAKTICKÁ ČÁST

Cílem praktické části je získání dat ze sociální sítě Twitter, filtrování těchto dat a následně jejich prostorová analýza. K získání dat jsme si vytvořili přístup k Twitter API, protože jiný způsob sběru dat licenční podmínky Twitteru neumožňují. Pro získání dat, prostřednictvím Twitter API, jsme zhodnotili a vybrali vhodné nástroje – program nebo skript v jazyce PHP nebo Python. Tyto nástroje pomocí klíčových slov a omezením datových atributů poskytly data potřebná pro konkrétní prostorovou analýzu. Získaná data byla uložena do souboru CSV, dále importována do programu Excel, kde byla filtrována pomocí filtrů a vzorců. Upravená data byla vizualizovaná v prostoru na základě prostorových dat a klíčových slov. Vznikly tak mapy nebo grafy popisující výsledky konkrétní případové studie.

2.1 Zdrojový kód pro stahování dat v Streaming API

Pro zprovoznění tohoto skriptu byl využit operační systém Fedora dle instrukcí na GitHubu. Ukázka zdrojového kódu je přiložena v tištěné příloze 3 „Zdrojový kód twitter-streamer“ (str. 52-56). Celý program je v elektronické podobě uložen na DVD-ROM jako příloha 7 „program twitter-streamer“.

Jednoduchý skript využívá python knihovnu Tweepy[27] pro práci se streamem tweetů. Nastavení parametrů a klíčových slov pro omezení streamovaných dat lze provést v konfiguračním souboru config.ini. Tweety jsou ukládány v CSV souboru a vypsány ve standardním výstupu stdout. Jako oddělovač sloupců se používá čárka.

Aby tento Python skript fungoval, je potřeba nainstalovat knihovnu Tweepy, kterou lze nainstalovat jako balíček například ve Fedoře následujícím příkazem:

```
sudo dnf install python3-tweepy
```

Nebo lze použít příkaz pip:

```
pip3 install tweepy
```

Konfigurace je provedena zkopírováním config.ini.example do config.ini, doplněním aplikačních API klíčů a tokenů a upravením klíčových slov dle potřeby:

```
cp config.ini.example config.ini
```

[27] TWEEPY, *Tweepy Documentation,* <http://tweepy.readthedocs.io/en/v3.5.0/>.

Spuštení aplikace lze provést zavoláním příkazu v terminálu:

```
./twitter_streamer.py
```

Nástroje využité pro stahování tweetů tedy jsou operační systém Linux (např. Fedora, CentOS), programovací jazyk Python[28] a knihovny s připravenými funkcemi (Tweepy, OAuth).

2.2 Metody sběru dat a vizualizace pro případové studie

Proces sběru dat a vizualizace získaných dat proběhl v těchto následujících krocích:

1. Sběr dat
2. Filtrace a úprava dat pro prostorové analýzy
3. Vizualizace pomocí zvoleného nástroje

Některé kroky se vzájemně prolínají, protože zvolením atributů obsažených v tweetu, do jisté míry vyfiltrujeme data. Při samotné vizualizaci lze například u Gooogle Fusion Tables filtrovat data, které pro konkrétní mapy chceme zobrazit.

2.2.1 Sběr dat

Sběr dat začíná spuštěním Python skriptu s definovanými klíčovými slovy. Streaming API nám na základě těchto klíčových slov kontinuálně posílá tweety, jejichž obsah odpovídá definovaným klíčovým slovům. Dále Python skript ukládá pouze atributy tweetu, které si do funkce „csv.write“ nadefinujeme, tím kromě klíčových slov provádíme předfiltraci dat, to ukazuje tento kód:

```
csv.write('"%s","%s","%s","%s"\r\n' % (user_name, place, coord, text)
```

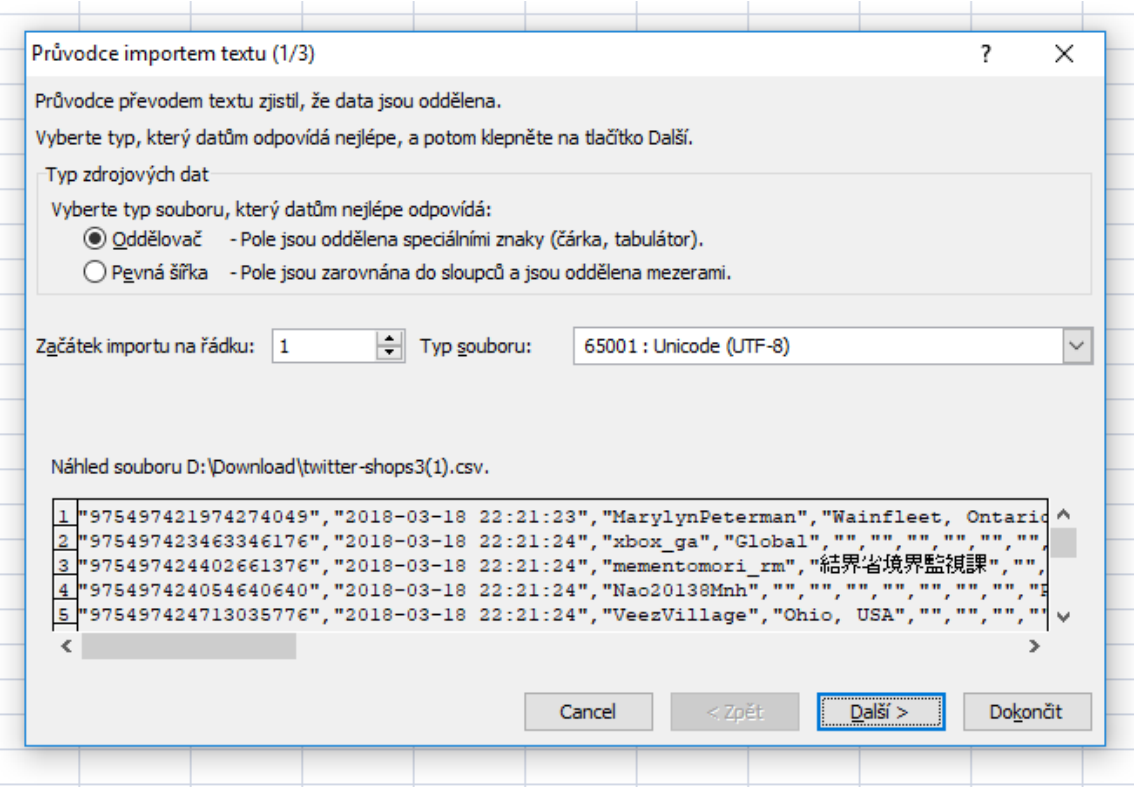
2.2.2 Filtrace dat a úprava pro prostorovou analýzu

Vlastní CSV soubor obsahuje hodnoty oddělené čárkami, takto je každý tweet uložený na samostatném řádku a jednotlivé atributy tweetu odděluje čárka. Pokud je hodnota položky uvedena ve dvojitých uvozovkách, jedná se o text, kdy čárka v tomto textu neodděluje jednotlivé atributy tweetu.

[28] PYTHON, *Welcome to Python*, <https://www.python.org/>.

CSV soubor byl následně importován do nového sešitu aplikace Microsoft Office Excel 2007 v položkách menu: „Data“, „Načíst externí data z textu“ a „Průvodce importem textu“.

Při importu textu se zvolí, že jsou pole oddělená speciálním znakem (čárka, tabulátor) a zvolí se kódování souboru UTF-8. Okno importu CSV souboru ukazuje obrázek 4.



Obr. 4 – Import CSV souboru do aplikace Excel[29]

Importovaný CSV soubor obsahuje volitelné sloupce jako „tid“ (tweet_id), „date“ (datum), „user“ (uživatel), „place“ (lokace), „text“ (text zprávy). Také jsou do CSV souboru přidány sloupce, které umožňují data dále filtrovat. Sloupec „poradi“ je vyplněn a stanovuje pořadí, v jakém byly tweety do souboru uloženy. Následně je použita funkce Excelu seřazení dat s rozšířením vybrané oblasti ve sloupci „place“, tím na prvních pozicích seřadíme tweety, které obsahují protorový údaj. Twitter v rámci streamování dat prostřednictvím Streaming API nevrací u tweetu údaj, idenfikující klíčové slovo, proto je zapotřebí přidat několik sloupců s funkcemi, které v textu tweetu najdou klíčové slovo.

[29]Zdroj vlastní

Popis všech přidaných slopců a funkcí je uveden v tabulce 1.

Název sloupce	Účel sloupce	Funkce nebo podmínka
gf_shop_icon	číslo ikonky pro fusion tables	
@	první znak textu (@)	=ZLEVA(L2;1)
RT	první tři znaky textu (RT)	=ZLEVA(L2;3)
AMAZON	zjistí zda text obsahuje amazon	=KDYŽ((JE.ČISLO(HLEDAT("amazon";L2)));"amazon";"ostatní")
EBAY	zjistí zda text obsahuje ebay	=KDYŽ((JE.ČISLO(HLEDAT("ebay";L2)));"ebay";"ostatní")
TWEET	zjistí, zdali je zpráva tweet, retweet nebo přímý tweet	=KDYŽ(F2=$F$1;"direct tweet";KDYŽ(G2=$G$1;"retweet";"twee t"))
FILTR	zjistí, zdali zpráva obsahuje slovo amazon, ebay nebo ostatní	=KDYŽ(H2=I2;"ostatní";KDYŽ(A(H2="am azon";I2="ostatní");"amazon";KDYŽ(A(H 2="ostatní";I2="ebay");"ebay";"amazon +ebay")))

Tab. 1 – Tabulka funkcí a vzorců přidaných v Excel souboru[30]

Uvedené sloupce zjišťují, zdali je text zprávy přímá zpráva jednoho uživatele jinému uživateli, to platí, pokud začíná „@“. Pokud začíná „RT “, pak se jedná o retweet. Následně ve zprávě hledáme výskyt klíčových slov. Některé zprávy však mohou obsahovat klíčové slovo v URL adrese, kterou tweet může také obsahovat. Vzniká problém, že Twitter odkazy automaticky zkracuje na vlastní, tzv. „tiny URL“, Bylo nutno využít atribut tweetu obsahující originální URL adresu, jestliže tweet obsahuje zkrácenou adresu v textu tweetu, a touto originální URL adresou nahradit zkrácenou URL adresu, aby mohlo být rozhodnuto o obsahu klíčových slov v textu zprávy tweetu.

V případě potřeby je možné geokódovat data pro převod textových názvů lokalit na souřadnice. Tuto službu nabízí řada online nástrojů, z nichž uvádím ty, které byly využity v praktické části této práce. Twitter nabízí geokódování pomocí API pro převod textových názvů lokalit na souřadnice pomocí funkce „GET geo/reverse_geocode*“*.[31] Obdobným způsobem lze zjistit název místa pomocí koordinátů či IP adresy díky funkci „GET geo/search*“*. Python aplikace twitter-streamer tyto funkce podporuje díky knihovně Tweepy. Google nabízí Geocoding API, které převádí textové názvy na souřadnice. Nástroj Google Fusion Tables toto aplikuje na sloupce označené formátem „Locations“ automaticky. V rámci Google Maps API jsme povolili verzi Standard, která

[30] Zdroj vlastní

[31] TWITTER, *Twitter Developers – Docs,* <https://developer.twitter.com/en/docs/>.

odstraňuje nízký limit geokódovaných údajů. Bezplatně totiž Google geokóduje pouze 2500 názvů na den. Každých dalších 1000 požadavků stojí pouze 0,5 USD až do celkového denního limitu 100 000 dotazů. [32]

2.3 Případová studie obchodní společnosti

Pro praktickou ukázku analýzy prostorových dat můžeme zvolit různé oblasti. Teoretickým příkladem může být hledání lokalit s nejvíce vyskytujícím se klíčovým slovem pro stanovení oblastí nákupu reklamy, rozšíření obchodní sítě atp. Lze sledovat konkurenci v prostoru a vyhledávat možné obchodní příležitostí v prostoru. Prostorová analýza je založena na definici dotazů, nejčastěji pomocí klíčových slov. V obchodním světě je uživatel Twitteru aktuální nebo potenciální zákazník. Vlastní analýza s využitím dostupných služeb a programů nám může vytvořit jako bodovou mapu nebo heatmapu (mapa zastupující četnost výskytu barevnou škálou teplot), kdy konkrétní hotspot (území rozlišené teplotní škálou) ukazují četnost výskytu definovaných dat.

2.3.1 Sběr dat

Klíčová slova pro sběr dat na téma Amazon nebo eBay byla stanovena pouze dle názvů těcho obchodních společností, prodávajících produkty nebo služby. Nastavení klíčových slov bylo provedeno v souboru config.ini v python aplikaci twitter-streamer takto:

```
filter_track = amazon, ebay
```

Sběr dat probíhal od 18.3.2018 do 19.3. 2018. CSV soubor nasbíraných dat je v elektronické podobě uložen na DVD-ROM v souboru příloh 10 „zdrojova data obchodni spolecnosti“. Celkem bylo do souboru ze streaming API uloženo 492 111 tweetů.

[32] GOOGLE, *Pricing and Plans,* <https://developers.google.com/maps/pricing-and-plans/#details>.

2.3.2 Filtrace a úprava dat pro prostorové analýzy

Soubor CSV byl importován do aplikace Excel, kde byl rozdělen na list „zdrojova_data“, který kromě přidaného sloupce „poradi“ obsahoval z CSV souboru importované sloupce:

tid, date, user, user_place, coordinates, place, polygon1, polygon2, polygon3, polygon4, text

Následně byl vytvořen list s názvem „filtrovana_data“, kde byly vloženy pouze tweety obsahující data v sloupci „place“, což je textový název v prostoru – lokalita, kterou tweetu přidělil Twitter. Tímto byly vyfiltrované tweety s geokódovanou informací. Sloupec „user_place“ obsahoval uživatelem vyplněný popis místa, ze kterého pochází, ale tato data byla nerelevantní geodata, protože obsahovala texty jako „mléčná dráha“ nebo „aladinova lahev“. Sloupce s polygony také vygeneroval Twitter z názvu v atributu „place“, jako možnost zobrazit oblast ze sloupce „place“ v polygonu. Polygon je více souřadnicemi stanovená oblast, kterou lokalita ze sloupce „place“ pokrývá. Převedení těchto lokalit do prostorových dat bylo provedeno až v rámci Google Fusion Tables, který umí data ze sloupce označeného formátem „Location“ geokódovat na souřadnice z textových názvů v sloupci „place“.

Pro vzorečky a filtry v Excelu byly přidány sloupce:

gf_shop_icon, @, RT, AMAZON, EBAY, TWEET, FILTR

Tyto vzorce a filtry nám podle textu zprávy ve sloupci „text“ zjistily, zdali je text zprávy tweetem, přímou tweet zprávou nebo retweetem. Také nám umožnily stanovit, zdali text zprávy obsahuje názvy obchodů „amazon“ nebo „ebay“. V případě, že z tweetu nebylo možné zjistit, zda text zprávy obsahuje „amazon“ a „ebay“, označil jej vzorec ve sloupci „FILTR“ jako „ostatní“. Bylo zjištěno, že u zpráv označených jako „ostatní“ použili uživatelé citace cizích tweetů. Tyto citace v textu zprávy Twitter zastupuje URL adresou citovaného tweetu, z které nebylo možné zjistit klíčová slova. Tyto geokódované tweety tak byly v analýze dat uvedeny jako „ostatní“ tweety. Excel soubor upravených dat je v elektronické podobě uložen na DVD-ROM v souboru příloh 10 „zdrojova data obchodni spolecnosti“.

2.3.3 Vizualizace a prostorová analýza

V souboru dat Excel jsme vytvořili další dva listy. První list „kontingencni_data“ je souborem dat z tabulky „filtrovana_data“ obsahující všechny lokace z tweetů. Tento soubor dat jsme použili pro vytvoření kontingenční tabulky v druhém listu „kontingencni_tabulka“, ve které je součet všech 2 639 míst, tedy geolokačních údajů. Zároveň je na stejném listu uvedena tabulka, kde jsou lokality seřazeny dle celkového množství tweetů v konkrétních lokalitách. Ze všech tweetů tak bylo získáno 0,55 % geotagovaných tweetů.

Data z finálně upraveného Excel souboru byla importována z listu „filtrovana_data“ do nástroje Google Fusion Tables (dynamické tabulky), který byl zvolen pro prostorovou vizualizaci případové studie. Pro tento účel musí být uživatel přihlášený do služby Google prostřednictvím Google účtu. Import zdrojových dat pro dynamickou tabulku lze provézt třemi způsoby: vložením dat z počítače, z Google Tabulky nebo vytvořením prázdné dynamické tabulky. Vložený soubor může mít velikost až 250 MB a podporované formáty jsou CSV, TSV, TXT nebo KML. Datový limit pro všechy uživatelské dokumenty ve službě Google Docs (Google dokumenty) je v současné době 1 GB.[33]

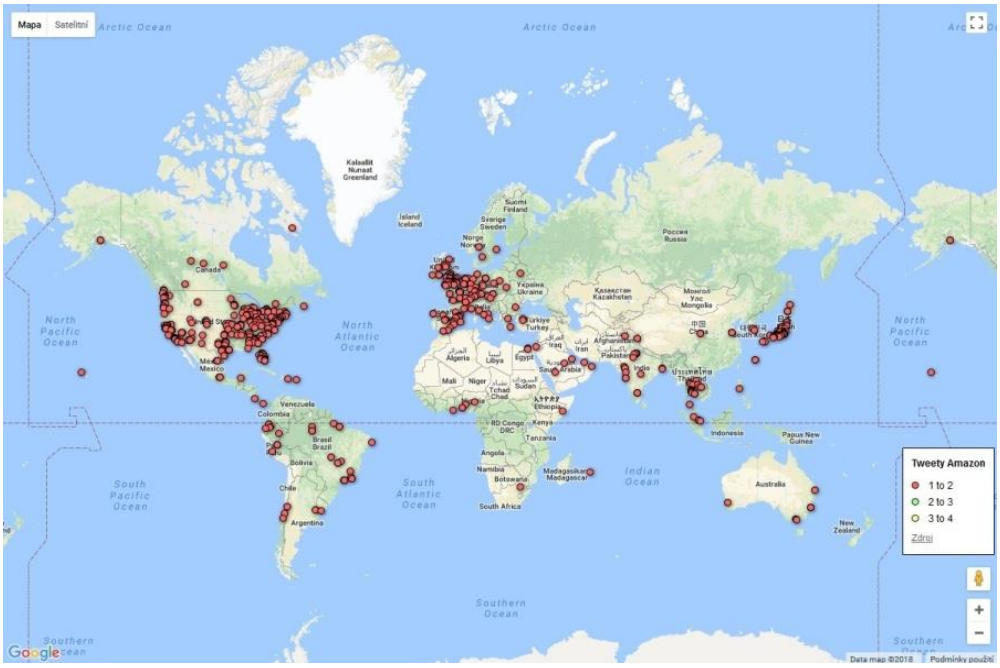
V dynamické tabulce[34] jsou „amazon“ a „ebay“ klíčová slova. Pro prostorovou analýzu jsme v importované tabulce „Data“ využili automatického geokódování lokalit v sloupci „place“. Následně byla vygenerována bodová mapa pojmenovaná „Featured map“, která jednotlivé tweety zobrazila dle jejich lokace. Vygenerována byla také heatmapa. Výsledná data byla dle sloupce „gf_shop_icon“ rozdělena podle číselného označení. Úprava stylu mapy a ikonek umožnila data podle číselného označení barevně rozlišit. Aplikací filtrů dat vznikly mapy tweetů, které jsou vizualizovány pomocí mapových podkladů Google Maps. Vytvořené obrázky map s větším rozlišením s jejich html interaktivní variantou jsou v elektronické podobě uloženy na DVD-ROM v souboru příloh 8 „soubor map obchodni spolecnosti“. Google Fusion Tabulka je uložena na DVD-ROM v příloze 12 „odkaz na google fusion tabulku“.

[33] GOOGLE, *Type and size of files to import*, <https://support.google.com/fusiontables/answer/171181>.

[34] GOOGLE Fusion Tables, *Amazon a eBay klíčová slova pro prostorovou analýzu*, <https://www.google.com/fusiontables/DataSource?docid=1dHzwbmJ8IexcV0P4oLPVsajbSJ0SHe6X2nLmOQf0>.

Obr. 5 – Lokace všech tweetů[35]

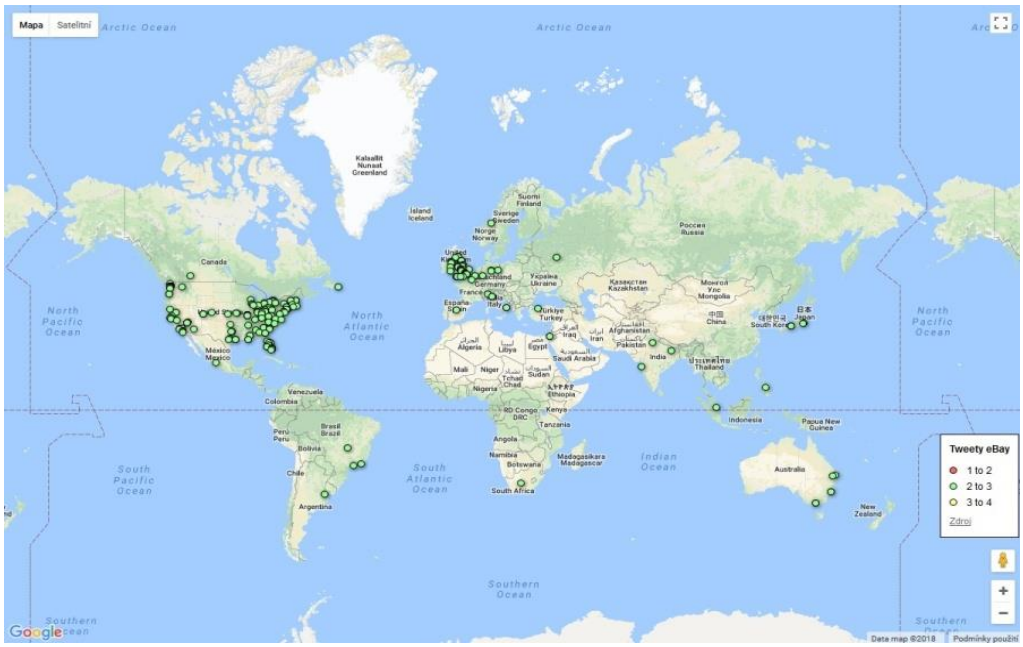
Obrázek 5 vizualizuje na mapě světa lokality tweetů všech katerogrií ze sloupce FILTR, tedy „amazon“, „ebay“, „amazon+ebay“ nebo „ostatní“. Obrázek 6 zobrazuje lokace tweetů „amazon“.



Obr. 6 – Lokace tweetů Amazon[36]

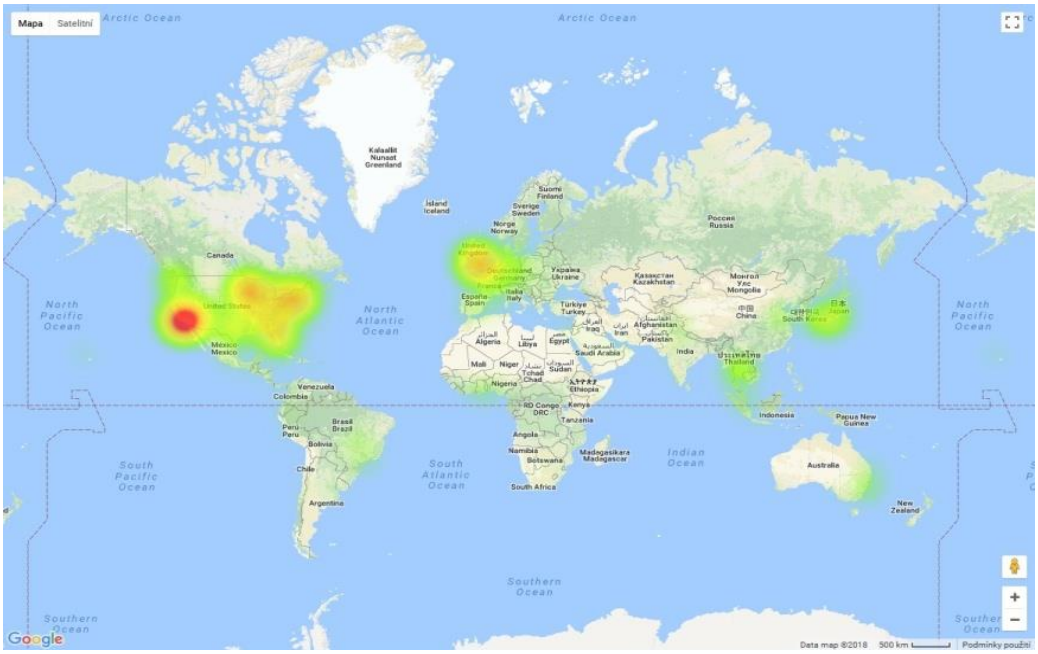
[35] Zdroj vlastní

[36] Zdroj vlastní



Obr. 7 – Lokace tweetů eBay[37]

Přehled všech lokalit s tweety „ebay“ je znázorněn na obrázku 7. Heatmapa na obrázku 8 znázorňuje lokality dle barevné teploty výskytu tweetů od největšího množství všech tweetů (červená barva) až po lokality s nejméně tweety (zelená barva).



Obr. 8 – Heatmapa všech tweetů[38]

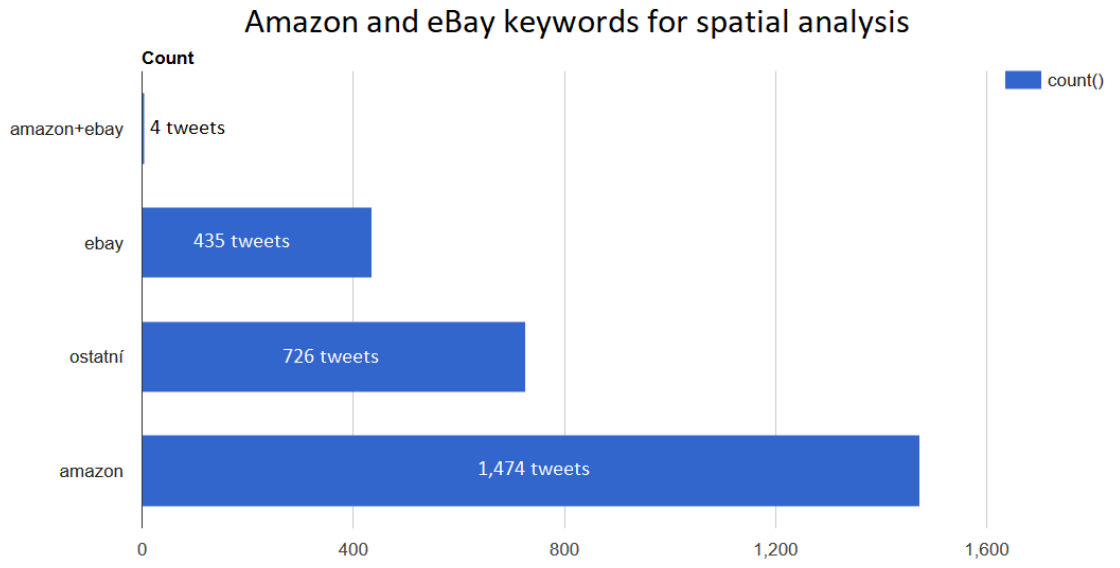
[37] Zdroj vlastní.

[38] Zdroj vlastní.

2.3.4 Výsledky

Výsledek prostorové analýzy geotagovaných tweetů znázorňuje graf 1 vygenerovaný v Google Fusion Tables. Nejvíce tweetovaným klíčovým slovem byl „amazon“ s celkovým množstvím 1 474 tweetů (55,9 %). Pouze 4 tweety (0,1 %) zmiňují jak „amazon“, tak „ebay“ v jedné zprávě. Klíčové slovo „ebay“ bylo zmíněno u 435 geotagovaných zpráv (16,5 %). U 726 zpráv (27,5 %) pak nebylo z textu zprávy možné určit, zdali obsažený URL odkaz či citovaný tweet pod touto adresou obsahuje jedno z definovaných klíčových slov. Podle heatmapy byla největší koncentrace tweetů v Severní Americe, Evropě a Japonsku. Nejmenší koncentrace tweetů se objevovala v Jižní Africe, Rusku, střední Asii a Kanadě. Konkrétní místo, které vykázalo nejvíce tweetů, je město Los Angeles.

Graf 1 Součet všech tweetů podle klíčových slov



Zpracováno v Google Fusion tables
Zdroj: vlastní

2.4 Případová studie nálada veřejnosti

Druhou praktickou ukázkou analýzy prostorových dat je oblast nálady veřejnosti. Sterne se zabývá v souvislosti s rozeznáváním nálady veřejnosti problematikou měření nálady. Pro měření nálady doporučuje jako užitečné mít barevný graf a teploměr. Barevný graf zobrazuje polaritu a temploměr intenzitu nálady.[39]

Stránka Twitterart.com sestavuje seznam pozitivních a negativních klíčových slov hodnotících náladu. Konkrétní kladná nebo záporná slova mohou být vyjádřena i emotikony (smajlíky). Takové seznamy slov lze následně použít k automatizovanému chápání nálady a sledování veřejného mínění. Neustálým rozšiřováním a vylepšováním algoritmů zpracovávajících sociální média pak přináší lepší výsledky a cit při analýze nálady veřejnosti. Některé nástroje jako Twitterart.com jsou na dobré cestě hodnocení nálady.[40]

Klíčovými slovy pro druhou praktickou ukázku analýzy byly usmívající a mračící se smajlíci. Zajímaly nás výsledky četnosti zastoupení jednotlivých smajlíků v konkrétních lokalitách. Smajlík usmívající se vyjadřoval slovo s pozitivní náladou a smajlík mračící se slovo s negativní náladou.

2.4.1 Sběr dat

Klíčová slova v souboru config.ini byla nastavena takto:

```
filter_track = :-), :), :-(, :(
```

Sběr dat probíhal 6 dní od 19.3.2018 do 25.3.2018. CSV soubor nasbíraných dat je v elektronické podobě uložen na DVD-ROM v souboru příloh 11 „zdrojova data nalada verejnosti“. Celkem bylo do souboru ze streaming API uloženo 38 241 tweetů. Tweety bez geotagovaných dat nebyly do souboru CSV ukládány.

2.4.2 Filtrace a úprava dat pro prostorové analýzy

Soubor CSV byl importován do aplikace Excel, kde byl vytvořen list „zdrojova_data“, který kromě přidaného sloupce „pořadí“ obsahuje sloupce atributů ze souboru CSV:

[39] Srov. STERNE, J., *Měříme a optimalizujeme marketing na sociálních sítí*, s. 111-118.

[40] Srov. tamtéž, s. 119-124.

tid, date, user, user_place, coordinates, centroid, place, polygon1, polygon2, polygon3, polygon4, text

V tomto případě, již data obsahovala atribut centroid, který byl vygenerován Twitterem a ukládán v průběhu stahování dat ze Streaming API. Plugin 3D mapy aplikace Microsoft Excel 2016 umí provést geokódování atributu „place“. U 82 % lokalit Excel geokodoval lokace se 100% jistotou správnosti místa dle poskytnutého názvu ze sloupce „place“. Zbývající lokality byly manuálně zkontrolovány a textové názvy odpovídaly navrhovaným lokacím ve všech případech.

Pro vzorečky a filtry v Excelu byly přidány sloupce:

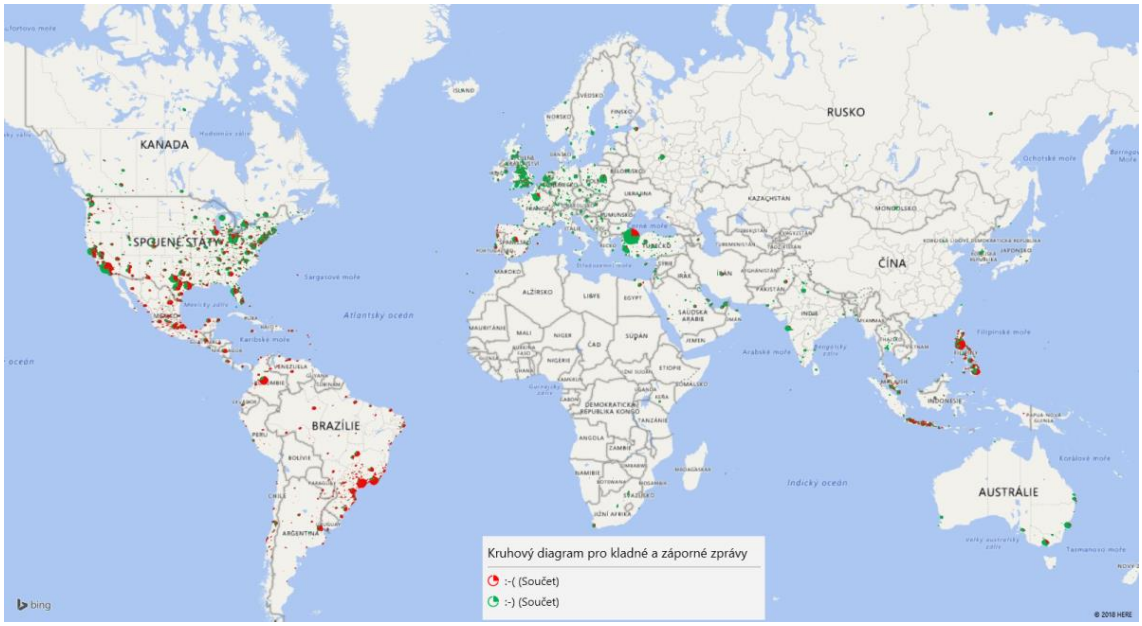
emoticon_id, FILTR, @, RT, :-), :-(, TWEET

Účel těchto sloupečků je filtrace dat. Opět se ze všech tweetů nepodařilo vyfiltrovat klíčová slova a takové tweety jsou označny jako „ostatní“. Excel soubor upravených dat je v elektronické podobě uložen na DVD-ROM v souboru příloh 11 „zdrojova data nalada verejnosti“.

2.4.3 Vizualizace a prostorová analýza

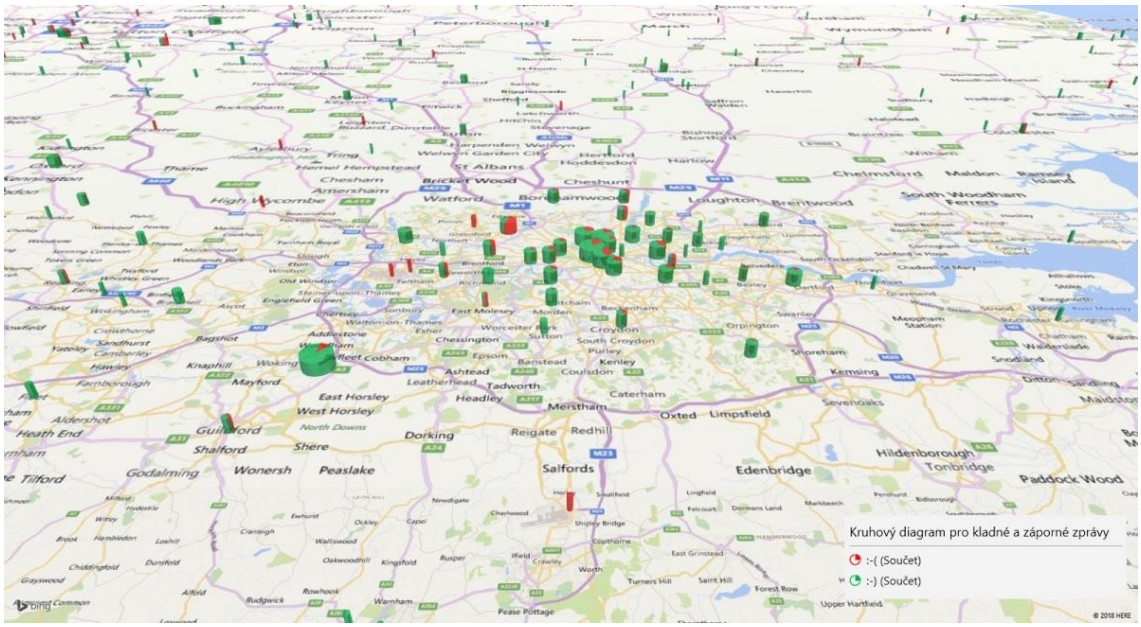
Na dalším listu s názvem „kontingencni_tabulka“ je vytvořená kontingenční tabulka, která sčítá všechny zprávy s usmívajícími se a mračícími se smajlíky ke všem unikátním lokalitám. Zprávy, kde nebylo možno vyhledat emotikony jsou v tabulce označeny jako „ostatní“. V 9 062 lokalitách bylo celkem 38 241 tweetů z nichž 18 944 tweetů (49.5 %) obsahovalo smějícího se smajlíka a 12 784 tweetů (33,5 %) obsahovalo mračícího se smajlíka. U 6 513 tweetů (17 %) nebylo možné smajlíka určit.

Po spuštění nástroje „3D mapa“ z menu „Vložení“ v aplikaci Excel, byly zvoleny data z listu „kontingnecni_tabulka“ a pomocného listu „top_mesta“. První prostorová vizualizace byla zvolena s kruhovým diagramem se strukturou rozdělenou na zprávy obsahující pozitivního (usmívající se) a negativního (mračící se) smajlíka. Zelenou barvou byla označena část výseče usmívajícího se smajlíka a červená barva znázorňuje mračícího se smajlíka. Výsledné mapy s kruhovým diagramem jsou znázorněny na obrazku 10-12.

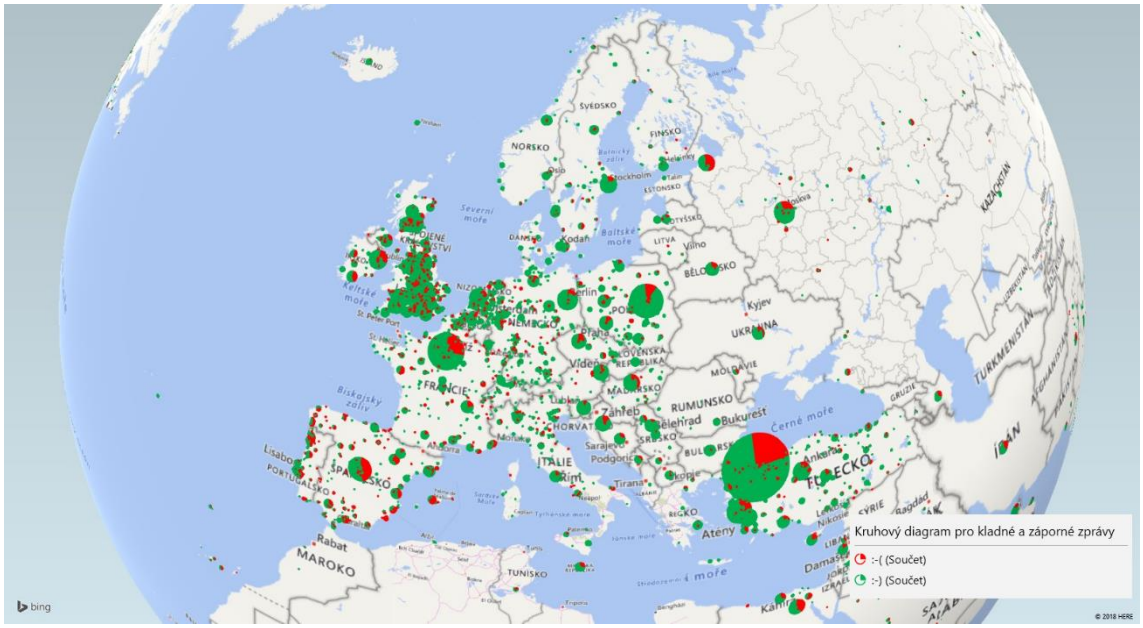


Obr. 10 – Mapa všech tweetů na mapě celého světa[41]

[41] Zdroj vlastní



Obr. 11 – Mapa tweetů z lokalit v okolí Londýna[42]



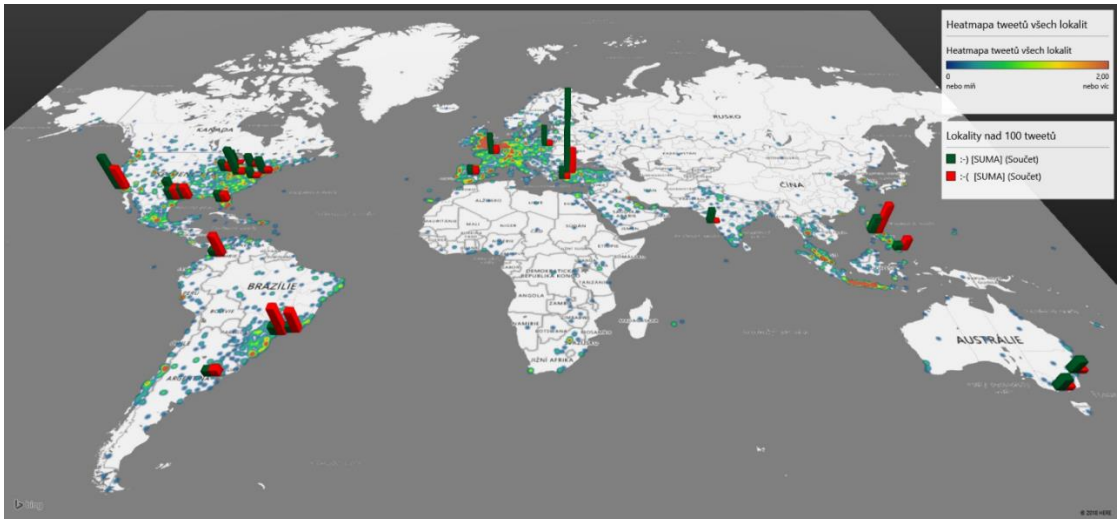
Obr.12 – Detail světové mapy tweetů v Evropě[43]

[42] Zdroj vlastní

[43] Zdroj vlastní

2.4.4 Výsledky

Kombinace heatmapy všech tweetů na světě spojená se skupinovým sloupcovým grafem na obrázku 13 znázorňuje výsledky této případové studie. Sloupcový graf zobrazuje zprávy s usmívajícím se a mračícím se smajlíkem z lokalit s více než 100 tweety. Zdrojem dat pro sloupcový graf je „Tabulka lokalit s více než 100 tweety“ v příloze 2 (str. 51) vycházející z tabulky v listu „top_mesta“.



Obr.13 – Výsledky prostorové analýzy ve vizualizaci[44]

Celkem bylo zvizualizováno 9 062 lokalit z celkového počtu 38 241 geokódovaných tweetů. Největší počet tweetů byl soustředěn na velká města. Lokalitou s nejvíce smajlíky bylo město Instanbul, kde bylo ve všech zprávách se smajlíky 66 % usmívajích smajlíků a 20 % mračících se smajlíků. Oblastí z lokalit, kde bylo získáno více než 100 tweetů s největším procentem usmívajících se smajlíků byl stát Ohio ve Spojených státech amerických (82 %). Nejvíce mračících se smajlíků obsahovaly zprávy ve městě Bogotá v Kolumbii (72 %) a v Brazíli ve městech Rio de Janeiro (65 %) a Sao Paulo (68 %). Mezi lokality s velkým zastoupením negativních smajlíků patří i ostrovní stát Filipíny.

[44] Zdroj vlastní

DISKUZE

V bakalářské práci jsme shromáždili, vyfiltrovali a provedli prostorovou analýzu dat pomocí sofistikovaných webových a desktopových nástrojů pro vizualizaci dat.

Od zavedení soc. sítě Twitter v roce 2006 se zdokonalily jeho sofistikované mapovací a vyhledávací funkce a v současné době umožňuje svým uživatelům rychle sdílet informace, názory, diskuze. Kromě toho Twitter také sděluje informace o poloze a tím se stává předmětem studia pro rozhodování s ohledem na sociální chování, trendy a predikce.

V první případové studii jsme zjišťovali a následně prostorově vymezili místa, která mají vysokou tweetovou aktivitu s výskytem názvů dvou obchodních společností jako klíčových slov „amazon" a „ebay". Na první pohled je tak patrné, že největší množství tweetů o obchodních společnostech Amazon a eBay bylo posláno z lokality Los Angeles, California a v dalším pořadí se pak objevují města z oblasti Spojených států amerických. Překvapivě se mezi prvních deset lokalit s nejvyšším počtem tweetů zařadila i lokalita v provincii Šan-tung v Čínské lidové republice s výskytem 31 tweetů. Domnívali jsme se, že se může jednat o více příspěvků jednoho konkrétního uživatele, což se po vyhledání názvu této lokality a kontrole všech příspěvků potvrdilo ve zdrojových datech. Z vytvořených map vyplývá, že obecně z Číny nejsou tweety téměř žádné, zjevně z důvodu omezení přístupu Číny od zbytku celosvětových služeb velkým čínským firewallem.

I naše výsledky ukazují, že Twitter, jako slibný zdroj dat včetně zachycení množství geografických dat z každodenního života, představuje zlomek společenských událostí v geografickém prostoru a může mít předvídatelné i nepředvídatelné nedostatky například ve smyslu důvěryhodnosti dat právě s ohledem na existující omezení přístupu k těmto službám.[45] Také MacEachren uvádí, že většina uživatelů nemá geolokaci tweetů prostřednictvím GPS povolenou. Pouze jednotky uživatelů mají GPS lokaci zapnutou.[46]

Oussalah, Bhaf, Challis aj. uvádějí rostoucí počet služeb vznikajících za účelem poskytování vyhledávání, shromažďování a zpracovávání dat, proto navrhli a implementovali softwarový systém, který umožňuje shromažďovat velké množství

[45] Srov. WESTERHOLT, R., STEIGER, E., RESCH, B., aj., Abundant Topological Outliers in Social Media Data and Their Effect on Spatial Analysis, *PLoS One*, 2016, roč. 11, č. 9, s. 26-27, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162360>.

[46] MACEACHREN, A., aj., *Geo-twitter analytisc: Applications in crisis management*, s. 4.

geografických dat.[47] Bao, Liu, Yu aj. uvádějí, že stále více lidí z různých věkových skupin a zázemím se stávají novými uživateli Twitteru zejména v městských oblastech a ve svém výzkumu se zaměřili na aktivity lidí související s možností předpovědi výskytu nehod a havárií, kdy nakupování uvedli jako jednu z aktivit ovlivňující počet dopravních nehod.[48] Prostřednictvím analýzy dat ze sociálních médií se autoři Steiger, Westerholt, Resch aj. zaměřili na odraz určitého kolektivního chování člověka v tweetu dle příslušného prostorového místa a dokázali v nich odhalit významné clustery související se sociální aktivitou lidí. Výzkum podpořil, že prostřednictvím dat ze sociálních sítí lze odhadnout soukromé, ale i kolektivní body zájmů uživatelů a výsledky mohou být využity v chápání nejen sociální dynamiky, ale i pro prevenci nežádoucího chování. [49]

Výsledky naší první případové studie porovnávají dvě velké obchodní společnosti z pohledu četnosti výskytu jejich názvu na Twitteru, kdy se potvrdilo, že největší výskyt tweetů s názvem společností je v lokalitách velkých měst. Toto srovnání může sloužit jako podklad k posouzení zastoupení těchto společností v různých lokalitách a jeví se jako vhodný podklad pro marketingové účely.

V naší druhé případové studii jsme se zaměřili na sledování nálady sdělení – tweetu v souvislosti s výskytem pozitivních a negativních smajlíků (emotikonů) v tweetu a zjistili jsme, že se jejich množství liší dle prostorového umístění. Největší procento výskytu pozitivních emotikonů jsme registrovali ve státě Ohio v USA. Nejvíce negativních emotikonů bylo zaznamenáno ve městě Bogotá v Kolumbii. Můžeme z toho usuzovat, že lze detekovat tweety ukazující na určité kolektivní prožívání reality. Problémem zůstává, že i pozitivní nálada jednotlivce může být hodnocena symbolem - emotikonem, který další jedinec chápe jako symbol negativního prožitku.[50] Také

[47] Srov. OUSSALAH, M., BHAT, F., CHALLIS, K., aj., A software architecture for Twitter collection, search and geolocation services, *Knowledge-Based Systems*, 2013, roč. 37, s. 105, <https://www.sciencedirect.com/science/article/pii/S0950705112002055>.

[48] Srov. BAO, J., LIU, P., YU, H., aj., Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas, *Accident Analysis and Prevention Journal,* 2017, č. 106, s. 358-369, <https://doi:10.1016/j.aap.2017.06.012>.

[49] Srov. STEIGER, E., WESTERHOLT, R., RESCH, B., aj., Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data, *Computers, Environment and Urban Systems*, 2015, roč. 54, s. 259-264, <https://www.sciencedirect.com/science/article/pii/S0198971515300181>.

[50] Srov. STERNE, J., *Měříme a optimalizujeme marketing na sociálních sítích*, 116-124.

Kounadi, Lampoltshammer, Groff aj. zkoumali tweety vyjadřující obavy a názory související s úrovní strachu z kriminality a dokázali, že vliv informací o zločinech silně závisí na lokalitě, mluveném jazyku a incidentech, ale také s nerovným přístupem Twitteru mezi různými zeměmi.[51] Z toho plyne, že uživatelé Twitteru nemusí reprezentovat názory, nálady a problémy celé společnosti.

[51] Srov. KOUNADI, O., LAMPOLTSHAMMER, T., J., GROFF, E., aj., Exploring Twitter to Analyze the Public's Reaction Patterns to Recently Reported Homicides in London, *PLoS One,* 2015, roč. 10, č. 3, s. 1-17, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4374728/>.

ZÁVĚR

Sociální sítě mají potenciál přinášet nové cenné poznatky pro marketingové agentury a další organizace, které se zajímají o pochopení chování lidí a sledování online sociálních trendů.

Tato bakalářská práce se věnuje sociální sítí Twitter a práci s prostorovými informacemi generovanými na této síti jejími uživateli.

Bakalářská práce nejdříve teoreticky popisuje sociální síť Twitter a využití dat z této sítě pro prostorovou analýzu. Byl vybrán vhodný způsob stahování dat prostřednictvím Twitter API pomocí skriptu v programovacím jazyku Python. Praktická část práce definuje klíčová slova pro dvě případové studie. Byla stažena data s prostorovými údaji. Získané soubory dat byly dále filtrovány a upraveny za účelem jejich importu do aplikací Google Fusion Tables a pluginu MS Excel 3D Mapy, kterými byly vizualizovány v prostorových mapách. Výsledky těchto vizualizací sloužily pro analýzu dat v prostoru. Na základě těchto výsledků, byly zjištěny lokality, kde se tweety definované klíčovými slovy vyskytují nejčastěji.

V první případové studii „obchodní společnosti“ se více jak v polovině tweetů (55,9 %) objevilo klíčové slovo „amazon“, pouze 4 tweety (0,1 %) zmiňují „amazon+ebay“ v jedné zprávě, ale skoro u 30 % stažených tweetů nebylo možné určit, zda se v nich klíčová slova vyskytují. Podle heatmapy byla největší koncentrace tweetů v Severní Americe, Evropě a Japonsku. Nejmenší koncentrace tweetů se objevovala v Jižní Africe, Rusku, střední Asii a Kanadě.

V druhé případové studii „nálada veřejnosti“ byl největší počet tweetů se sledovanými klíčovými slovy (emotikomy) soustředěn na velká města. Konkrétní lokalitou s nejvíce smajlíky bylo město Instanbul, kde ve všech zprávách bylo 66 % usmívajích se smajlíků a 20 % mračících se smajlíků. V lokalitách, z nichž bylo získáno více než 100 tweetů, bylo více jak 80 % usmívajících se smajlíků ve státě Ohio v USA a přes 70 % mračících se smajlíků obsahovaly zprávy ve městě Bogotá v Kolumbii.

V diskuzi byly zhodnoceny možnosti využití dat ze sociální sítě Twitter. Několik studií demonstruje užitečnost a platnost geoprostorových dat z Twitteru. Předpokládáme, že i tato práce by mohla motivovat další výzkum způsobu využití geodat s ohledem na existující omezení při analýze těchto dat. V současné době je poměrně obtížnou otázkou, jak měřit význam slov. V budoucnosti by tento problém měla řešit automatizace

textových analytických nástrojů. Správná definice klíčových slov z pohledu jejich významu a celkového kontextu zprávy určí přesnost a úspěšnost provedení nejen prostorové analýzy zpráv na sociální síti Twitter.

Přestože data ze sociálních sítí mohou vykazovat různá omezení, poskytují nové informace, které pomáhají lépe rozumět aktivitám a chování lidí. Údaje získané ze sociálních sítí poskytují cenný zdroj informací pro pochopení sentimentu jedince nebo kolektivní veřejné nálady.

ANOTACE

Příjmení a jméno autora:	Tomáš Daniel
Instituce:	Moravská vysoká škola Olomouc

Název práce v českém jazyce:	Využití dat Twitteru pro prostorové analýzy
Název práce v anglickém jazyce:	Usage of Twitter data for spatial analysis

Vedoucí práce:	Mgr. Vít Pászto, Ph.D.
Počet stran:	56
Počet příloh:	12
Rok obhajoby:	2018
Klíčová slova v českém jazyce:	využití, data, sociální, sítě, Twitter, prostorová, analýza
Klíčová slova v anglickém jazyce:	usage, data, social, network, Twitter, spatial, analysis

Sociální síť Twitter je cenným zdrojem informací o postojích, chování a aktivitách uživatelů. Geotagované tweety mohou poskytnout vstupní data prostorové anýlyzy. Cílem této práce je najít způsob dolování dat z Twitteru, která zahrnují specifická klíčová slova ze dvou různých případových studií. Výsledky budou vizualizovány v prostorových mapách, které slouží jako výstup prostorové analýzy.

The social network Twitter is a valuable resource of information about user attitudes, behavior and activities. Tweets with geographical information can provide input data for spatial analysis. The aim of this thesis is to find out how to process gathered data from Twitter, containing specific keywords, from two different case studies. The results will be visualized in spatial maps, which serve as an output of spatial analysis.

LITERATURA A PRAMENY

APIGEE, *Twitter API console* [online]. [cit. 2018-02-18]. Dostupné na WWW: <https://apigee.com/console/twitter>.

BAO, Jie, LIU, Pan, YU, Hao, aj. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. *Accident Analysis and Prevention* [online]. September 2017, č. 106 [cit. 2018-02-20], s. 358-369. Dostupné na WWW: <https://doi:10.1016/j.aap.2017.06.012>.

GITHUB, *abraham/twitteroauth* [online]. © 2018 [cit. 2018-01-05]. Dostupné na WWW: <https://github.com/abraham/twitteroauth>.

GITHUB, *twitter-streamer* [online]. © 2018 [cit. 2018-01-05]. Dostupné na WWW: <https://github.com/pschiffe/twitter-streamer>.

GOOGLE, *Google Fusion Tables, Amazon a eBay klíčová slova pro prostorovou analýzu* [online]. [cit. 2018-03-18]. Dostupné na WWW: <https://www.google.com/fusiontables/DataSource?docid=1dHzwbmJ8IexcV0P4oLPVsajbSJ0SHe6X2nLmOQf0>.

GOOGLE, *Pricing and Plans* [online]. [cit. 2018-03-10]. Dostupné na WWW: <https://developers.google.com/maps/pricing-and-plans/#details>.

GOOGLE, *Type and size of files to import* [online]. [cit. 2018-03-18]. Dostupné na WWW: <https://support.google.com/fusiontables/answer/171181>.

KOCICH, David, a HORÁK, Jiří. Twitter as a source of big spatial data. In *Proceedings of the International Multidisciplinary Scientific GeoConference SGEM* [online]. September 2016, sv. 2, č. 1 [cit. 2018-03-22], s. 921-928. Dostupné na WWW: <https://sgemworld.at/sgemlib/spip.php?article8546>.

KOUNADI, Ourania, LAMPOLTSHAMMER, Thomas J., GROFF, Elizabeth, aj. Exploring Twitter to Analyze the Public's Reaction Patterns to Recently Reported Homicides in London. *PLoS One* [online]. March 2015, roč. 10, č. 3 [cit. 2018-03-22], s. 1-17. Dostupné na WWW: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4374728/>.

KWAK, Haewoon., aj. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* [online]. 2010 [cit. 2018-02-21], s. 591-600. Dostupné na WWW: <http://www.ambuehler.ethz.ch/CDstore/www2010/www/p591.pdf>.

MACEACHREN, Alan M., aj. Geo-twitter analytics: Applications in crisis management. In: *25th International Cartographic Conference* [online]. 2011 [cit. 2018-02-20], s. 1-8. Dostupné na WWW: <https://www.geovista.psu.edu/publications/2011/MacEachren_ICC_2011.pdf>.

NOVOTNÁ, Marie, ČECHUROVÁ, Monika, a BOUDA, Jakub. *Geografické informační systémy ve školách*. 1. vyd. Plzeň: Aleš Čeněk, 2012. 154 s. ISBN 978-80-7380-385-8.

OUSSALAH, Mourad, BHAT, F., CHALLIS, Keith, D., aj. A software architecture for Twitter collection, search and geolocation services. *Knowledge-Based Systems* [online]. January 2013, roč. 37 [cit. 2018-03-22], s. 105-120. Dostupné na WWW: <https://www.sciencedirect.com/science/article/pii/S0950705112002055>.

PÁSZTO, Vít. *Prostorová informace a vybrané metody geocomputation pro její hodnocení.* 1. vyd. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2015. 156 s. ISBN 978-80-244-4821-3.

PYTHON, *Welcome to Python* [online]. © 2001-2018 [cit. 2018-01-20]. Dostupné na WWW: <https://www.python.org/>.

RAPANT, Petr. *Geoinformační technologie*. 2. vyd. Ostrava: Vysoká škola báňská – Technická univerzita Ostrava, Hornicko-geologická fakulta, Institut geoinformatiky, 2006. 102 s. ISBN 80-248-1263-0.

STEIGER, Enrico, WESTERHOLT, Rene, RESCH, Bernd, aj. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems* [online]. November 2015, roč. 54 [cit. 2018-03-22], s. 255-265. Dostupné na WWW: <https://www.sciencedirect.com/science/article/pii/S0198971515300181>.

STERNE, Jim. *Měříme a optimalizujeme marketing na sociálních sítí.* 1. vyd. Brno: Computer Press, 2011. 280 s. ISBN 978-80-251-3340-8.

TECHIELLA, *Twitter Search using the Twitter API v1.1 & PHP* [online]. [cit. 2017-10-08]. Dostupné na WWW: <http://techiella.x0.com/twitter-search-using-the-twitter-api-php/>.

TWEEPY, *Tweepy Documentation* [online]. © 2018 [cit. 2018-01-07]. Dostupné na WWW: <http://tweepy.readthedocs.io/en/v3.5.0/>.

TWITTER, *API documentation* [online]. © 2017 [cit. 2017-09-22]. Dostupné na WWW: <https://dev.twitter.com/>.

TWITTER, *Developer terms* [online]. © 2018 [cit. 2018-02-24] Dostupné na WWW: <https://developer.twitter.com/en/developer-terms>.

TWITTER, *Twitter Developers – Docs* [online]. © 2017 [cit. 2017-12-27]. Dostupné na WWW: <https://developer.twitter.com/en/docs>.

WESTERHOLT, Rene, STEIGER, Enrico, RESCH, Bernd, aj. Abundant Topological Outliers in Social Media Data and Their Effect on Spatial Analysis. *PLoS One* [online]. September 2016, roč. 11, č. 9 [cit. 2018-03-22], s. 1-31. Dostupné na WWW: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162360>.

YUE, Li, QINGHUA, Li, a JIE, Shan. Discover Patterns and Mobility of Twitter Users – A Study of Four US College Cities. *ISPRS International Journal Of Geo-Information* [online]. February 2017, roč. 6, č. 2 [cit. 2018-03-22], s. 1-17. Dostupné na WWW: <http://www.mdpi.com/2220-9964/6/2/42> .

SEZNAM ZKRATEK

API	Application Programming Interface
CSV	Comma-separated values
GPS	Global Position Systém
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol)
ID	unikátní identifikační číslo
MIT	Massachusetts Institute of Technology
MS	Microsoft
SDK	Software Development Kit
STDOUT	Standard input
TSV	Tab-Separated Values
TXT	Textový soubor
URL	Uniform Resource Locator
USA	Spojené státy americké
WWW	World wide web

SEZNAM OBRÁZKŮ

SEZNAM TABULEK A GRAFŮ

SEZNAM PŘÍLOH

Vázané přílohy:

Příl. 1 – Mapa struktury tweetu

Příl. 2 –Tabulka lokalit s více než 100 tweety

Příl. 3 – Zdrojový kód twitter-streamer

Volné přílohy:

Příl. 4 – DVD-ROM (volná příloha)

Přílohy na DVD-ROM

Příl. 5 – Poster

Příl. 6 – program twitter-search

Příl. 7 – program twitter-streamer

Příl. 8 – soubor map obchodni spolecnosti

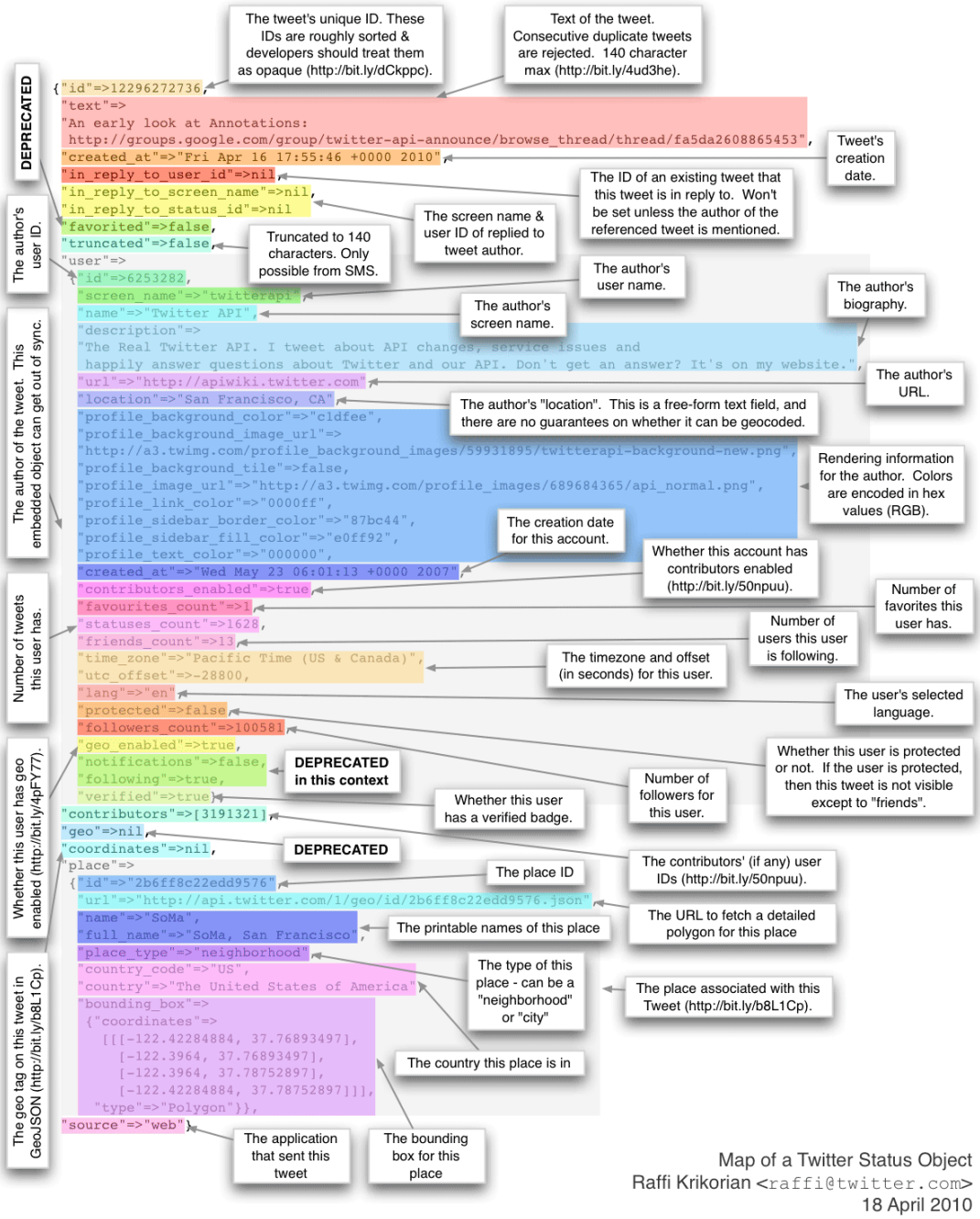
Příl. 9 – soubor map nalada verejnosti

Příl. 10 – zdrojova data obchodni spolecnosti

Příl. 11 – zdrojova data nalada verejnosti

Příl. 12 – odkaz na google fusion tabulku

Příl. 1 - Mapa struktury tweetu



https://www.scribd.com/doc/30146338/map-of-a-tweet

Příl. 2 - Tabulka lokalit s více než 100 tweety

Tabulka lokalit s více než 100 tweety					
Poř.	Místo	:-) [SUMA]	:-([SUMA]	:-) [%]	:-([%]
1	İstanbul, Türkiye	554	166	66	20
2	Los Angeles, CA	226	151	51	34
3	Manila City, National Capital Region	74	190	24	63
4	Quezon City, National Capital Region	106	155	35	51
5	São Paulo, Brasil	56	181	21	68
6	Paris, France	137	47	63	21
7	Rio de Janeiro, Brasil	42	142	19	65
8	Ohio, USA	179	28	83	13
9	Texas, USA	104	59	50	28
10	Warszawa, Polska	133	18	65	9
11	Bogotá, D.C., Colombia	40	139	21	72
12	Houston, TX	76	86	40	45
13	Makati City, National Capital Region	64	87	37	50
14	San Antonio, TX	69	76	40	44
15	San Francisco, CA	117	28	68	16
16	Manhattan, NY	105	33	63	20
17	Mumbai, India	87	13	61	9
18	Florida, USA	62	54	44	38
19	Toronto, Ontario	80	30	57	21
20	Chicago, IL	65	46	48	34
21	Davao City, Davao Region	39	79	29	59
22	Sydney, New South Wales	81	22	62	17
23	Melbourne, Victoria	76	28	60	22
24	Michigan, USA	85	9	67	7
25	İzmir, Türkiye	77	24	69	22
26	Madrid, España	52	40	47	36
27	Austin, TX	60	36	56	33
28	Washington, DC	59	21	58	21
29	Ciudad Autónoma de Buenos Aires, Argentina	38	45	38	45

Příl. 3 – Zdrojový kód twitter-streamer

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

import sys
import json
import tweepy
import configparser

# override tweepy.StreamListener to add logic to on_status
class MyStreamListener(tweepy.StreamListener):
 _tweets_csv_path = None

 def __init__(self, api=None, tweets_csv_path=None):
 super().__init__(api=api)
 self._tweets_csv_path = tweets_csv_path

 def on_status(self, status):
 # store tweets in a csv file and print them to stdout
 with open(self._tweets_csv_path, mode='a') as csv:
 # https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object
 try:
 tid = status.id_str
 except AttributeError:
 tid = ''
 try:
```

Příl. 3 – pokračování

```
created_at = status.created_at

except AttributeError:

created_at = ''

try:

user_name = status.user.screen_name.replace('"', '""')

except AttributeError:

user_name = ''

try:

user_place = status.user.location.replace('"', '""').replace('\n', ' || ').replace('\r', '')

except AttributeError:

user_place = ''

try:

place = status.place.full_name.replace('"', '""')

place_obj = self.api.geo_id(status.place.id)

centroid = ','.join(reversed(list(map(str, place_obj.centroid))))

except AttributeError:

return

try:

coord = ','.join(reversed(list(map(str, status.coordinates.coordinates))))

except AttributeError:

coord = ''

try:

polygon0 = ','.join(reversed(list(map(str,
status.place.bounding_box.coordinates[0][0]))))

polygon1 = ','.join(reversed(list(map(str,
status.place.bounding_box.coordinates[0][1]))))
```

Příl. 3 – pokračování

```
 polygon2 = ','.join(reversed(list(map(str,
status.place.bounding_box.coordinates[0][2]))))

 polygon3 = ','.join(reversed(list(map(str,
status.place.bounding_box.coordinates[0][3]))))

 except AttributeError:

 polygon0 = ''

 polygon1 = ''

 polygon2 = ''

 polygon3 = ''

 try:

 text = status.extended_tweet['full_text']

 for url in status.extended_tweet.entities['urls']:

 text = text.replace(url['url'], url['expanded_url'])

 except AttributeError:

 try:

 text = 'RT @{0}: {1}'.format(status.retweeted_status.user.screen_name,
status.retweeted_status.extended_tweet['full_text'])

 for url in status.retweeted_status.extended_tweet.entities['urls']:

 text = text.replace(url['url'], url['expanded_url'])

 except AttributeError:

 try:

 text = 'RT @{0}: {1}'.format(status.retweeted_status.user.screen_name,
status.retweeted_status.text)

 for url in status.retweeted_status.entities['urls']:

 text = text.replace(url['url'], url['expanded_url'])

 except AttributeError:

 text = status.text if status.text else ''
```

Příl. 3 – pokračování

```
 for url in status.entities['urls']:

 text = text.replace(url['url'], url['expanded_url'])

 text = text.replace("", """).replace('\n', ' || ').replace('\r', '')

 csv_row =
'"{0}","{1}","{2}","{3}","{4}","{5}","{6}","{7}","{8}","{9}","{10}","{11}"\r\n'.form
at(tid, created_at, user_name, user_place, coord, centroid, place, polygon0, polygon1,
polygon2, polygon3, text)

 print(text)

 #print(csv_row)

 #print(json.dumps(status._json, indent=4) + '\n')

 csv.write(csv_row)



def main():

 config = configparser.ConfigParser()

 config.read('config.ini')

 auth = tweepy.OAuthHandler(config['DEFAULT']['consumer_key'],
config['DEFAULT']['consumer_secret'])

 auth.set_access_token(config['DEFAULT']['access_token'],
config['DEFAULT']['access_token_secret'])



 api = tweepy.API(auth_handler=auth, wait_on_rate_limit=True,
wait_on_rate_limit_notify=True)



 myStreamListener = MyStreamListener(api=api,
tweets_csv_path=config['DEFAULT']['tweets_csv_path'])

 myStream = tweepy.Stream(auth=api.auth, listener=myStreamListener)



 try:
```

Příl. 3 – pokračování

```
 # parameters: https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html

 myStream.filter(track=config['DEFAULT']['filter_track'].split(','), languages=config['DEFAULT']['filter_lang'].split(','))

 except KeyboardInterrupt:

 sys.exit()




if __name__ == '__main__':

 main()
```