

**Univerzita Palackého v Olomouci**  
**Přírodovědecká fakulta**  
**Katedra geoinformatiky**

**Dana FUSKOVÁ**

**GEOINFORMACE V PROSTŘEDÍ  
SOCIÁLNÍCH SÍTÍ**

**Bakalářská práce**

**Vedoucí práce: Mgr. Lukáš MAREK**

**Olomouc 2014**

## **Čestné prohlášení**

Prohlašuji, že jsem bakalářskou práci bakalářského studia oboru Geoinformatika a geografie vypracovala samostatně pod vedením Mgr. Lukáše Marka.

Všechny použité materiály a zdroje jsou citovány s ohledem na vědeckou etiku, autorská práva a zákony na ochranu duševního vlastnictví.

Všechna poskytnutá i vytvořená digitální data nebudu bez souhlasu školy poskytovat.

V Olomouci 19. srpna 2014

---

Děkuji vedoucímu práce Mgr. Lukáši Markovi za podněty a připomínky při vypracování práce. Dále děkuji konzultantovi Ondřeji Růžičkovi za cenné rady.

Vložený originál **zadání** bakalářské práce (s podpisy vedoucího katedry, vedoucího práce a razítkem katedry). Ve druhém výtisku práce je vevázána fotokopie zadání.

# OBSAH

SEZNAM POUŽITÝCH ZKRATEK .....	7
ÚVOD .....	8
<b>1 CÍLE PRÁCE.....</b>	<b>9</b>
<b>2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ .....</b>	<b>10</b>
2.1 Postup zpracování .....	10
2.2 Vizualizace .....	12
2.3 Použitý software .....	13
<b>3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY .....</b>	<b>14</b>
<b>4 SOCIÁLNÍ SÍŤ A JEJÍ DEFINICE .....</b>	<b>15</b>
4.1 Sociální síť, definice, historie .....	15
4.2 Prostorová data .....	16
4.3 Vybrané sociální síť .....	18
4.3.1 Twitter.....	18
4.3.2 Facebook.....	21
4.3.3 LinkedIn.....	26
<b>5 DATA MINING .....</b>	<b>29</b>
5.1 Dolování geografických dat ze sociálních sítí .....	30
<b>6 VIZUALIZACE DAT ZE SOCIÁLNÍCH SÍTÍ .....</b>	<b>31</b>
6.1 Možnosti vizualizace.....	31
6.1.1 ArcGIS .....	31
6.1.2 Prostředí a jazyk R.....	33
6.1.3 Placebook.....	34
6.1.4 GeoSocialApp aggregator.....	35
6.2 Geokódování .....	36
<b>7 DOLOVÁNÍ DAT ZE SOCIÁLNÍ SÍŤE TWITTER .....</b>	<b>37</b>
7.1 Twitter a prostředí R .....	37
7.1.1 Zisk dat Search API .....	37
7.1.2 Zisk dat Streaming API .....	39
7.2 Krize na Ukrajině a Twitter.....	42
7.3 Vizualizace dat o Ukrajině .....	43
7.4 Výsledky případových studií.....	44

<b>8 DOLOVÁNÍ DAT ZE SOCIÁLNÍ SÍTĚ FACEBOOK.....</b>	<b>45</b>
8.1 Facebook a prostředí R.....	45
8.1.1 Zisk informací o přátelích.....	46
8.1.2 Zisk veřejných příspěvků s prostorovým atributem .....	47
8.2 Vizualizace dat ze sociální sítě Facebook .....	48
8.3 Výsledky .....	50
<b>9 VÝSLEDKY .....</b>	<b>51</b>
<b>10 DISKUZE .....</b>	<b>53</b>
<b>11 ZÁVĚR .....</b>	<b>55</b>
<b>LITERATURA .....</b>	<b>57</b>
<b>SUMMARY .....</b>	<b>63</b>
<b>PŘÍLOHY.....</b>	<b>64</b>

## SEZNAM POUŽITÝCH ZKRATEK

<b>Zkratka</b>	<b>Význam</b>
API	Application Programming Interface
CSS	Cascading Style Sheets
CSV	Comma Separated Values
DW	Data Warehouse
FQL	Facebook Query Language
GeoJSON	Geographic JavaScript Object Notation
GIS	Geografický informační systém
GPS	Global Position System
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
LBSN	Location Based Social Network
OLAP	Online Analytical processing
REST	Representational State Transfer
SQL	Structured Query Language
XML	Extensible Markup Language

## ÚVOD

Díky rozvoji a rozšíření webu a internetových technologií se webové sociální sítě staly populárním tématem současnosti. Díky nim mezi sebou mohou denně komunikovat miliony uživatelů a to nejen prostřednictvím zpráv, ale také sdílením fotografií, videí, odkazů a v neposlední řadě i polohy. Hlavním důvodem ke kontaktu osob v sociálních sítích je pak většinou společný zájem mezi uživateli.

Významnou roli v rozvoji sociálních sítí, umožňujících sdílet také pozici uživatele či zprávy, umožnil zejména nástup přenosných zařízení, která poskytují možnost rychlé geolokace (s různou přesností) ať už pomocí GPS, Wi-Fi či přímo z údajů mobilní sítě. Takto dochází ke vzniku velkých databází, které obsahují také prostorové informace a mají obrovský potenciál pro další vizualizace či analýzy.

Prostorová data v prostředí sociálních sítí jsou uložena a zpracovávána hned několika způsoby. Kromě možnosti automatické geolokace, tedy přidání polohy daného uživatele přímo k příspěvku na sociální síti, jsou u sociálních sítí významné také prostorové atributy, které vyplňují uživatelé sami. Pro případné analýzy se tak nabízí velké množství různých typů dat, ať už přímo o uživateli nebo o jejich aktivitě. Stejně jako existuje velké množství rozdílných typů dat, existuje i mnoho metod k jejich získání. Na získání (geo)dat z prostředí sociálních sítí přímo navazuje i vizualizace prostorových informací v datech obsažených. Získané informace jde dále zobrazovat v podobě (webových) map či grafů, které jsou obzvláště vhodné k vizualizaci přímého vztahu mezi uživateli.



# 1 CÍLE PRÁCE

Cílem bakalářské práce je vypracovat pojednání o možnostech získávání a vizualizace prostorových dat z prostředí sociálních sítí a medií.

Úvodní část se věnuje rešeršní části, která bude zahrnovat zejména popis sociálních sítí a typy prostorových dat, které je z nich možné získat. Teoretická část bude kromě toho také obsahovat popis možností data miningu (dolování dat) prostorových informací ze sociálních sítí a způsoby vizualizace získaných dat.

Cílem praktické části práce je provedení případových studií na datech z vybraných sociálních sítí, ve kterých se projeví zkušenosti nabyté v teoretické části práce a to zejména získávání dat, jejich úprava, zpracování, případně provedení geokódování. Výsledkem bude vytvoření webových map ze získaných dat (Příloha 14, 15, 16).

## 2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ

Následující kapitola popisuje postup a metody, které byly použity při zpracování této bakalářské práce a které jsou seřazeny chronologicky tak, aby přiblížily celý harmonogram práce.

### 2.1 Postup zpracování

Základem této práce bylo studium odborné literatury a jiných dostupných zdrojů. Na základě nastudovaných znalostí byly vybrány vhodné sociální sítě, na kterých pak probíhaly případové studie.

U těchto sociálních sítí bylo potřeba osvojit si nejen samotné fungování sítě, ale také jejich vazby s prostorovými prvky. Po ovládnutí funkčnosti bylo třeba se seznámit s jednotlivými API rozhraními a licenčními podmínkami, které jednotlivé sítě poskytují. Na základě nastudovaných informací byly zvoleny způsoby data miningu. Po vydolování dat byla data dále upravena tak, aby mohla být vizualizována.

#### Rešeršní činnost

Jak bylo uvedeno výše, začátkem práce byla rešeršní činnost, která směřovala na témata sociální sítě, dolování dat ze sociálních sítí a na možnosti vizualizace dat z těchto sítí.

#### Výběr sociálních sítí pro případové studie

Pro další podrobnější popis byly vybrány tři sociální sítě – Twitter, Facebook a LinkedIn. Motivací výběru sítě Twitter<sup>1</sup> byla jednak velká rozšířenost ve světě a také omezenost zprávy pouze na 140 znaků, díky čemuž se tzv. „tweet“ stává krátkou zprávou s přesnou vypovídací hodnotou. Facebook<sup>2</sup> je zde zahrnut jako nejrozšířenější sociální síť současnosti jak u nás, tak ve světovém měřítku (tab 2.1). LinkedIn<sup>3</sup> se dá nazvat profesní sociální sítí, ve které uživatelé píší profesní zkušenosti a dovednosti. Všechny tyto sociální sítě mají prostorové prvky a zisk dat z těchto prostředí může sloužit k dalšímu zpracování či vizualizaci.

#### Seznámení se s API

Všechny vybrané sociální sítě mají k dispozici rozhraní API, díky kterému lze dolovat informace. API (*Application Programming Interface*) označuje rozhraní pro programování aplikací. API je definováno na úrovni zdrojového kódu a poskytuje úroveň abstrakce mezi aplikací a jádrem s cílem zajistit přenositelnost kódu (Foldoc, 2014). Možnosti dolování dat pomocí API je u všech zvolených sociálních sítí omezeno licenčními podmínkami, kterými se musí řídit každý vývojář aplikací.

---

<sup>1</sup> [www.twitter.com](http://www.twitter.com)

<sup>2</sup> [www.facebook.com](http://www.facebook.com)

<sup>3</sup> [www.linkedin.com](http://www.linkedin.com)

Tab 2. 1 Přehled nejnavštěvovanějších sociálních sítí ve vybraných zemích v roce 2013 (Zdroj: Dvoran.com, 2014)

Země	1. místo	2. místo	3. místo	4. místo	5. místo
ČR	Facebook	YouTube	LinkedIn	Lidé	Twitter
Slovensko	Facebook	YouTube	LinkedIn	Tumblr	Twitter
Austrálie	Facebook	YouTube	LinkedIn	Twitter	Pinterest
Brazílie	Facebook	YouTube	LinkedIn	Twitter	Instagram
Čína	Qzone	Sina Weibo	Renren	Douban	PengYou
Egypt	Facebook	YouTube	Twitter	Vube.com	Ask.fm
Indie	Facebook	YouTube	LinkedIn	Twitter	Pinterest
Japonsko	YouTube	Facebook	Twitter	Mixi	Ameba
Jižní Afrika	Facebook	YouTube	LinkedIn	Twitter	Pinterest
Německo	Facebook	YouTube	Xing	LinkedIn	Tumblr
Nizozemsko	Facebook	YouTube	LinkedIn	Twitter	Tumblr
Polsko	Facebook	YouTube	LinkedIn	Tumblr	Vube.com
Rusko	Vkontakte	Odnoklassniki	Facebook	LiveInternet	LiveJournal
Španělsko	Facebook	YouTube	Twitter	LinkedIn	Tumblr
USA	Facebook	YouTube	LinkedIn	Twitter	Pinterest
Velká Británie	Facebook	YouTube	LinkedIn	Twitter	Tumblr

### Data mining

Data mining je do češtiny překládán jako dolování nebo vytěžování dat. Jde o využití specifických algoritmů pro extrakci vzorů z dat (Fayad, 1996). U sociálních sítí Twitter, Facebook, LinkedIn je nejrozumnějším řešením dolování dat pomocí jejich API, proto je tohoto využito i v této práci. U sociálních sítí Twitter a Facebook bylo využito prostředí jazyka R, kde byly použity již předem vytvořené knihovny/balíky pro práci se sociálními sítěmi. U LinkedIn byla data získána pomocí jazyka Python.

### JSON

Získaná data ze sociálních sítí byla uložena do formátu JSON (*JavaScript Object Notation*, česky *JavaScriptový objektový zápis*). JSON je odlehčený formát pro ukládání a výměnu dat. Je jednodušší alternativou k XML (*Extensible Markup Language*). Je snadno čitelný i zapisovatelný člověkem a dá se jednoduše analyzovat i generovat strojově. Je sice založen na syntaxi jazyka JavaScript, ale formát je pouze textový a může tak být čten a používán libovolným programovacím jazykem (JSON, 1999).

## GeoJSON

Pro snadnější vizualizaci dat do webových map bylo potřeba převést získaná data z formátu JSON do formátu GeoJSON, což je formát pro kódování geografických datových struktur. GeoJSON je vytvořen na bázi formátu JSON a může reprezentovat geometrii, funkci nebo soubor funkcí. Podporuje geometrie: Point, LineString, Polygons, Multipoint, MultiLineString a Multipolygon. Seznamy geometrií jsou reprezentovány v GeometryCollection. Seznamy funkcí jsou zastoupeny ve FeatureCollection (GEOJSON, 2008).

## Geokódování

Ačkoli sociální síť Twitter obsahuje přímo funkci pro geolokaci, díky které mohou uživatelé připojit svou polohu přímo k dané zprávě, tak pro vizualizaci dat nebyla data s možností geolokace získána v dostatečném množství, a proto bylo využito geokódování jednotlivých záznamů. Geokódování je služba, díky které je možné přiřadit zeměpisné souřadnice k zadanému textovému řetězci, nejčastěji adrese nebo názvu měst či jiných geografických prvků (API Mapy.cz, 2014). V případě této práce bylo využito geokódování pomocí Nominatim u OpenStreetMap<sup>4</sup>.

## 2.2 Vizualizace

Jedním z cílů této bakalářské práce je vizualizace vytěžených dat. Pro zobrazení všech vlastností dat byla zvolena metoda vytvoření webových interaktivních map s využitím JavaScriptu a knihoven Leaflet.

### JavaScript

JavaScript je všestranný, všudypřítomný programovací jazyk. Na začátku byl pouze doplňkem společnosti Netscape k jazyku VisualBasic, který vyrostl do jednoho z nejoblíbenějších programovacích jazyků na světě. Jedná se o skriptovací jazyk pro web. Má sice syntaktickou podobnost s jazykem Java, nicméně se jedná o samostatný programovací jazyk (PEHLIVANIAN, 2014).

### Leaflet

Leaflet je moderní open-source JavaScript knihovna pro tvoření mobilních interaktivních map. Má všechny funkce, které většina vývojářů potřebuje pro tvorbu on-line map. Je navržen tak, aby byl co nejjednodušší a mohl efektivně pracovat na všech hlavních počítačových a mobilních platformách - využívá HTML5 a CSS3. Má snadno použitelné a dobře zdokumentované API, díky kterému může uživatel jednoduše číst zdrojový kód (Leaflet, 2010-2014).

---

<sup>4</sup> <http://nominatim.openstreetmap.org/>

## 2.3 Použitý software

Kromě výše zmíněných popsaných programů a programovacích jazyků (JavaScript, Leaflet, aj.) bylo nutné pro získání a úpravu dat využít i dalších programovacích prostředí a jazyků.

### Prostředí a jazyk R

R je svobodný software a programovací jazyk pro statistické výpočty a grafiku. Jedná se o projekt GNU, který je implementací komerčního jazyka S. Nabízí širokou škálu statistických a grafických technik a díky volné licenci je velmi rozšířený. Používat se dá přímo z příkazového řádku, nicméně v této práci byl použit s grafickým rozhraním RStudio. Prostředí lze rozšířit pomocí jednotlivých knihoven, nazývaných *packages* a do češtiny překládaných jako „balíčky“ (The R Project, 2009).

V případě této práce byly využity zejména balíčky *twitterR*, *rFacebook*, *streamR*, *sqldf*, *httr*, *rjson*, *plyr*, *rgdal*, *tcltk*.

### Python

Python je dynamický objektově-orientovaný programovací jazyk, který je možné využít v mnoha oblastech. Jeho filosofie je založena na snadné čitelnosti kódu tak, aby se dal použít i v kombinaci s jinými programovacími jazyky. Je vyvíjen jako open source projekt. Python je využíván zejména jako skriptovací jazyk (Python.org, 2001-2014). K mnoha jeho pozitivním vlastnostem patří například jednoduchost z hlediska učení, proto byl tento jazyk použit pro dolování dat ze sociální sítě LinkedIn. Pro získání dat bylo třeba do jazyka Python nainportovat knihovnu *LinkedIn*.

### IPython Notebook

IPython Notebook je webové interaktivní výpočetní prostředí, kde je možné kombinovat psaní kódu, spouštění kódu, výpočty i grafické výstupy v rámci jednoho dokumentu. Umožňuje tak přístup k interaktivním dokumentům a kombinuje dvě základní složky: webovou aplikaci a textové dokumenty (IPython, 2014).

### 3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Sociální sítě jsou jedním z nejvíce skloňovaných témat dnešní doby v souvislosti s internetem. Z hlediska návštěvnosti na internetu předstihly dokonce i webové stránky s pornografií, které si dlouhodobě držely první místo (Kubelka, 2012). Proto i zájem o tuto problematiku roste. Cílem této práce jsou však prostorová data v sociálních sítích. Prostorové informace se staly významnou součástí hlavně s vývojem a rozšířením mobilních přístrojů s možností automatické geolokace. Geodata jsou tak momentálně součástí téměř všech sociálních sítí. Zisk těchto dat tak nabízí spoustu možností pro prostorové analýzy. Ve světě je o tomto tématu věnováno několik publikací. Teorií sítí a data miningem se zabývají Nasrullah a kol. (2010), vizualizací sociálních sítí například Nathan Yau (2011). Praktickou publikací o dolování dat prostřednictvím programovacího jazyka Python je kniha *Mining the Social Web* od autora Matthew A. Russell (2011).

V České republice se problematice sociálních sítí věnuje obor Studia nových médií na Filosofické fakultě Univerzity Karlovy v Praze. Zde je jednou ze zajímavých studií diplomová práce Martina Kubelky, která se věnuje data miningu sociálních sítí (Kubelka, 2012). Praktickým příkladem zisku dat a aplikace těchto dat nejen na prostorové analýzy je diplomová práce Tomáše Gregoroviče (2010) zaměřená na sociální síť Couchsurfing<sup>5</sup>. Na katedře geoinformatiky přírodovědecké fakulty Univerzity Palackého v Olomouci vznikla v roce 2013 bakalářská práce zaměřená na pohyb obyvatel prostřednictvím nejvýznamnější světové geolokační sítě Foursquare<sup>6</sup> (Jarcovják, 2013).

---

<sup>5</sup> [www.coursurfing.com](http://www.coursurfing.com)

<sup>6</sup> [www.foursquare.com](http://www.foursquare.com)

## 4 SOCIÁLNÍ SÍŤ A JEJÍ DEFINICE

Jedním z hlavních principů sociální sítě je zejména vytváření skupiny osob, které mezi sebou pojí nebo vytváří určité vztahy. Tyto vztahy nemusí být založeny jen na samotném přátelství, ale hlavním sjednocujícím prvkem mohou být také společné zájmy, etnická nebo náboženská podobnost, atd. Mezi takto vytvořenými skupinami osob pak dochází k přímé interakci, kdy mezi sebou lidé komunikují, předávají si informace, sdílí fotografie, články či videa.

### 4.1 Sociální sítě, definice, historie

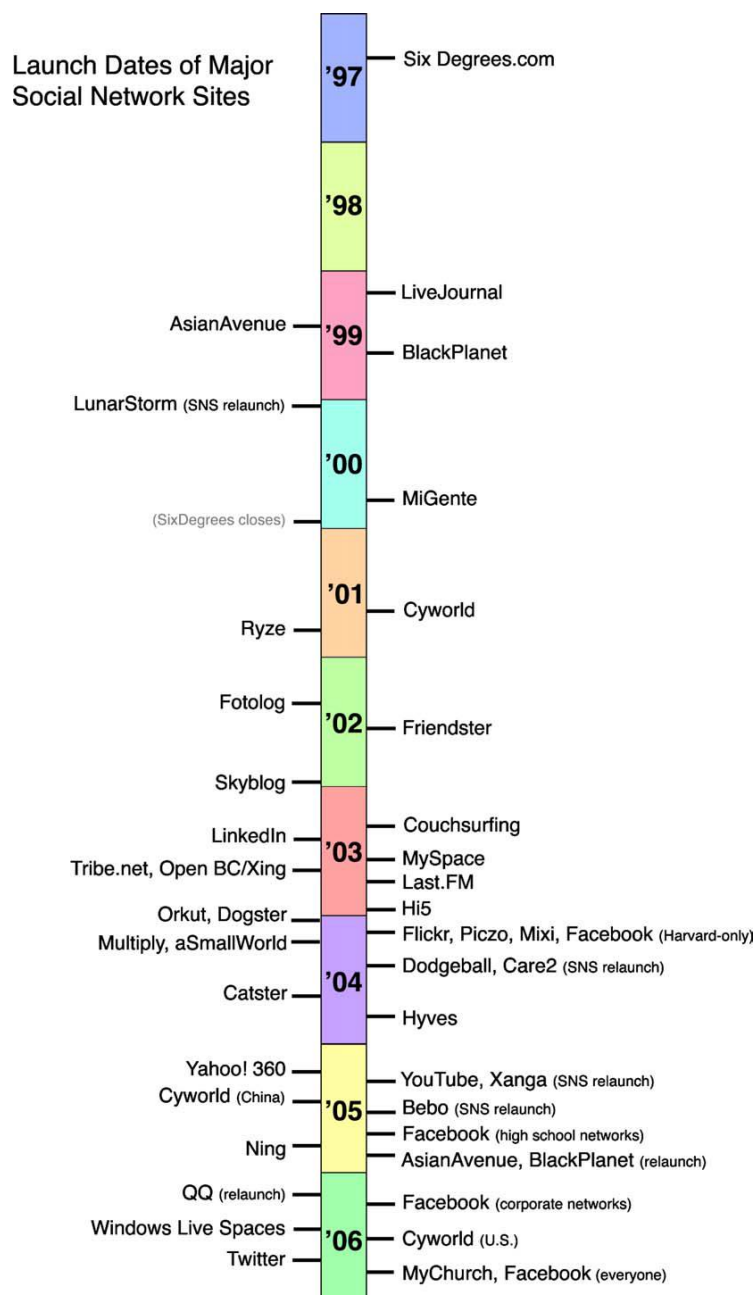
Samotný pojem sociální síť pochází ze sociologie. Lidé v takovéto síti představují tzv. „uzly“ (*node* nebo *vertex*) a vytváří mezi sebou vztahy, charakterizované jako „hrany“ (*edge*). S rozvojem internetu, a zejména pak Webu 2.0, se vyvinul i pojem webová sociální síť a vytvořily se tak i nové definice původního pojmu (Haythornthwaite, 2005).

Boyd (2008) definuje sociální sítě jako webově založené služby, které umožňují jednotlivcům budovat veřejný nebo napůl veřejný profil v limitovaném rozhraní, vytvářet seznamy dalších uživatelů, s nimiž jsou ve spojení, které je možné procházet a to vše v rámci jednoho ohraničeného systému. Základem takové sociální sítě je uživatel, který si vytváří svůj profil. Profil je generován pomocí odpovědí na (často osobní) otázky, které obvykle obsahují dotazy na jméno, věk, místo pobytu či zájmy. Většina sítí také doporučuje využívat profilové fotografie. Záleží pak většinou na vlastníkovvi profilu, zda vyplněné informace zveřejnění každému nebo jen určité skupině lidí. Jak bylo popsáno výše, klíčovou vlastností sociálních sítí je vytváření listu kontaktů (přátel, následovníků, ...). U většiny sítí si uživatel přidává přátele, kteří profil již v rámci sítě založený mají, ale možností je i pozvat neregistrované přátele například prostřednictvím e-mailu. Zvláštností webových sociálních sítí je, že se zde potkávají i uživatelé, kteří se nemusí v reálném životě nikdy potkat.

Výše uvedené vlastnosti definují sociální sítě v moderním pojetí. Vůbec první sociální síť splňující všechna uvedená kritéria byla v roce 1997 založená SixDegrees.com. Její prostředí umožňovalo vytváření profilů, seznamu přátel a možnost odesílání zpráv mezi uživateli. Ačkoli web přitáhl miliony uživatelů, nepodařilo se ho udržet a zanikl tak v roce 2000. Sám zakladatel se domnívá, že síť SixDegrees prostě jen předběhla svou dobu, protože většina uživatelů internetu v té době nevyužívala internet v takové míře jako dnes. V období mezi lety 1997 a 2001 se množství sociálních sítí začalo zvětšovat, nicméně žádná z nich se do světového podvědomí nedostala.

Změnu přinesl až vznik sítě Friendster v roce 2002, jejíž hlavní pointou byla on-line seznamka. Tato seznamka byla založena na předpokladu, že uzavřít vztah s někým cizím je pravděpodobnějším prostřednictvím společného přítele. Na rychle vzrůstající množství uživatelů však nebylo prostředí sítě připraveno. Díky snaze vyřešit velké technické problémy, vznikaly v síti často i nesmyslná omezení, což vyústilo v úpadek zájmu. Od této doby však začalo vznikat mnoho projektů (obr. 4.1), z nichž ty nejúspěšnější

přetrvaly až do dnes (MySpace, CouchSurfing, Last.FM, YouTube, ...). Fenomén dnešní doby, kterým je Facebook, pak vznikl v roce 2004. Aktuálně existuje více než 200 různých sociálních sítí (Pavliček 2010).



Obr. 4.1 Časová osa data spuštění hlavních sociálních sítí (zdroj: Boyd, 2007)

## 4.2 Prostorová data

Prostorová data neboli geodata jsou data vztažená k určitému místu na Zemi. V prostředí sociálních sítí se objevují v různých formách. Kromě údajů o bydlišti nebo místě pobytu uživatele se s rostoucím vývojem mobilních technologií staly součástí



sociálních sítí také geolokační prvky. Uživatelé tak mohou často ovlivňovat chod svého okolí nebo se podílet na šíření informací z okolí. Podle uplatnění prostorového prvku v síti je tak sociální síť možné rozdělit do tří skupin:

- Primárně geolokační sociální síť (geosociální síť)
- Sociální síť s možností geolokace
- Sociální síť s prostorovým prvkem ve formě textu

U primárně geolokačních sítí (anglicky označovány jako LBSN – *Location Based Social Networks*) je nejdůležitější prostorová pozice uživatele. Proto jsou tyto sítě založeny na informaci o tom, kde uživatel je, co tam případně dělá, a jaké má z toho pocity s cílem sdílet tyto informace s ostatními uživateli. S rostoucím rozšířením tzv. „chytrých“ mobilních telefonů, jsou pak populárnější i geolokační služby, kde základním požadavkem je právě mobilní zařízení s možností určení polohy uživatele a možností následného odeslání do dané sociální sítě.

Za zakladatele geosociální sítí se považují Dennis Crowley a Alex Rainert, kteří v roce 2000 spustili službu Dodgeball. Tato služba byla založena na posílání SMS zpráv s polohou uživatele, která se následně rozšířila všem přátelům uživatele. Nicméně po skoupení služby firmou Google v roce 2005 byla sociální síť v roce 2009 zrušena (Ziv, 2006).

Nejznámější a nejrozšířenější geolokační sítí je *Foursquare*. Tato sociální síť je založena na tzv. *check-inech* na jednotlivých místech (*venues*), díky kterým pak uživatelé vidí všichni jeho Foursquare přátelé. Lokaci na daném místě provádí uživatel sám a to použitím GPS ve svém zařízení. Mapovým podkladem pro Foursquare jsou pak data z OpenStreetMap. Za každé checknutí získá uživatel také body, díky kterým je pak tato sociální síť i geolokační hrou. Další možností je také vyhledávání objektů v daném okolí s různými atributy. Takto je Foursquare nejen geosociální sítí uživatelů, ale poskytuje i možnosti k lokalizované reklamě či jako podpora turismu. Další podrobnosti popisuje kapitola 4.2.

Další populární primárně geolokační sociální síť byla v minulosti Gowalla, která však svou činnost ukončila v roce 2012 po tom, co byla odkoupena sociální sítí Facebook. V té době bylo uživatelů Gowally téměř dvacet miliónů (Machulda, 2012). Dalšími geosociálními sítěmi jsou Brightkite, Yelp, IJustMadeLove nebo Loopt.

Mezi sociální sítě s možností geolokace patří dnes nejrozšířenější a nejznámější sociální síť jako Facebook, Twitter. Dále pak Couchsurfing, Google+<sup>7</sup>, Instagram, Flickr, Panoramio a další.

Možnost geolokace v těchto sítích funguje jako doplňující funkce k některým z prvků. Například u Facebooku je možnost přidat geolokaci ke statusu, fotografii nebo k videu.

---

<sup>7</sup> <https://plus.google.com/>

U fotografií a videí je i možnost přidat prostorový prvek až po nahrání fotografie. Sociální síť Facebook a její prostorová data jsou podrobněji popsány v kapitole 4.3.2.

Díky až sekundární funkci u těchto sociálních sítí není množství takovýchto příspěvků výrazné. Například počet geolokovaných tweetů v sociální síti Twitter se pohybuje v rozmezí mezi 0,5 % až 3 % z celkového počtu (Karimi, 2014).

Poslední skupinou jsou sociální sítě s prostorovým atributem, který je nejčastěji přítomen ve formě textu či doplňující informace např. v uživatelském profilu či přímo textu zprávy. Typ záleží na druhu sociální sítě. Často je spojován přímo s uživatelem (město, region pobytu), nebo jako doplňující informace k prvku (místo vzniku filmu na ČSFD), adresa firmy či školy (LinkedIn), apod.

### 4.3 Vybrané sociální sítě

Pro detailnější popis sociálních sítí a prostorových dat v nich byly vybrány sítě Twitter, Facebook a LinkedIn. Motivace výběru jednotlivých sítí byla nastíněna již v kapitole 2.1. Všechny tyto sociální sítě mají možnost přístupu k prostorovým atributům a získaná data tak mohou posloužit k dalším vizualizacím či analýzám.

#### 4.3.1 Twitter

Sociální síť Twitter fungující od roku 2006 (Carlson, 2011) je založena na krátkých textových zprávách (maximálně 140 znaků), ve kterých může uživatel kromě textu sdílet i fotografie nebo odkazy webové stránky. Příklad jednoho tweetu je na obr. 4.2.



Obr. 4. 2 Příklad jednoho tweetu uživatele @dancafuskova (Zdroj: vlastní)

#### Historie a popis sítě Twitter

Podle autorů Boyda a Ellisona (2007) vznikl Twitter jako vnitřní komunikační služba pro zaměstnance společnosti Odeon. Odesílané textové zprávy se nazývají tweety. Právě zmíněná stručnost příspěvku dělá Twitter jedinečným, protože díky omezení velikosti mají jednotlivé zprávy silnou sdělovací hodnotu. Spříznění uživatelé v této sociální síti nejsou nazýváni „přátelé“, jak je známe z jiných sociálních sítí, ale jde o tzv. „následovníky“ (*followers*). Uživatel může pak příspěvky ostatních označit za

oblíbené nebo je předat dále (*retweet*). V těle tweetu lze označit libovolného uživatele symbolem *@uzivatelskejmeno*, nebo použít klíčová slova, tzv. *hashtagy*, které jsou označeny symbolem *#hashtag*. Pro připojení odkazu používá twitter vlastní zkracovač adres. Twitter umí také shromažďovat příspěvky daného tématu do seznamů (*lists*), díky kterým odběratelé dané skupiny mohou získávat příspěvky od uživatelů s podobným zaměřením či zájmy. Přidání místa tweetu – tedy jeho geolokace, bohužel není u Twitteru příliš populární. Další možností je přidání fotografie k tweetu.

Registrací do sociální sítě získá uživatel uživatelské jméno, které je vázáno na konkrétní e-mailovou adresu, která ale není primárně veřejná. Přidané příspěvky jsou pak zveřejňovány pod tímto uživatelským jménem. Uživatel si může vybrat, zda chce, aby jeho tweety byly veřejné nebo soukromé pouze pro povolené uživatele. Podobně může uživatel zvolit i přidání geografické polohy k jednotlivým tweetům. V profilových informacích může uživatel také vypsát polohu (*location*), která je definována Twitterem jako „*Kde na světě se nacházíte*“. Uživatelem nemusí být pouze fyzické osoby, ale také firmy, sportovní kluby, aplikace, hudební skupiny aj.

Samotné rozhraní sociální sítě je strukturováno do několika částí. Na hlavní stránce (*Home*) uživatel vidí tweety těch uživatelů, které uživatel sleduje. Může je rozbalit a podívat se na podrobnější informace, přidat mezi oblíbené, reagovat na něj vlastní zprávou, nebo nechat zobrazit na svůj profil (*retweetovat*). V případě cizojazyčných příspěvků je v českém prostředí možnost zobrazení překladu. Dalšími částmi jsou Oznámení (*Connect*) a Objevujte (*Discover*), Já (*me*), Hledání na twitteru (*Search Twitter*), Soukromé zprávy (*Direct message*), Nastavení (*Settings and help*), Koho sledovat (*Who to follow*), Trendy (*Trends*). Všechny tyto části se snaží vytvářet intuitivní pohodlné prostředí pro snadnou práci.

Zajímavostí je český portál *klaboseni.cz*<sup>8</sup>, který vznikl jako vyhledávací služba v archivu českého a slovenského Twitteru. Krom vyhledávání tweetů jsou zde také statistiky nejoblíbenějších hashtagů, nejoblíbenějších uživatelů, tweetů a jiné (Klábosení, 2014).

## Twitter API

API u Twitteru povoluje přístup k datům prostřednictvím tří částí aplikačního rozhraní. *Search API*, *Streaming API* a *REST API*.

*Search API* slouží k vyhledávání tweetů podle klíčových slov a to nejen samotného textu, frází nebo hashtagů, ale také k vyhledávání fotografií, URL adres. Zmenšit okruh výsledných tweetů lze také pomocí geolokace, jazyka, nebo doby uveřejnění příspěvku. Podle nastávajících pravidel Twitteru, lze vyhledávat tweety staré maximálně 8 dní (Twitter Developers Rules, 2014).

Pomocí *REST API* lze vyhledat uživatele, jejich timelines (uveřejněné tweety v závislosti na době zveřejnění), nebo následovníky (*followers*), seznamy (*lists*), místa,

---

<sup>8</sup> <http://www.klaboseni.cz/>

trendy, oblíbené tweety. Díky REST API lze také zprávy na Twitter zveřejnit, nebo smazat, odpovídat na některé tweety, přeposílat tweety (retweet), nebo přidávat tweety do oblíbených.

Streaming API slouží k získání tweetů v reálném čase. Tweety lze získat pro dané klíčové slovo, nebo také podle počtu následovníků, lokace, jazyka, nebo časové zóny. Omezení u streaming API je 1 % pro klasického uživatele. Je také možnost získat pouze geolokované tweety, nebo tweety napsané do určitého okruhu od dané polohy.

Pro práci s API u Twitteru je potřeba nejdříve získat autorizační protokol *OAuth*. Ten uživatel získá po registraci aplikace v části Twitteru pro vývojáře. Uživatel získá přístupové údaje *Token credentials*. Součástí je *API key* a *API secret*, *Access token*, *Access token secret*, které se skládají z písmen a číslic. Po zadání těchto přístupových údajů společně s *Request token URL* se provede online autorizace, která má neomezenou dobu platnosti. Lze ji ale v nastavení aplikace na webu zrušit. Počet dotazů je limitován na 350 za hodinu na jednu autorizaci.

Získané informace se liší podle charakteru dotazu a podle použití dané části API. Search API sice vyhledá nejen poslední aktuální tweety, ale také příspěvky z historie to až osm dní staré, nicméně pro detailnější informace (například informace o státě, jazyce, lokaci...) je potřeba využít část REST API a v případě většího počtu tweetů je tato metoda velmi časově náročná. Pro dolování většího množství tweetů je tak pohodlnější použít prostředí Streaming API. Jak už bylo řečeno, tweety se stahují v aktuálním čase běhu aplikace. Díky licenčním podmínkám je možnost získat pouze 1 % všech tweetů. Je však možnost zažádat o navýšení na 10 %. V žádosti je potřeba podrobně popsat důvody a následné využití dat a také se žadatel zavazuje, že získané data a analýzy s nimi spojené poskytne zpětně Twitteru. Získaný tweet prostřednictvím Streaming API bez omezení počtu atributů tak obsahuje 42 vlastností (*text*, *retweet\_count*, *favorited*, *truncated*, *id\_str*, *in\_reply\_to\_screen\_name*, *source*, *retweeted*, *created\_at*, *in\_reply\_to\_status\_id\_str*, *in\_reply\_to\_user\_id\_str*, *lang*, *listed\_count*, *verified*, *location*, *user\_id\_str*, *description*, *geo\_enabled*, *user\_created\_at*, *statuses\_count*, *followers\_count*, *favourites\_count*, *protected*, *user\_url*, *name*, *time\_zone*, *user\_lang*, *utc\_offset*, *friends\_count*, *screen\_name*, *country\_code*, *country*, *place\_type*, *full\_name*, *place\_name*, *place\_id*, *place\_id*, *place\_lon*, *lat*, *lon*, *expanded\_url*, *url*). Ne každý tweet ale všechny informace obsahuje. Vždy záleží na autorovi tweetu, které informace jsou o zprávě dostupné. Velikost databáze 100 tweetů uložených do formátu JSON se pak pohybuje mezi 420 – 470 kB.

Vzhledem k licenčním podmínkám, které zakazují web scraping u Twitteru, je API jediným neplaceným způsobem jak se k datům dostat. Komerčním způsobem je využití služeb např. Gnip nebo DataSift, které jsou oficiálními partnery Twitteru. Uživatel si také může stáhnout archiv svého profilu. Tato data po odeslání žádosti dostane ve formátech JSON, nebo CSV (Twitter developers, 2014).

### 4.3.2 Facebook

Facebook je s více než miliardou uživatelů (1,23 miliarda aktivních účtů k 31. 12. 2013) největší a nejpobulárnější sociální síť současnosti (Facebook Reports, 2014). Jeho historie, popis prostředí a API rozhraní je popsáno v následující kapitole.

#### Historie a popis sítě Facebook

Mark Zuckerberg se spoluzakladateli Dustinem Moskovitzem, Chrisem Hughesem a Eduardem Severinem založili Facebook v roce 2004 původně pouze pro studenty Harvardské univerzity. Postupně se pak k síti mohli připojovat i další uživatelé s mimo harvardskými univerzitními adresami. První univerzitou v České republice, která byla schválena pro vstup studentů do systému, byla Masarykova univerzita. Později se začaly připojovat i velké obchodní společnosti až se nakonec od 11. srpna 2006 může připojit kdokoli starší třinácti let. V současnosti je Facebook plně přeložen do šedesáti osmi jazyků.

První funkcionalitou bylo pouze vyhledávání a prohlížení si profilů ostatních uživatelů v rámci univerzity. Později se k prostředí přidala populární „zed“ a postupně bylo umožněno přidávání fotografií, komentování, možnost „Líbí se mi“ (*like*), geolokace, sdílení cizích příspěvků, vyjádření pocitů u statusů.

K registraci stačí zadat křestní jméno, příjmení, e-mail, datum narození a pohlaví. Krom registrace osob lze také zaregistrovat stránku pro celebrity, skupinu nebo společnost. Přihlašovacími údaji se tak stal e-mail a heslo vložené při registraci. Každý uživatel získá svůj osobní profil, ve kterém může sdělit informace členěné do několika kategorií (základní informace, práce a vzdělání, místa kde žil, kontaktní údaje, vztah, rodina, ...). U všech sdělených informací může uživatel nastavit, kdo má právo vidět dané informace (veřejné, soukromé pro přátele, soukromé pro určitou skupinu lidí, aj.). Podobně uživatel může nastavit i nástroje, kdo ho může kontaktovat nebo vyhledat. Postupem času se do profilu přidávají informace, které uživatel získal díky své aktivitě na této síti (přidání přátel, hraní her, zvolení oblíbenosti různých stránek, přidání fotografií, aj.).

Sociální síť Facebook slouží především k propojení uživatelů, kteří se znají nebo mají společné zájmy. Příkladem může být síť vztahů přátel uživatele Danča Fusková vytvořené aplikací TouchGraphPhoto (2014). Na obrázku 4.3 lze vidět 100 přátel uživatelky a jejich vzájemné vztahy. Aplikace rozdělí přátele do skupin podle množství vztahů. Na tomto příkladu lze vidět třeba modrou skupinu přátel z *kanoistiky*, červeně jsou znázorněni přátele z *gymnázia*, žlutě z *Univerzity Palackého*. Jde také vidět, že uživatel *Radek Fusek* (bratr uživatelky) má například společné vztahy téměř s každou skupinou.



powered by  
TouchGraph

Obr. 4.3 Příklad sítě vztahů uživatky Danča Fusková (zdroj: vlastní)

Pro Facebook jsou důležité dvě aktivity - psaní nebo sdílení statusů, fotografií, stránek, videí a také dávání najevo oblíbenosti jiných statusů, stránek, fotografií pomocí tzv. To se mi líbí (*like*), komentováním, nebo sdílením. Veškerá aktivita se pak zobrazuje na tzv. zdi (*timeline*), což je také hlavní stránka každého uživatele. Další možností je také zakládání skupin uživatelů a to jak veřejných, tak uzavřených. Uživatelé, stránky, nebo i skupiny mají právo zakládat události, kterých se pak uživatelé mohou zúčastnit. Další aktivitou je psaní soukromých libovolně velkých zpráv mezi jednotlivými uživateli.

Geolokační služba nebyla na Facebooku prvotně zamýšlena. Až díky konkurenci v sítích Foursquare, či Gowalla se v roce 2010 rozhodli vývojáři sítě přidat službu Facebook Places. Podobně jako v jiných sítích tak bylo možné sdílet se svými přáteli informaci o místě pobytu. Nicméně se tato služba netěšila velkému zájmu, převážně proto, že lidé, kteří by měli o tuto službu zájem, většinou už využívali konkurenční Foursquare. Proto v roce 2011 byla služba zrušena a geolokace byla implementována jako doplňková informace pro populární sdílení statusů nebo fotografií (Moon, 2014).

Další prostorovou informací v této sociální síti je informace o současném, či minulém bydlišti, místě narození, nebo aktuální adrese uživatele. Většina firem, nebo podniků pak sdílí veřejně místo působnosti. Tyto informace se pak dají buď zobrazit přímo v prostředí sociální sítě, nebo v prostředí Bing map.

## API Facebook

Primární způsob pro získání dat z Facebooku je pomocí *Facebook social graph*. Social graph je to proto, že hlavní pointou je spojení daného uživatele s tím, co ho zajímá – sociální graf a to prostřednictvím hran (*edges*), uzlů (*nodes*) a polí (*fields*). Hranami rozumíme vztahy mezi objekty, například *líbí se mi* u příspěvků, účasti na událostech, sdílení, nebo komentáře. Uzly jsou pak samotné objekty (uživatel, stránka, skupina, fotografie, album, událost, ...). Pole jsou pak informace o jednotlivých objektech jako například narozeniny uživatele, nebo název stránky nebo skupiny. Graph API má několik verzí, z nichž nejnovější je Graph API v2.0. API je založeno na protokolu HTTP low-level, který se používá pro získání dat, odesílání zpráv, nahrávání fotografií a řadu dalších úkolů. Každý objekt v prostředí sociální sítě Facebook má své vlastní ID, díky kterým lze přistupovat k jednotlivým vlastnostem a přes které lze pak dále pracovat s daty.

To, co je na Facebooku veřejně, je veřejně i přes API bez potřeby autorizace aplikace. Příkladem je volání použitím url adresy: `graph.facebook.com/ID_uzivatele`

Příklad pro volání veřejných informací o autorovi práce:

```
http://graph.facebook.com/1169572976,
```

výsledek dostaneme ve formátu JSON jako:

```
{
  "id": "1169572976",
  "first_name": "Dan\u010da",
  "gender": "female",
  "last_name": "Fuskov\u00e1",
  "locale": "cs_CZ",
  "name": "Dan\u010da Fuskov\u00e1",
  "username": "dancaf"
}
```

K dalšímu postupu je ale třeba získat přístupový kód a následná práce běží přes identifikaci *OAuth Access Token* (dále přístupový klíč). Přístupový klíč je řetězec, který identifikuje uživatele, aplikace nebo stránky a umožňuje přístup třetích stran k jejich datům bez potřeby sdílet citlivé informace jako jsou hesla (Russell, 2011).

Přístupových klíčů je několik druhů a existují různé způsoby jejich získání. Daný klíč obsahuje informace o platnosti klíče a také, na které aplikaci nebo části API se vztahuje.

Typy přístupových klíčů (Facebook Developers, 2014):

### 1. User Access Token

- Tento uživatelský přístupový klíč je nejčastěji používaný typ. Má omezenou dobu platnosti. Lze s ním číst, upravovat nebo přidávat data prostřednictvím jména daného uživatele. Je získán pomocí registrace běžného uživatele jako vývojář a následné potvrzení uživatele k přístupu k jeho informacím.

## 2. App Access Token

- Pomocí tohoto druhu klíče lze data nejen číst, ale také upravovat nastavení aplikace. Je generován pomocí předem dohodnutého postupu mezi aplikací a Facebookem.

## 3. Page Access Token

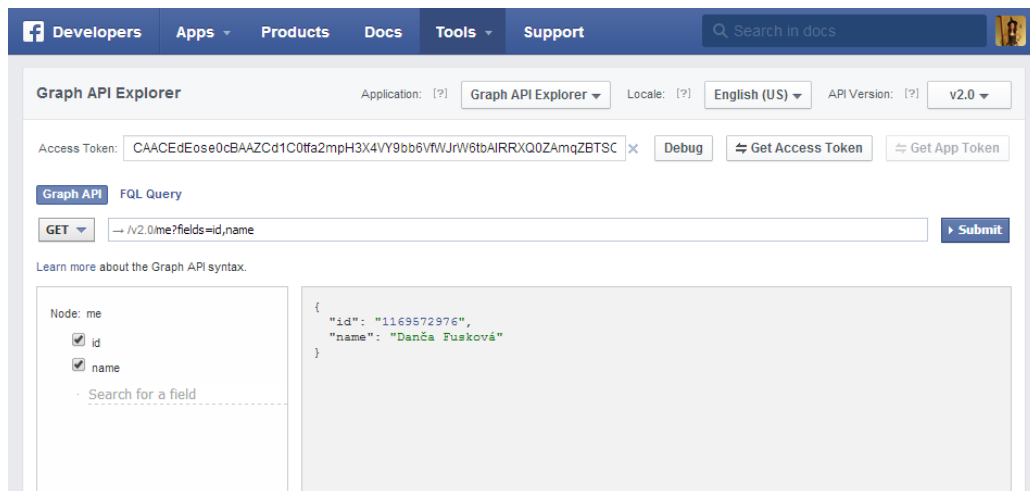
- Přístupový klíč ke stránkám na facebooku. Umožňuje přístup přímo k rozhraní API jednotlivých stránek. K získání je nejprve třeba požádat o tzv. správcovství stránky a následně až o Page Access Token.

## 4. Klient Token

- Tento klíč je identifikátor, který může uživatel vložit do nativních binárních souborů, nebo do desktopových aplikací k identifikaci aplikace. Používá se pro přístup k rozhraní aplikace. Používá se však velmi málo.

### Příklad získání User Access Token

Příklad získání User Access Token, tedy nejčastěji používaného přístupového klíče je zobrazeno na obr. 4.4. Po přihlášení do vývojářského prostředí je uživatel dotázán na typy oprávnění (obr. 4.5), se kterými chce pracovat. Platnost vytvořeného klíče je dvě hodiny. Uživatel může vytvořit až 500 klíčů pro jednu aplikaci.



Obr. 4.4 Získání přístupového klíče v prostředí developers Facebook (zdroj: vlastní)



Select Permissions

User Data Permissions
Extended Permissions

<input checked="" type="checkbox"/> user_about_me	<input checked="" type="checkbox"/> user_actions.books	<input checked="" type="checkbox"/> user_actions.music
<input checked="" type="checkbox"/> user_actions.news	<input checked="" type="checkbox"/> user_actions.video	<input checked="" type="checkbox"/> user_activities
<input checked="" type="checkbox"/> user_birthday	<input checked="" type="checkbox"/> user_education_history	<input checked="" type="checkbox"/> user_events
<input checked="" type="checkbox"/> user_friends	<input checked="" type="checkbox"/> user_games_activity	<input checked="" type="checkbox"/> user_groups
<input checked="" type="checkbox"/> user_hometown	<input checked="" type="checkbox"/> user_interests	<input checked="" type="checkbox"/> user_likes
<input checked="" type="checkbox"/> user_location	<input checked="" type="checkbox"/> user_photos	<input checked="" type="checkbox"/> user_relationship_details
<input checked="" type="checkbox"/> user_relationships	<input checked="" type="checkbox"/> user_religion_politics	<input checked="" type="checkbox"/> user_status
<input type="checkbox"/> user_tagged_places	<input checked="" type="checkbox"/> user_videos	<input checked="" type="checkbox"/> user_website
<input checked="" type="checkbox"/> user_work_history		

Public profile included by default.
[Get Access Token](#)
[Clear](#)
[Storno](#)

Obr 4.5 Typy oprávnění ve vývojářském prostředí sociální sítě Facebook (zdroj: vlastní)

Další možností pro získání dat je přes Facebook Query Language (FQL). Jde o dotazovací mechanismus založený na principu SQL, který vyhledává v relační databázi Facebooku. Odpovědí je JSON objekt, který obsahuje vlastnosti požadovaného objektu. Kromě známé formy *SELECT FROM WHERE* je možnost použít i příkaz *IN* pro vložení vnitřního příkazu. Příkladem je dotaz na ID, jméno a místo narození uživatele s přístupovým klíčem a jeho přátel zobrazený na obrázku 4.6.

The screenshot shows the Facebook Graph API Explorer interface. At the top, there are navigation tabs for Developers, Apps, Products, Docs, Tools, and Support. Below this, the application name is 'Graph API Explorer' and the locale is 'English (US)'. The API version is set to 'v2.0'. An access token is provided: 'CAACEdE0se0cBAEapxM2uT3UYkCksHID28IMZCMoOINdbNmMGfXHWfcWEBBI'. There are buttons for 'Debug', 'Get Access Token', and 'Get App Token'. The main section is titled 'Graph API' and has a sub-tab for 'FQL Query'. The query entered is:
 

```
SELECT uid, name, hometown_location FROM user
WHERE uid = me() OR uid IN
(SELECT uid2 FROM friend WHERE uid1 = me())
```

 A 'Submit' button is visible to the right of the query. Below the query, there is a link to 'Learn more about the Graph API syntax'. The response is a JSON object:
 

```
{
  "data": [
    {
      "uid": "1169572976",
      "name": "Danča Fušková",
      "hometown_location": {
        "city": "Valasske Mezirici",
        "state": "Zlínský kraj",
        "country": "Czech Republic",
        "zip": "",
        "latitude": 49.4667,
        "longitude": 17.9667,
        "id": "108517352513384",
        "name": "Valasske Mezirici"
      }
    },
    {
      "uid": "1246030674",
      "name": "Petr Čulík",
      "hometown_location": null
    }
  ]
}
```

Obr 4.6 Příklad získání dat o autorovi práce prostřednictvím FQL (zdroj: vlastní)

### 4.3.3 LinkedIn

Profesní sociální síť LinkedIn sdružuje profesionály z různých oborových odvětví a umožňuje jim tak přímý zejména profesní kontakt. Krom toho je využíván i firmami v případě náborových kampaní nebo pro oznámení aktuálních volných pozic.

#### Popis a historie sítě

LinkedIn je sociální síť fungující od roku 2003 a orientuje se na databázi profesionálů z různých oborů. Za zakladatele je považován Reid Hoffman (spoluzakladatel také platebního systému PayPal). Společnost je v soukromém vlastnictví a její obchodní model je založen na kombinaci příjmů z předplatného, reklamy a placených náborových řešení. Největší rozmach má v USA, Indii a Brazílii (LinkedIn : Press Center, 2014).

Uživatel se do služby může registrovat pomocí e-mailu a hesla, kde zadá své jméno a příjmení nebo může využít registraci přes účet Facebooku. Podle Tůmové (2013) je hlavní pointou vytvoření sítě odborníků. Proto by uživatel měl vyplnit co nejvíce informací do svého profilu a to takové informace, které se běžně uvádí v životopisech (především pracovní zkušenosti, vzdělání, jazykové zkušenosti, získané certifikáty nebo ocenění a hlavně dovednosti, které ovládá). Funkcionalitou sítě je i možnost vybrání zájmu o zaměstnání v daném oboru, nebo firmě. Právnícké osoby se do systému registrují podobně jako lidé, jen vyplňují jiné informace (informace o působení, produktech, nabídky na zaměstnání, ...). Uživatel si pak vytváří síť kontaktů, mezi kterými vznikají spojení (*connections*). Síť mohou tvořit nejenom známí uživatelé, ale také firmy nebo jednotlivci, kteří pro uživatele mají nějakou hodnotu (potenciální zaměstnání, zájem o služby, ...). V hlavním prostředí jsou pak zveřejňovány nové informace o uživatelích v síti kontaktů (nová spojení, nová zaměstnání, aktualizace profilů, ...). Dalšími částmi jsou sekce Profil, Síť, Pracovní příležitosti, Zájmy a Obchodní služby. Typy vztahů jsou rozlišovány na stupně:

- **Stupeň 1** – vztah uživatelů, kteří jsou v přímém kontaktu (navzájem přijali žádost o přátelství)
- **Stupeň 2** – uživatelé, kteří jsou přátelé některého z uživatelů v síti kontaktů
- **Stupeň 3** – lidé, kteří jsou ve spojení s přáteli Stupně 2
- **Ostatní** – tito uživatelé nemají žádné spojení s kontakty majitele účtu, nelze se s nimi přímo spojit, nicméně je lze kontaktovat prostřednictvím zprávy

Prostorové prvky v této sociální síti nejsou geolokačního charakteru, ale pouze jako volitelné prostorové atributy. Prvním prvkem je lokalita uživatele řazena podle poštovního směrovacího čísla. Další prostorové prvky jsou prvky jednotlivých institucí, nebo firem (místo působení). Russel (2014) ale upozorňuje, že tyto prostorové informace nejsou uloženy jako souřadnice, nýbrž pouze jako textový řetězec. To může být při následném geokódování problémem vzhledem k existenci více míst se stejným či obdobným názvem.

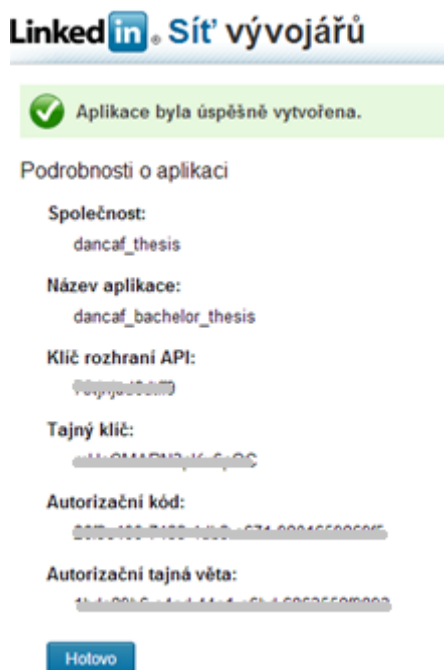
## API LinkedIn

API této sociální sítě je rozděleno do dvou částí. První částí je *JavaScript API*, která umožňuje integrovat javascriptové prvky do vlastních aplikací. Na stránkách pro vývojáře pak jde testovat funkcionalitu, jak lze vidět na obr 4.7 (LinkedIn Developer, 2014).



Obr. 4.7 JavaScript konzole sítě LinkedIn (zdroj: LinkedIn Developer, 2014)

Druhou částí API pro získání dat je *REST API*. Díky REST API lze vyhledávat informace o uživateli, konkrétních skupinách a vztazích mezi nimi. Pro práci s REST API je potřeba zaregistrovat aplikaci a získat přístupové klíče (obr. 4.8). Pro autorizaci je potřeba využít protokolu *OAuth*. V prostředí sítě LinkedIn má přístupový klíč neomezenou dobu platnosti s možností zrušení přes nastavení uživatele. Dotazy se vrací ve formátech JSON.



Obr.4.8 Získání přístupových klíčů k vytvoření aplikace v sociální síti LinkedIn (zdroj: vlastní)

V případě API u LinkedIn jsou definovány tři typy limitů.

1. **Limit pro aplikaci (*Application throttles*)** – omezuje počet dotazů celé aplikace pomocí API klíče
2. **Limit pro uživatele (*User throttles*)** – omezuje počet volání na každého uživatele dané aplikace
3. **Limit pro vývojáře aplikace (*Developer throttles*)** – slouží pro uživatele uvedené jako vývojáře (při registraci aplikace). Tito uživatelé mají přibližně čtyřikrát vyšší počet dotazů, než obyčejní uživatelé

Například posílat zprávy může celá aplikace 5000x za den, uživatel 10x denně, vývojář 40x denně. Denní počítadla se nulují vždy po půlnoci středoevropského centrálního času (Throttle Limits, 2014).

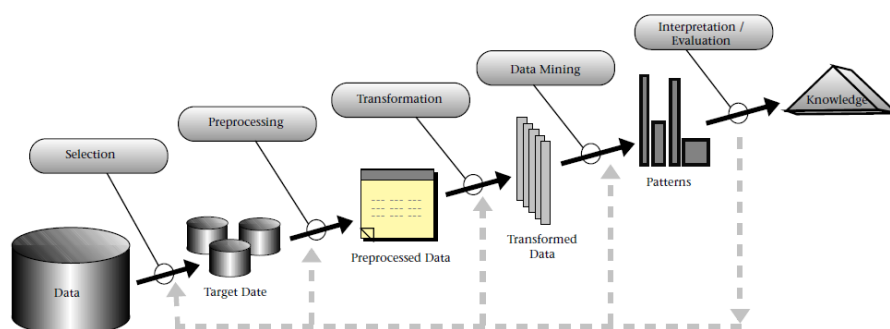
## 5 DATA MINING

Data mining je anglický termín, který se dá do češtiny přeložit jako „dolování dat“ nebo „vytěžování dat“. Usama Fayyad (1996) definuje data mining jako aplikaci specifických algoritmů pro extrakci vzorů z dat. Han a Kamber (2000) zase jako proces sloužící k objevování zajímavých znalostí z velkých objemů dat. Každopádně však jde o mezivědní obor, který zahrnuje znalosti z oborů jako je statistika, strojové učení, získávání informací, rozpoznávání nebo bioinformatika. Používán je tak v oblastech jako je obchod, finančnictví, telekomunikace a v neposlední řadě i v sociálních médiích.

Podle Hana (2009) je však samotný pojem mining (dolování) u dat nesprávný, správně by to mělo být *dolování znalostí z dat*. Právě proto mnoho lidí používá nesprávně jako synonymum pro data mining termín Knowledge Discovery in Databases (KDD), česky „získávání znalostí z databází“. Data mining je však pouze jednou z částí KDD. KDD je proces vyšší úrovně získávání informací prostřednictvím dolování dat a to společně s přípravou před dobýváním znalostí až po následnou prezentaci.

KDD se skládá z několika kroků, znázorněné také na obrázku 5.1.

- Selektce a čištění dat – odstranění irelevantních informací, integrování více zdrojů
- Předzpracování
- Transformace – přetransformování dat do vhodných forem pro dolování dat
- Dolování dat – zásadní proces, při kterém se používají inteligentní metody s cílem získat data



Obr. 5.1 Přehled kroků, které tvoří KDD proces (Zdroj: Fayyad, 1996)

KDD předpokládá, že podstatné informace jsou skryty ve velmi rozsáhlých databázích a tak je třeba využít i jiné než tradiční metody získání dat (např. dotazníkové šetření). Součástí procesů jsou rozsáhlé statistické analýzy, strojové učení, rozpoznávání vzorů, numerické hledání a vědecké vizualizace, které mají snahu se přizpůsobit

konkrétním datovým objemům a to prostřednictvím informačních technologií (Harvey, 2009).

Základní jednotkou procesu KDD je data warehouse (dále jen DW), do češtiny můžeme přeložit jako *skladiště dat*. DW je úložiště, které integruje data z jedné nebo více databází a je výsledkem několika technologických a ekologických trendů, včetně třeba snahy snížení nákladů na ukládání a zpracování dat, zároveň se snahou zvýšit hodnotu obsažených informací.

Další součástí KDD jsou často online analytické nástroje na zpracování (OLAP – *online analytical processing*). Tyto prostředky poskytují souhrnné pohledy na data a umožňují uživatelům pracovat s těmito shrnutými názory. Mezi standardní nástroje patří *roll-up* (zvyšování úrovně agregace), *dril-down* (snížování úrovně agregace), *slice* (řez) a *dice* (výběr a projekce) a *pivot* (přeorientování vícerozměrných zobrazení v datech). Nástroje jsou tak sice o něco účinnější než klasický dotazovací jazyk SQL, jsou ale méně silné než data miningové metody, protože neodhalují skryté vztahy.

## 5.1 Dolování geografických dat ze sociálních sítí

Webové sociální sítě jsou vlastně webové stránky, proto i na dolování dat ze sociálních sítí se díváme jako na získání dat z webu.

Prvním a asi nejpohodlnějším způsobem získání dat ze sociálních sítí je použití jejich veřejného aplikačního rozhraní (API – *Application Programming Interface*). API je definováno na úrovni zdrojového kódu a poskytuje úroveň abstrakce mezi aplikací a jádrem s cílem zajistit přenositelnost kódu (Foldoc, 2014). API u sociálních sítí je často založeno na architektuře REST (*Representational State Transfer*), která je postavena na faktu, že každý objekt v síti má své jedinečné označení. Výhodou může být vytvoření automatizovaného procesu pro získávání dat, nicméně API většiny sociálních sítí je omezeno počtem dotazů, časovými omezeními, nebo jinými podmínkami. Více o API u vybraných sociálních sítí v kapitole 4.3.

Další možností je tzv. web scraping. Ten funguje na systému parsování webových stránek, kdy jsou extrahovány veškeré informace ze všech sítí, které se na webových stránkách danému uživateli objevují. Tento způsob vykonává mnoho komerčních (např. Mozenda) i volně dostupných programů (Web-harvest).

Web scraping sice nemá časové omezení a ani není omezen počtem dotazů, avšak u většiny sociálních sítí je zakázán přímo v licenčních podmínkách (např. Twitter i LinkedIn). Není také vhodný pro vytěžování prostorových dat, protože získává pouze informace, které se zobrazují přímo na stránce.

Prostorová data, jak bylo řečeno v kapitole 4.2, se vyskytují v prostředích sociálních sítí v různých formách - od toho, že jsou podstatou sítě, až po stav kdy jde pouze o doplňkovou vlastnost jednotlivého objektu či uživatele. Stejně tak každá sociální síť má různé podmínky, jakým způsobem data můžete získat a následně je využít. Proto neexistuje všeobecný návod na dolování prostorových informací ze sociálních sítí.

## 6 VIZUALIZACE DAT ZE SOCIÁLNÍCH SÍTÍ

Prostorová data či prostorové atributy v různých podobách jsou v dnešní době součástí téměř všech sociálních sítí. Vzhledem k jejich stávající a stále se rozšiřující popularitě vzrůstá i počet aplikací a analytických nástrojů zpracovávajících data, která z těchto sítí lze získat. To platí samozřejmě i pro (geo)vizualizaci dat, kdy vzniká velké množství projektů využívajících nejrůznější způsoby zobrazení geodat. V této kapitole je popsáno několik z nich.

### 6.1 Možnosti vizualizace

#### 6.1.1 ArcGIS

System ArcGIS od firmy Esri představuje celou řadu produktů sloužících k analýze (geo)dat, vizualizaci dat a tvorbě map. Vzhledem k velkému množství produktů systému a širokému zájmu veřejnosti o sociální sítě je i zde několik možností, jak data ze sociálních sítí interaktivně zobrazovat.

#### Aplikace Social Media<sup>9</sup>

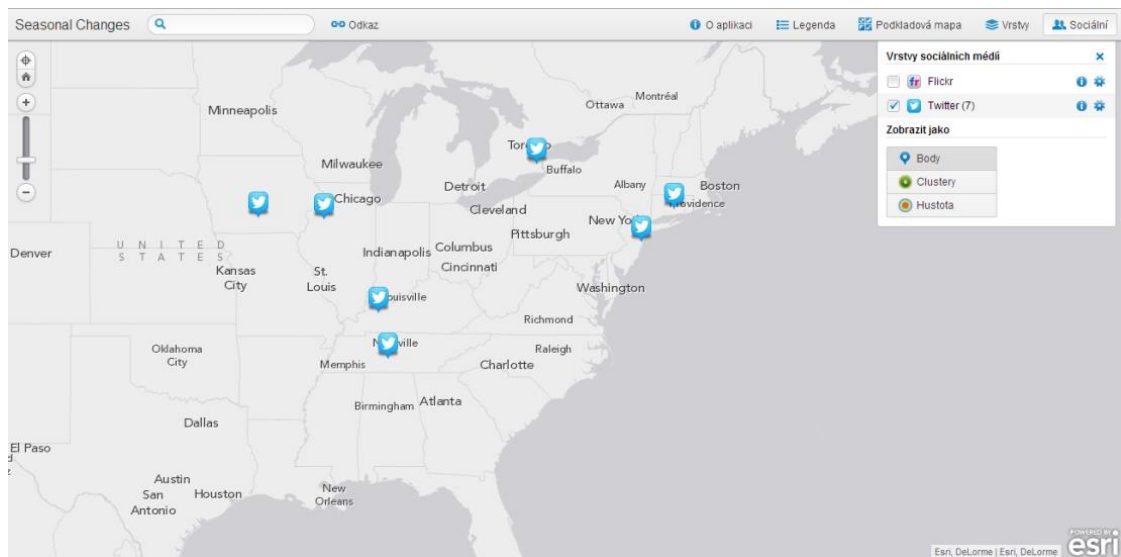
Social Media je šablona pro webové aplikace postavené na ArcGis, která umožňuje zobrazit geolokovaný obsah z Flickr, YouTube a Twitteru prostřednictvím jejich API (ArcGIS – Social Media, 2014). Výsledky mohou být zobrazeny jako body, shluk bodů nebo pomocí tzv. heat map. Šablonu mapy lze nastavit pomocí následujících možností:

- **Mapa:** možnost zvolení mapy, využité v aplikaci
- **Obsah:** zapnutí zobrazení vrstev z Flickr, YouTube nebo Twitteru
- **Hledat:** vyhledávání adres a míst
- **Sdílet:** povoluje sdílení na sociální sítě Twitter a Facebook
- **Basemapy:** zvolení podkladové vrstvy
- **Vrstvy:** povoluje zobrazit přepínač vrstev
- **Legenda:** zobrazuje legendu
- **Vrstva rendering:** zvolí požadované zobrazení bodů (body, shluky bodů, heat mapy)
- **Dialog:** Zobrazí daný text po kliknutí
- **Klíčová slova:** Zvolení klíčových slov pro hledání daných příspěvků
- **Časové období:** zadá rozsah hledání v čase

Práce s aplikací je možná dvěma způsoby. Jednoduchou metodou, která nevyžaduje širší znalosti práce s GIS klienty, je využití aplikace online přímo ve webovém prohlížeči. Veškerá výše zmíněná funkcionality zůstane zachována a aplikace postupně prohledává vybranou oblast po dobu pěti sekund. Výsledkem je tak online mapa se zobrazenými příspěvky ze zvolených sociálních sítí s daným tématem jako na obrázku 6.1 (ArcGIS – Social Media, 2014).

---

<sup>9</sup> <http://www.arcgis.com/home/item.html?id=18230986538047b4b761715f0b5dd913>



Obr. 6.1 Prostředí aplikace Social Media s příkladem vyhledání pojmu Obama. (zdroj: vlastní)

Pro uživatele, kteří mají znalost práce s JavaScriptem a jsou seznámeni s ArcGIS JavaScript API, existuje možnost stažení celého zdrojového kódu. Aplikace pak může prohledávat oblast až do maximálního poloměru, které API jednotlivých sítí dovoluje. U Twitteru je to 1500 km, Flickr 600 km a u Youtube je maximální hodnota 999 km. Aplikace je licencována na základě licence Apache, verze 2.0 (GitHub: Esri, 2014).

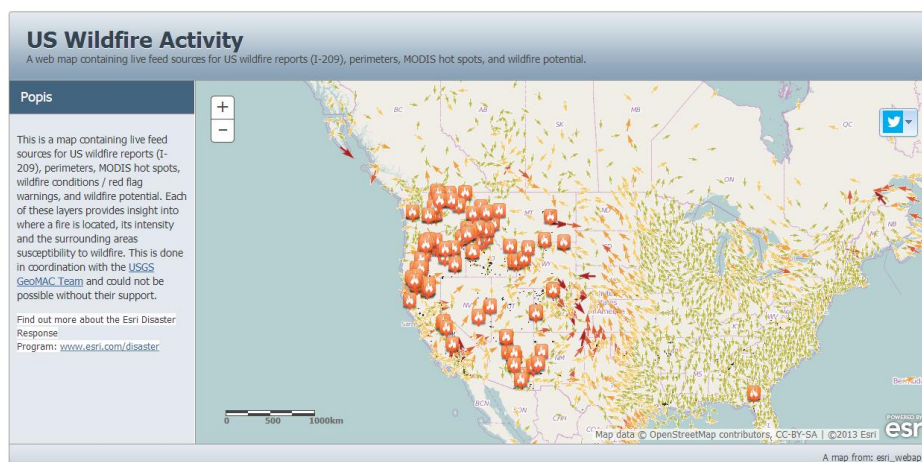
### Chrome – Twitter

Další možností zobrazení dat z Twitteru je pomocí konfigurovatelné šablony aplikace Chrome Twitter. Toto rozhraní poskytuje jednoduchou možnost zobrazení geolokovaných tweetů (ArcGIS - Chrome – Twitter, 2014). Šablonu lze nastavit pomocí následujících možností:

- **Mapa:** vybere podkladovou mapu použitou v aplikaci
- **Titul:** nastaví titul mapy
- **Vlastník:** Přepíše výchozí název vlastníka mapy na zadaný text
- **Popis:** Změní popis aplikace v postranním panelu
- **Klíčové slovo:** Hledaný textový řetězec, podle kterého se načtou geolokované tweety

Podobně jako v předchozím případě je pro plné využití možné stáhnout zabalený soubor a mít základní znalosti práce s JavaScriptem. Výsledkem práce může být třeba mapa s tweety o požárech v USA (obr. 6.2).





Obr. 6.2 Tweety o požárech v USA zobrazené pomocí aplikace Chrome-Twitter (zdroj: US Wildfire Activity, 2013)

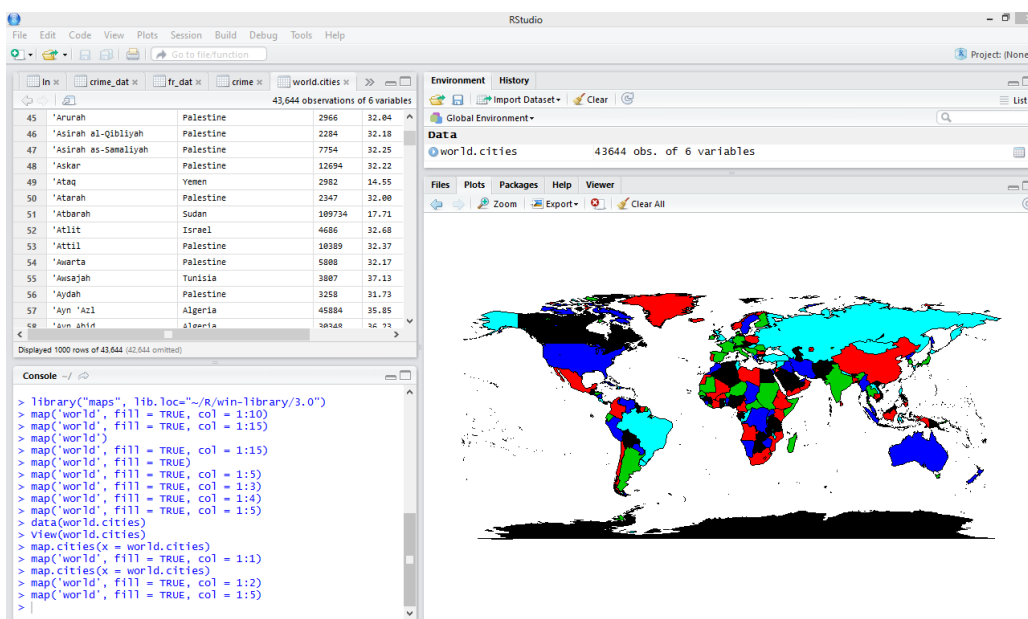
## 6.1.2 Prostředí a jazyk R

Jazyk R byl již popsán v kapitole 2.3. Díky velkému množství již vytvořených knihoven, je ale nejen programovacím jazykem pro statistické výpočty, ale také vhodným nástrojem pro vytváření grafiky a vizualizaci dat. Základní knihovnou pro tvorbu map může být balíček *maps*<sup>10</sup>. Tato knihovna, krom jiných funkcí, obsahuje databázi bodů, které reprezentují hranice států světa (*world*), nebo bodovou vrstvu měst ve světě (*world.cities*).

Zdrojový kód pro zobrazení států světa v prostředí R:

```
svet<- map('world', fill = TRUE, col = 1:5)
```

Výsledek jednoduchého příkazu je zobrazen na obrázku 6.3.



Obr. 6.3 Zobrazení států světa v prostředí R s pomocí knihovny maps (zdroj: vlastní)

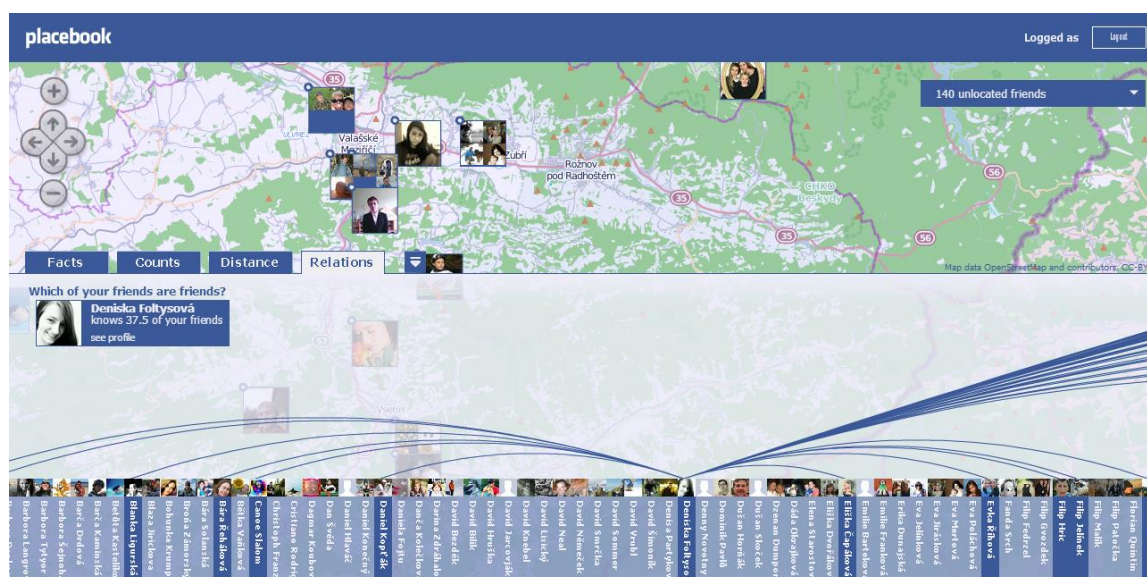
<sup>10</sup> <http://cran.r-project.org/web/packages/maps/maps.pdf>

Další knihovny, které se mohou využít pro vizualizaci prostorových dat, jsou *rmaps*, *rworldmap*, *plotGoogleMaps*, *mapdata*, *ggplot2*, *ggmap* (Statacumen.com, 2014).

Vzhledem k populárnosti sociálních sítí je i do prostředí jazyka R vytvořeno několik knihoven, zabývající se sociálními sítěmi (*twitterR*, *rFacebook*, *SocialMediaMineR* aj.). Příkladem vizualizace dat z Twitteru je studie o tweetování v době německých voleb uživatelky Sascha W. na webu R-bloggers (2014).

### 6.1.3 Placebook<sup>11</sup>

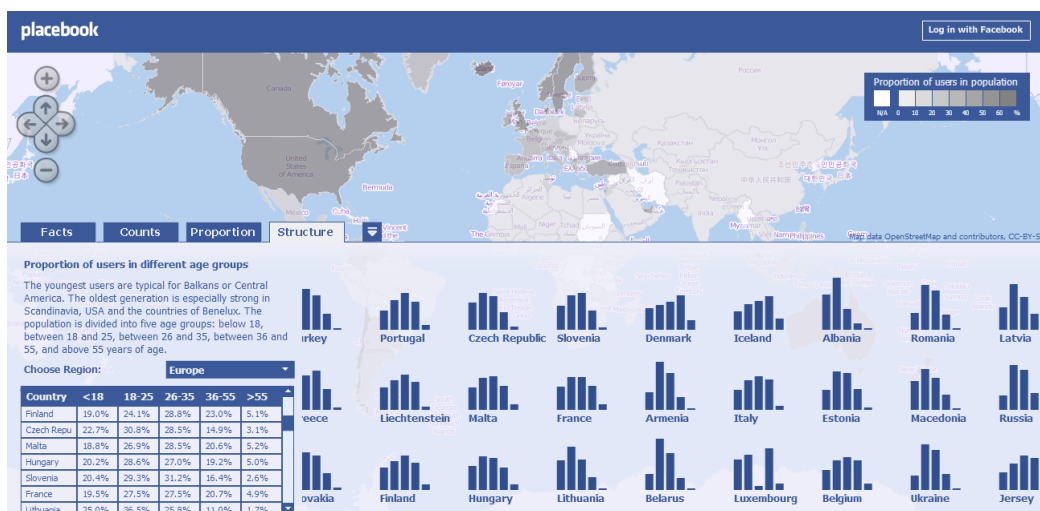
Zajímavou variantou zobrazení přátel a vztahů v sociální síti Facebook je pomocí online aplikace Placebook (Social Map, 2010). Aplikace je prací dvou studentů Masarykovy Univerzity Zdeňka Hynka a Martina Pulicara. Po přihlášení ke svému uživatelskému účtu Facebooku se načtou do mapy jednotliví přátelé. Aplikace však kromě zobrazení přátel umí také znázornit vztahy mezi jednotlivými přáteli přihlášeného uživatele (obr 6.4).



Obr. 6.4 Prostředí aplikace Placebook a interaktivní znázornění uživatele Danča Fusková (zdroj: vlastní)

Kromě informací o počtech přátel a jejich rozložení ve světě jsou zde také informace o počtech všech uživatelů a věkové struktury všech uživatelů v jednotlivých zemích. Tato data jsou zobrazena nejen v tabulkách, ale také v mapě (obr 6.5).

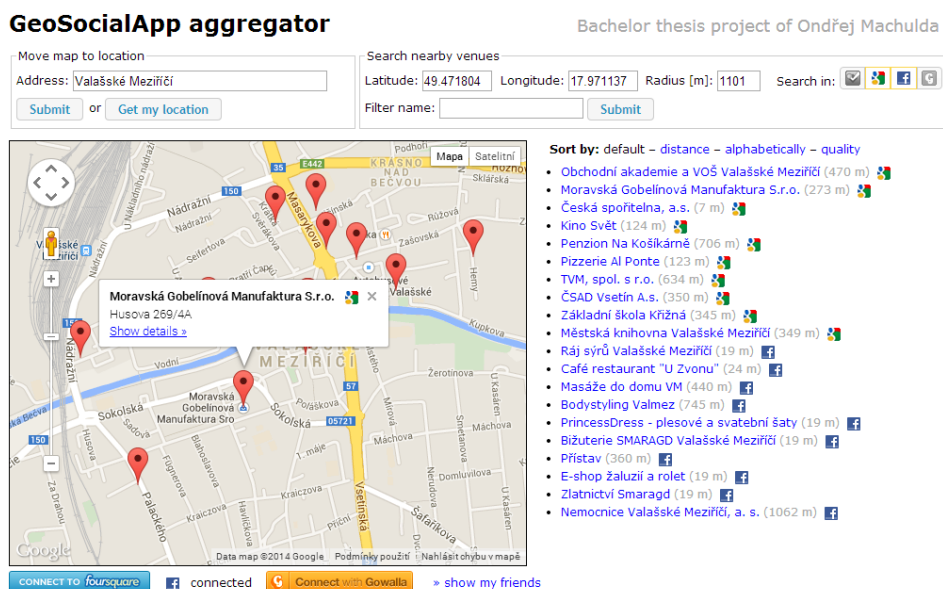
<sup>11</sup> <http://geographics.cz/socialMap/>



Obr. 6.5 Prostředí aplikace Placebook a demografické statistiky všech uživatelů Facebooku

## 6.1.4 GeoSocialApp aggregator<sup>12</sup>

Další možností zobrazení geodat z více sociálních sítí současně je pomocí aplikace Ondřeje Machuldy. Tato aplikace vznikla jako součást bakalářské práce v roce 2012. Jde o agregační službu umožňující zobrazení zájmových bodů ze sociálních sítí Foursquare, Google Places, Facebook a Gowalla na daném místě (Machulda, 2012). Možnost zadat lokaci lze dvěma způsoby a to buď přímým zadáním adresy nebo zadáním souřadnic. Příkladem funkčnosti aplikace je dotaz na objekty vyskytující se v oblasti Valašské Meziříčí (obr. 6.6).



Obr. 6.6 Prostředí online aplikace GeoSocialApp aggregator (zdroj: vlastní)

<sup>12</sup> <http://gsaa.vlastovka.eu/>

## 6.2 Geokódování

V případě nevyužití některé z možností online znázornění dat ze sociálních sítí a přistoupení k data miningu dat, existují dva hlavní způsoby zaznamenání polohy zprávy nebo uživatele. U sítí s možností přímé geolokace je to nejčastěji prostřednictvím GPS souřadnic, které se skládají ze zeměpisné šířky (*latitude*) a délky (*longitude*). Po získání vybraných dat je tak možnost přímé vizualizace prostřednictvím některého z GIS softwarů.

Druhým případem je zaznamenání místa určení v podobě textu. V tomto případě je potřeba využít tzv. geokódování, čímž se rozumí dvojice nabízených služeb: hledání zeměpisné pozice dle zadaného řetězce (dopředné geokódování), které je využito i v rámci práce, nebo hledání zeměpisných objektů na dané souřadnici (zpětné geokódování) (API Mapy.cz, 2014)

Nástrojů sloužících ke geokódování existuje celá řada. Může jít o webové aplikace, služby či desktopové programy. Častým řešením a relativně snadným způsobem k určení souřadnic zeměpisných objektů je pak využití API mapových portálů. Bohužel, většina API mapových služeb umožňuje pouze omezený počet dotazů. Výjimkou je API českého mapového portálu Mapy.cz. Zde sice lze vyhledávat až na úroveň adres s číslem popisným, nicméně je omezen pouze na české názvy. Geokódování s využitím Google Geocoding API je vhodným způsobem získání zeměpisných souřadnic jednotlivých domů s číslem popisným, ovšem omezení počtu dotazů na 2500 denně je značně limitující pro obsáhlé datové sady. Další možností geokódování je využití nástroje OpenStreetMap Nominatim, který sice nemá žádný denní limit, nicméně podle podmínek užívání není doporučen ke geokódování velkých objemů dat (OpenStreetMap Nominatim, 2014).

V případě potřeby geokódování velkého množství dat je jednou z možností využít databáze GeoNames. Tato databáze obsahuje více než deset milionů názvů, které jsou členěny do různých kategorií (země, města, hory, poštovní směrovací čísla, ...) (Geonames, 2014).

U geolokace pomocí textových řetězců v sociálních sítích však často dochází k chybám. Pokud prostorový prvek vyplňuje přímo uživatel, snadno se dopustí chyb jako jsou překlepy, nepřesnosti nebo různé zkratky názvů (např. Valašské Meziříčí jako „Valmez“). Často také uživatelé uvádějí jako místa pobytu neexistující či imaginární místa nebo místa s citovým zabarvením. (např. Peklo, V sedmém nebi, Za sedmero horami, ...). Geokódovat tyto názvy je pak často nemožné. Posledním častým nedostatkem u geokódování je existence několika míst se stejnými názvy. V případech malých množství dat je možné výsledné souřadnice snadno zkontrolovat, u velkých dat však chyby mnohdy odstranit nelze.

## 7 DOLOVÁNÍ DAT ZE SOCIÁLNÍ SÍTĚ TWITTER

Jak už bylo uvedeno v kapitole 4.3.1, jedinou možností jak získat data ze sociální sítě Twitter je přes jeho API. K dolování dat je potřeba se autorizovat pomocí protokolu OAuth. API twitteru je rozděleno do 3 částí, *Search API*, *Streaming API* a *REST API*. Pro získání dat se dají použít první dva případy. Pro práci s daty a získání dat je třeba se zaregistrovat jako vývojář, vytvořit a popsat aplikaci, pro kterou data potřebujete. Po zaregistrování uživatel obdrží dva kódy: *consumer Key* a *consumer secret*, díky kterým již může data dolovat. Stejně jako u mnoha jiných sociálních sítí, je pro práci s daty u Twitteru vytvořeno několik knihoven pro práci. Jsou vytvořeny pro různé programovací jazyky a jsou vestavěny jak pro práci s daty, tak pro práci s autorizačním protokolem *OAuth*.

Twitter je jednou z největších sociálních sítí dnešní doby. Lidé zde píší nejen o svých zážitcích, ale často se zde vyjadřují k aktuálnímu dění kolem nich, nebo ve světě. Proto je i vhodným prostředím pro případové studie. Tématem studie však musí být oblast, která zajímá široké množství lidí tak, aby získaná data byla vhodná pro prostorové analýzy. Příkladem je třeba událost Olympijských her. Toto téma zajímá nejen samotné sportovce a příznivce daného sportu, ale vzhledem k velikosti akce také novináře, politiky a v neposlední řadě i ostatní veřejnost. O olympijských hrách se ale nemluví jen v pozitivním slova smyslu. Velké množství lidí díky ohromným výdajům tyto události odsuzuje a v případě Olympijských her 2014 je spousta negativních ohlasů mířena i na pořádající zemi – Rusko.

Důvody popsané v závěru kapitoly 7.1.1 vysvětlují, proč studie ohledně olympijských her nebyla úspěšná. Proto bylo třeba vybrat jiné téma. Nepříznivá situace na Ukrajině se vyvinula v tzv. „Ukrajinskou krizi“. Události na poloostrově Krym a vměšování Ruska do těchto záležitostí vyvolalo v prostředích sociální sítí nevídaný zájem. Mapování této oblasti tak naskytlo skvělou možnost pro případovou studii.

### 7.1 Twitter a prostředí R

Pro práci s daty sociální sítě Twitter jsou vytvořeny i knihovny do prostředí jazyka R. Vytvořeny jsou knihovny pro získání dat jak z části *Search API*, tak z části *Streaming API*. Pro autorizaci je možnost využít balíček *ROAuth*.

#### 7.1.1 Získání dat Search API

Pro práci s částí *Search API* je vytvořena knihovna *twitteR*, jejímž autorem je Jeff Gentry (Cran – Package *twitteR*, 2014). Knihovna obsahuje 18 funkcí, které jsou vyjmenovány a popsány v tabulce 7.1. Pro získání tweetů je třeba se na začátku procesu autorizovat. Jak získat autorizační klíče bylo popsáno v kapitole 4.3.1. Pro práci v operačním systému Windows je třeba pak při každém dotazu volat také argument *cainfo*, který obsahuje dokument *cacert.pem*.

Tab 7.1 Funkce knihovny twitterR v prostředí programového jazyka R

Funkce	Popis funkce
decode_short_url	Najde původní adresu URL, která byla upravena zkracovačem
directMessage-class	Poskytne model představující přímé zprávy z Twitteru
dmGet	Umožňuje posílat a mazat přímé zprávy na Twitteru
favorites	Získá nejoblíbenější tweety od specifických uživatelů
friendships	Získá list následovníků (followers) přihlášeného uživatele
getCurRateLimitInfo	Načte aktuální limit pro získání tweetů pro přihlášeného autorizovaného uživatele
getTrends	Najde aktuální trendy v síti
getUser	Získá veřejné informace o daném uživateli Twitter
import_statuses	Funkce určená k importu dat do Twitteru z různých datových zdrojů
registerTwitterOAuth	Získá přístupové klíče z aplikace a ověří autorizaci uživatele
searchTwitter	Vyhledá Tweety podle zadaného textového řetězce
showStatus	Funkce navrácí jeden specifický klíč podle identifikačního čísla
status-class	Navrací další informace o daném tweetu
taskStatus	Funkce odesílající přímou zprávu vybranému uživateli a informace o jejím úspěšném, nebo neúspěšném odeslání
timelines	Navrací timeline uživatele
twListToDF	Převádí Twitter seznamy do data.frame
updateStatus	Zveřejní prostřednictvím aplikace status na Twitteru
user-class	Upravuje informace uživatele na Twitteru

Pro získání veřejných tweetů ostatních uživatelů z části Search API se využívá funkce *searchTwitter*. Možnost získat tweety lze maximálně osm dní staré. Argumentů funkce je devět (tab 7.2). Pro získání geolokovaných tweetů je potřeba využít argumentu *geocode*. Pro získání příspěvků z dané země je třeba podle kódování ISO 3166-1 zadat argument *locale*.

Příklad získání padesáti tweetů o Olomouci z oblasti Francie a uložení do proměnné *fr\_olomouc*:

```
Fr_olomouc<-searchTwitter('olomouc', locale="FR", n=50,
cainfo="cacert.pem")
```

Příklad získání Tweetů s hashtagem #ukraine v oblasti sto kilometrů od centra města Kyjev a uložení do proměnné *uk100* :

```
uk100<-searchTwitter('#ukraine', geocode=30.52,50.46,100km',
cainfo="cacert.pem")
```

V případě nevyužití argumentu  $n$ , či-li počet žádaných tweetů, se vrátí maximální možný počet tweetů podle aktuálního API limitu. Často je tak reakcí na dotaz varující zpráva, která zaznamená, že pro zvolenou oblast je dostupných pouze 303 tweetů, ačkoli bylo požadováno maximum, tedy 500 :

Warning message:

```
In doRppAPICall("search/tweets", n, params = params, retryOnRateLimit =
retryOnRateLimit, :
```

```
500 tweets were requested but the API can only return 303
```

Tab 7.2 Argumenty funkce searchTwitter knihovny twitterR

Argument	Popis
searchString	zadání textového řetězce
n	žádaný počet tweetů
lang	jazyk tweetu
since	od jaké doby chceme tweety získat
until	do kterého data chceme tweety získat
locale	země ze které tweet pochází
geocode	vrací tweety od uživatelů v určitém okruhu okolo dané zeměpisné šířky
sinceID	vrací tweety od daného uživatele
retryOnRateLimit	zobrazuje aktuální limit pro získání tweetů

Jak uvádí Karimi (2014) geolokovaných tweetů je však v prostředí sociální sítě Twitter pouze 0,5 – 3 %. V situaci, kdy funkce *searchTwitter* vrací pouze informace, které byly zadány jako argumenty (text tweetu, uživatelské jméno uživatele, případně jazyk nebo geolokaci), není tato funkce a část Search API vhodným způsobem získávání dat pro následnou vizualizaci a analýzu. Příkladem byl pokus o sběr dat o zimních olympijských hrách v Rusku v období 6. 2. 2014 – 23. 2. 2014. Počty získaných tweetů se pohybovaly sotva v řádech stovek za den a většina z nich byla z USA, mnohé z nich navíc jen retweetované. To je dané filosofií Twitteru a funkcí searchTwitter, kdy výsledkem dotazu uživatele není celkové množství tweetů, ale především ty nejoblíbenější.

### 7.1.2 Získání dat Streaming API

Dalším způsobem získání dat z Twitteru je přes Streaming API. V prostředí R je k tomu určena knihovna *streamR* od autora Pablo Barbera, která má 5 funkcí (tab 7.3). I v tomto případě je potřeba autorizačních kódů, které vývojář získá při zaregistrování aplikace. Tato možnost získává data v reálném čase s omezením získání 1 % tweetů, což je dáno podmínkami a funkcionalitou API.

Tab. 7.3 Funkce knihovny *streamR* v prostředí jazyka R

Funkce	Popis
example_tweets	Získá deset ukázkových tweetů publikovaných uživatelem @twitterapi
filterStream	Vyhledává tweety v reálném čase podle textového řetězce a zadaných kritérií
parseTweets	Funkce, která parsuje získané tweety z JSON souboru
readTweets	Převádí získané tweety z JSON formátu do prostředí R
sampleStream	Připojuje se k streaming API a vrací malé množství veřejných statusů
userStream	Získává tweety od jednoho konkrétního uživatele

Pro získání tweetů v reálném čase je určena funkce *filterStream*. Tato funkce má devět argumentů pro získání konkrétnějších příspěvků. V tabulce 7.4 jsou jednotlivé argumenty popsány. Z hlediska prostorových informací je významným argumentem *locations*. Tento argument nám vrací informace o zeměpisné šířce a délce. Získané tweety jsou uloženy do JSON souborů.

Tab. 7.4 Argumenty funkce *filterStream*

Argument	Popis
file.name	Jméno souboru, do kterého se mají vytěžené tweety uložit
track	Textový řetězec pro získání tweetu na dané téma
follow	Textový nebo číselný řetězec s ID uživatele, od kterého se mají tweety získávat
locations	Číselný formát obsahující zeměpisnou délku a šířku pro získání tweetů z určité oblasti
language	Získání tweetů napsaných v daném jazyce
timeout	Časové období získání tweetů ve vteřinách
tweets	Číslo udávající maximální počet získaných tweetů
oauth	Autorizační objekt obsahující přístupové klíče
verbose	Generuje výstup do prostředí R o informacích o probíhajícímu procesu

Příklad získání tweetů během jedné hodiny s hashtagy týkající se krize na Ukrajině a odtržení poloostrova Krym od Ukrajiny a uložení výsledků do souboru `ukraine1603.json`:

```
filterStream(file.name="ukraine1603.json", track=c("#crimea",
"#ukraine"), timeout=3600, oauth=twitCred)
```

Výsledkem je tak JSON soubor, který obsahuje textový řetězec s atributy jednotlivých tweetů. Výhodou této funkce je získání nejen textu konkrétního tweetu, ale také všechny dostupné veřejné informace o uživateli, který tweet publikoval.



Pro další práci v prostředí R je možnost využít funkce *parseTweets*, která jednotlivé atributy tweetu převede z formátu JSON do podoby `data.frame` v aplikaci, se kterým lze snadno pracovat v prostředí R.

```
uk1603.df<-parseTweets("ucraine1603.json")
```

Tweet získaný prostřednictvím Streaming API bez omezení počtu atributů obsahuje 42 proměnných vypsanych v tabulce 7.4. Ne všechny tweety ale obsahují každou z proměnných. V případě, že uživatel daný atribut nevyplnil, nebo ho nemá veřejný, v prostředí R se zobrazí tato proměnná jako NA.

Tab. 7.4 Proměnné jednoho tweetu získané funkcí `filterStream`

Proměnná	Datový typ	Proměnná	Datový typ
country	character	place_id	logical
country_code	character	place_lat	numeric
created_at	character	place_lon	numeric
description	character	place_name	logical
expanded_url	character	place_type	character
favorited	logical	protected	logical
favourites_count	numeric	retweet_count	numeric
followers_count	numeric	retweeted	logical
friends_count	numeric	screen_name	character
full_name	character	source	character
geo_enabled	logical	statuses_count	numeric
id_str	character	text	character
in_reply_to_screen_name	character	time_zone	character
in_reply_to_status_id_str	character	truncated	logical
in_replyto_user_id_str	character	url	character
lang	character	user_created_at	character
lat	numeric	user_id_str	character
listed_count	numeric	user_lang	character
location	character	user_url	character
lon	numeric	utc_offset	numeric
name	character	verified	logical

Geolokačními atributy jsou proměnné *lat* a *lon*, které jsou vytvořeny v číselném datovém formátu. Kromě geolokačních parametrů jsou zde také další prostorové atributy, nicméně zapsány textovým řetězcem. Nejvýznamnější z nich je *location*, který udává zvolenou pozici autora tweetu.

## 7.2 Krize na Ukrajině a Twitter

Pro získání tweetů o dění na Ukrajině je vhodné zvolit funkci `filterStream`, díky které dostaneme nejen velké množství příspěvků, ale také informace o příspěvovatelích, díky čemuž lze provádět vizualizace a analýzy.

První částí je načtení potřebných knihoven (`streamR`, `sqldf`, `httr`, `rjson`, `plyr`, `tidy`, `rgdal`). Dalším krokem je proces autorizace pomocí protokolu `OAuth`. Následně může proběhnout sběr dat s klíčovými slovy a požadavkem na geolokaci. Posledním krokem je vytvoření souboru GeoJSON. Úplný kód pro získání dat je součástí přílohy 9.

Sběr dat probíhal v období 10. – 30. 3. 2014, tedy v době odtržení polostrova Krym od Ukrajiny. I v tomto případě se však ukázalo, že počet geolokovaných tweetů je opravdu pouze 1 % z celkového počtu tweetů s daným tématem. Díky zvolení metody přes `Streaming API` je však možnost geokódovat umístění uživatele podle proměnné `location`. Na rozdíl od předešlého příkladu je ale zájem získat všechny tweety, bez omezení parametrem `location`, který vrací pouze geolokované tweety.

Geokódování velkého množství záznamů je však problematické. Většina možností geokódování má omezený počet dotazů. Proto bylo vybráno geokódování přes nominativ u OpenStreetMap. JSON soubory jednotlivých dnů se musely rozdělit na menší celky. Až poté mohlo být geokódování provedeno. Možnost geokódování je i v prostředí jazyka R. V případě této práce byla vytvořena funkce `finder` (Příloha 6).

Pro samotné geokódování je pak využito funkce `lapply` aplikované na prostorový prvek `location`.

```
Guk3003.df<-with(uk3003, data.frame(text, source, retweeted,
                                   created_at, lang, location, user_id_str,
                                   description, user_url, name, time_zone,
                                   user_lang, screen_name, country_code, country,
                                   full_name, lat, lon, expanded_url, url,
                                   lapply(location, function(val){finder(val)})))
```

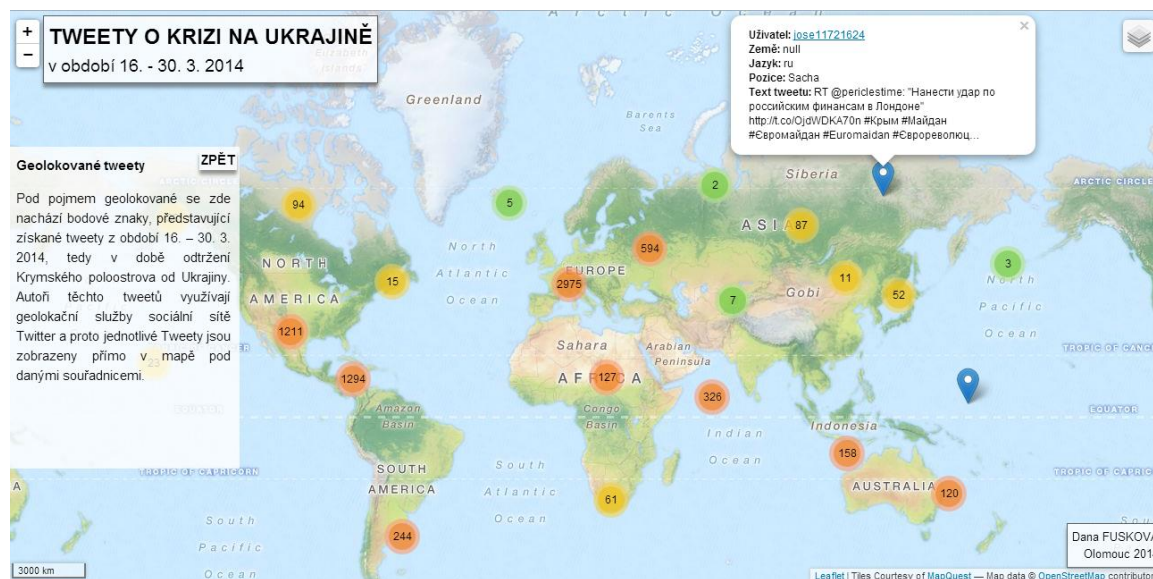
Výsledkem geokódování je pak opět datová sada rozšířena o dva atributy (`lat3`, `lon3`). Pro případné vizualizace ve webových mapách je třeba ještě data převést do formátu `GeoJSON`. Jednotlivé soubory dat mají však velkou velikost (0,5 GB – 1,5 GB pro data jednoho dne), proto musela být data ještě ořezána o méně významné atributy, aby došlo ke zmenšení.

Díky ne příliš přesným pravidlům Twitteru ohledně obsahu textu muselo před samotnou vizualizací dojít k dalším úpravám. V prvním případě musely být překódovány různé jazyky do latinky. Jednalo se především o azbuku. Toto bylo vyřešeno rekódováním textů do UTF-8. V textech tweetů však byly často také znaky s motivačním nábojem, které následný export nemohl zpracovat, a tak bylo třeba takové tweety odstranit. Podobný problém byl i v atributu `location`, díky kterému probíhalo geokódování. V této části však byly také problémy s tím, že tento atribut není nijak kontrolován a uživatelé si tam tak můžou napsat cokoli. Uživatelé tak uváděli jako místo

svého pobytu například „In heaven“, „Alien from Marse“, „End of the World“. Jediným bezproblémovým parametrem byl kód země (country\_id), kvůli formátu ISO 3166-1, díky kterému se dají dělat např. kartogramy.

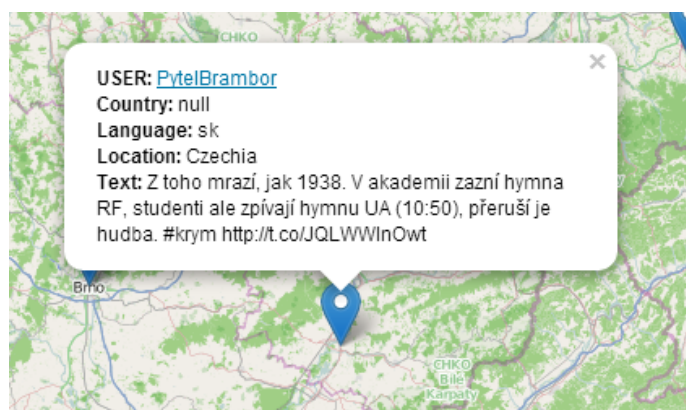
### 7.3 Vizualizace dat o Ukrajině

Získané tweety o Ukrajině uložené v GeoJSON souborech jsou bodové prvky s geolokačními atributy zeměpisná šířka a zeměpisná délka. Data se tak dají umístit do mapy. Jedním ze způsobů je využití open-source knihovny Leaflet (obr 7.1).



Obr. 7. 1 Ukázka mapy tweetů o krizi na Ukrajině (zdroj: vlastní)

Pro načtení GeoJSON souborů do prostředí Leaflet bylo využito JavaScriptové knihovny *GeoJSON.js*. Vzhledem k velkému množství tweetů v blízkých lokalitách, převážně pak u těch geokódovaných, bylo využito knihovny *markercluster.js*. Důležitým požadavkem bylo zobrazení vybraných informací o každém tweetu, proto byla vytvořena funkce *popUp*, která zobrazuje vybrané informace ke každému příspěvku. Příklad je zobrazen na obrázku 7.2.



Obr. 7.2 Příklad popUp okna u jednoho tweetu.

Pro popis mapy a intuitivní zacházení byl v rámci mapy vytvořen postranní sidebar pomocí vytvořeného JavaScriptu *sidebar.js*. Výsledkem je tak zobrazení nejen umístění tweetů, ale také další informace jednotlivého tweetu (Příloha 14).

## 7.4 Výsledky případových studií

Pomocí části Streaming API Twitteru bylo získáno ohromné množství příspěvků ze všech oblastí světa. Následná úprava dat byla sice náročná vzhledem k velkým objemům dat a neustále přibývajícím množství problémů s těmito daty, výsledná upravená data však přinesla mnoho zajímavých výsledků. Z grafu (obr 7.3) lze vidět zájem o problematiku na Ukrajině ve vybraných dnech. Je překvapivé, že největší počet získaných příspěvků není z 16. března, tedy ze dne, kdy obyvatelé Krymu rozhodli v referendu o odtržení Krymského poloostrova od Ukrajiny a připojení k Ruské federaci. Nejvíce setweetovalo 20. března, tedy v době, kdy se sešel americký ministr obrany Chuck Hagel s ruským ministrem obrany Sergejem Šojgou. Významným dnem z hlediska zveřejňování příspěvků o Ukrajině je ještě 25. březen, kdy se ruská vláda dohodla s představenstvem Ukrajiny o odchodu ukrajinských vojáků z Krymu.



Obr. 7.3 Počty tweetů v období 16. – 30. 3. 2014 (zdroj: vlastní)

Z tabulky počtu tweetů z jednotlivých zemí (Příloha 13) není překvapivým výsledkem, že největší absolutní počet tweetů pochází z USA. Sociální síť Twitter však není ve všech zemích stejně oblíbená (tab 2.1), proto pro správné vyhodnocení dat je vhodné data vztáhnout k počtu uživatelů. Při tomto přepočtu převládá významně samotná Ukrajina. To samozřejmě vzhledem k událostem je očekávaný výsledek. Na další pozici je Rusko. Zvláštní by se mohlo zdát zvýšené množství tweetů ze států Polsko, Slovensko nebo Rumunsko. Vzhledem k jejich těsné pozici s Ukrajinou to však také není překvapivý výsledek.

## 8 DOLOVÁNÍ DAT ZE SOCIÁLNÍ SÍTĚ FACEBOOK

Sociální síť Facebook byla více popsána v kapitole 4.3.2. Získání dat z této sítě je možné provést prostřednictvím Facebook API nazvaného *Graph API*, které je na velmi pokročilé úrovni. Kromě dobré dokumentace, je pro něj vytvořeno mnoho nástrojů pro vývojáře s velkými možnostmi pro přístupy k datům. Aby bylo možné získat data prostřednictvím *Graph API*, tak je nutné být autorizován. K dispozici jsou dva způsoby autorizace.

První možností je autorizační protokol *OAuth* vyžadující dva kódy: *app\_id* a *app\_secret*. Jeho výhodou je, že proces autorizace není s těmito přístupovými kódy vázán na časové období a pouze je třeba se řídit licenčními podmínkami Facebooku, týkajícími se počtu dotazů. Druhou možností je snadnější přihlášení pomocí klíče nazývaného *Access token*. Tento klíč je pouze jednorázový a poskytne uživateli rychlý přístup k datům. Jeho platnost je ale omezena pouze na dvě hodiny. Po uplynutí této doby lze ovšem snadno požádat a následně získat další klíč.

### 8.1 Facebook a prostředí R

V prostředí R je pro práci s daty Facebooku vytvořena knihovna *rFacebook* od autora Pablo Barbera. Tato knihovna má 11 funkcí pro práci s *Graph API* Facebook (tab. 8.1).

Tab 8.1 Funkce knihovny *rFacebook*

Funkce	Popis
fbOAuth	Slouží k přístupu k API Facebook přes protokol OAuth
getCheckins	Získá seznamy geolokací přátel na Facebooku
getFriends	Získá seznam přátel uživatele
getLikes	Získá seznam oblíbených stránek daného uživatele
getNetwork	Vytvoří síť vztahů mezi přáteli
getNewsfeed	Zobrazí příspěvky, které uživatel vidí v prostředí timeline
getPage	Získá příspěvky z libovolné veřejné Facebook stránky
getPost	Získá informace o daném veřejném příspěvku
getUsers	Vydoluje informace o daném uživateli
searchFacebook	Vyhledá veřejné příspěvky na dané téma
updateStatus	Přidá status na profil uživatele

Jako vhodný způsob pro získání informací o pohybu přátel uživatele se zdá být funkce *getCheckins*. Tato služba ale vrací pouze hodnoty polohy vytvořené při krátkém spuštění samostatné geolokační služby Facebook Places v roce 2010.

Zajímavou funkcí je také *getNetwork*, která získá informace o propojenosti přátel autorizovaného uživatele mezi sebou. Je tak umožněno získat informace o tom, kdo se s kým zná prostřednictvím Facebooku.

Všechny ostatní funkce získávají informace o stránkách, lidech nebo příspěvcích. Jak už bylo dříve zmíněno, geolokační prvek je v prostředí sítě Facebook přítomen pouze jako doplňující atribut k danému příspěvku. Bohužel Graph API je více zaměřené na vztahy mezi jednotlivými prvky než na prostorové prvky, a proto jedinou možností, jak získat informaci o poloze, je zaměřit se na předem známé geolokované příspěvky. Tento způsob však není efektivní k získání většího množství dat.

### 8.1.1 Získání informací o přátelích

Jednou ze dvou možností jak přes Graph API prostřednictvím jazyka R získat prostorová data, je přes podrobné informace jednotlivých přátel přihlášeného uživatele. Každý z přátel má možnost při vyplňování informací ve svém profilu vyplnit hned dva prostorové prvky. Jsou to informace o tom, odkud uživatel pochází (atribut *hometown*) a informace o tom, kde uživatel aktuálně žije (atribut *location*).

Příklad získání informací o uživateli s přístupovým klíčem a uložení do proměnných *ja*, *pratele* a *pratele\_info*:

```
library("Rfacebook")

token<- "XXXXXXXXXXXXXXXXXXXX"
ja <- getUsers("dancaf", token, private_info = TRUE)
pratele <- getFriends(token, simplify = TRUE)
pratele_info <- getUsers(pratele$id, token, private_info = TRUE)
```

Výsledkem funkce *getFriends* je databáze přátel autorizovaného uživatele s atributy *name* a *id*. Pro zjištění podrobných informací o těchto přátelích je třeba připojit informace, které získáme prostřednictvím funkce *getUsers*.

Informace, které tímto způsobem získáme, jsou data o daných přátelích. Celkem můžeme získat 14 proměnných o jednom uživateli (tab 8.2).

Získaná prostorová data (*hometown*, *location*) jsou však pouze textovým řetězcem. Pro následnou vizualizaci je tak třeba tyto prvky geokódovat. V případě přátel pouze z České republiky, by bylo vhodnějším způsobem využít geokódování prostřednictvím API *Mapy.cz*. V případě uživatelky Danča Fusková (autorky práce), která má více než 100 přátel ze zahraničí, to však nestačí. Při počtu záznamů do 1000 je příjemným, rychlým a efektivním způsobem využití R balíčku *ggmap*, který umožňuje geokódování pomocí Google Geocoding API.

Tab 8. 2 Získané informace prostřednictvím funkce *getUsers*

Název	Popis
id	identifikační číslo uživatele
name	jméno uživatele
username	uživatelské jméno
first_name	křestní jméno
last_name	příjmení
gender	pohlaví
locale	lokace a jazyk
likes	oblíbenost
picture	odkaz na profilový obrázek
birthday	datum narození
location	místo
hometown	rodné město
relationship_status	rodinný stav

### 8.1.2 Zisk veřejných příspěvků s prostorovým atributem

U zisku veřejných informací vztahených k prostoru je důležité zvolit správné klíčové slovo. V případě použití i ustáleného mezinárodního výrazu je téměř v každém případě největší počet příspěvků z anglicky mluvících zemí. Proto v případě analýz z této sociální sítě je lepším případem použití jména nadnárodního charakteru.

Druhou možností je zisk veřejných příspěvků na dané klíčové slovo a připojení informace o umístění uživatele, který tento příspěvek napsal. K získání veřejných statusů slouží funkce *searchFacebook*.

Výsledkem je databáze záznamů s deseti proměnnými (tab. 8.3). V případě nevyužití argumentu počtu příspěvků, funkce bude prohledávat pouze deset vteřin. V případě žádostí o větší množství, než API Facebooku je schopno poskytnout, vrátí se menší počet dotazů.

Ukázka kódu k získání veřejných příspěvků s klíčovým slovem *Putin*, zisk informací o uživateli, kteří tyto příspěvky publikovali a připojení informace o umístění uživatele:

```

Putin<-searchFacebook("Putin", token, n=5000)
P_users <- getUsers(Putin$from_id, token)
names(P_users)[1] <- "from_id"
Putin_data <- merge(Putin, P_users, by="from_id")
Putin_data$country <- substr(Putin_data$locale, 4, 5)

```

Tab 8.3 Proměnné výsledku hledání searchFacebook

<b>název proměnné</b>	<b>popis</b>
from_id	ID autora
from_name	jméno autora příspěvku
message	text zprávy
created_time	čas publikování
type	typ příspěvku (status, video, fotografie)
link	odkaz v případě fotografie, nebo odkazu
id	id příspěvku
likes_count	počet lajků
comments_count	počet okomentování
shares_count	počet sdílení

Umístění uživatele, které takto získáme, ale není geolokační službou, ale pouze získá informace o kódu země, kde se autor příspěvku nachází z atributu *locale*, který se skládá také z informace o jazyku. Jazyk zde ale není myšlen jako jazyk, ve kterém byl příspěvek napsán, ale jako nastavení prostředí sociální sítě Facebook.

## 8.2 Vizualizace dat ze sociální sítě Facebook

Pro vizualizaci dat je i v tomto případě vhodné využít prostředí Leaflet. Po aplikaci geokódování na prostorové prvky získané knihovnou rFacebook je vhodné data převést do formátu GeoJSON. Díky následnému využití skriptů GeoJSON.js jsou pak data načtena do prostředí Leafletu.

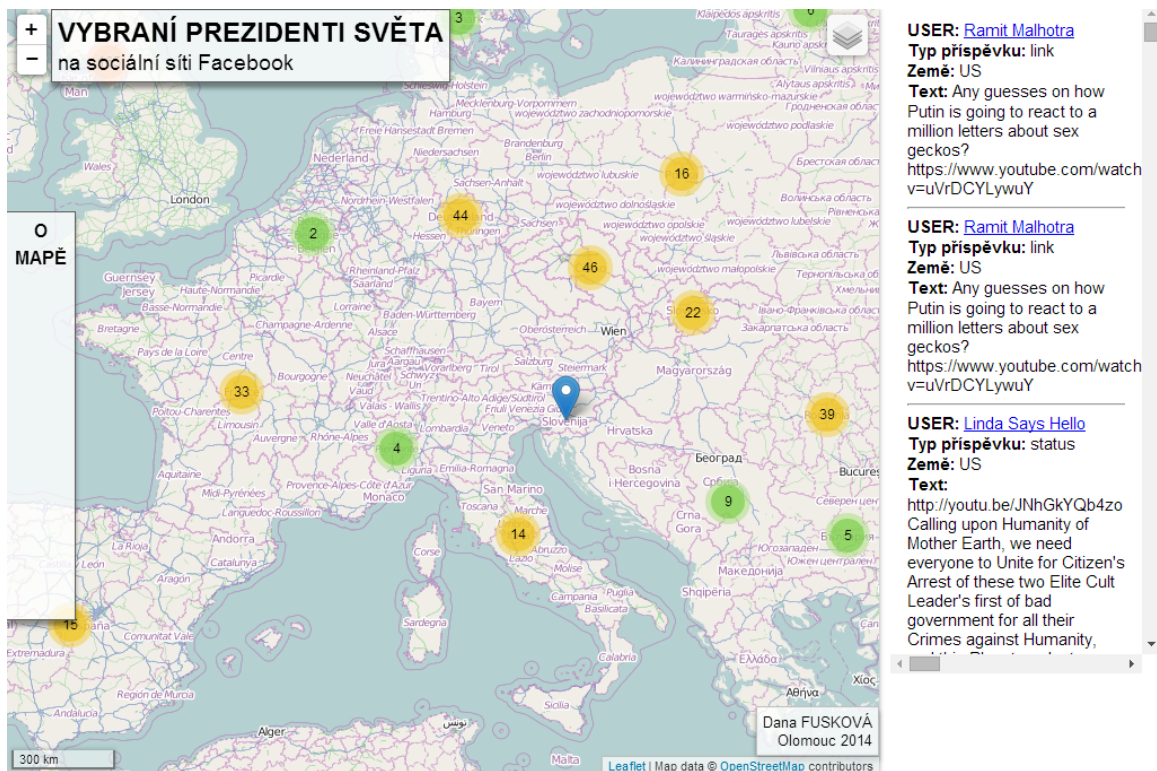
V případě prostorových informací o přátelích uživatelky Danča Fusková byly vytvořeny dva GeoJSON soubory s informacemi o aktuální poloze přátel a o rodném městě. Pro zobrazení informací o přátelích bylo i v tomto případě využito popUp okna. Pro popis informací v mapě slouží postranní sidebar (obr 8.1).





Obr. 8. 1 Ukázka vytvořené mapové aplikace zaměřené na přátele v síti Facebook (zdroj: vlastní)

U veřejných příspěvků bylo umístění uživatele získáno ze zkratky jednotlivé země. Dochází zde k zobrazení všech příspěvků dané země v jednom bodě. Pro tento případ tedy bylo využito možnosti zobrazit informace ze všech příspěvků do postranního panelu (obr 8.2).



Obr. 8. 2 Ukázka mapového prostředí se zaměřením na vybrané prezidenty

### 8.3 Výsledky

Grafickým výsledkem v tomto případě jsou dvě webové mapy vizualizované pomocí skriptu Leaflet. Zobrazují kromě umístění také další informace o uživatelích, nebo o příspěvcích, které publikovali (Příloha 15, 16).

Vizualizace mapy přátel uživatelky Danča Fusková přineslo výsledek ohledně koncentrace jejích přátel. Uživatelka má přátele koncentrované v pěti místech světa. Nejvíce přátel v obou případech prostorových dat (lokace i rodné město) má v místě jejího rodného města a současně stále místem trvalého bydliště ve Valašském Meziříčí. Další výrazný počet přátel je v hlavním městě České republiky, v Praze. Tato hodnota se dá vysvětlit koncentrací přátel z kanoistiky a z pionýra. Zajímavým výsledkem je Norské město Trondheim, ve kterém uvádí hned 35 přátel místo současné pozice. Rodné město je to však pouze pro jednoho z nich. Výsledek pro uživatelku, která toto město navštívila v rámci studijního pobytu Erasmus, není překvapením, vzhledem ke sběru dat v období školního roku.

## 9 VÝSLEDKY

Tato bakalářská práce se zabývala geoinformatickými daty v prostředích sociálních sítí. Jedním z cílů práce bylo zvolení vhodných metod pro získání prostorových dat z prostředí sociálních sítí. Hlavním výsledkem tak jsou praktické postupy, které obsahují návody jak z vybraných sociálních sítí data získat, upravit je a následně vizualizovat (Příloha 14, 15, 16).

První část práce se věnovala pojmu sociální síť a popisu typů prostorových dat v nich. Na základě získaných informací byly vybrány tři sociální sítě vhodné pro data mining: Twitter, Facebook a LinkedIn. U těchto sítí byla popsána jejich filosofie, historie, prostředí a API rozhraní.

V praktické části bylo využito možnosti data miningu přes API rozhraní jednotlivých sítí. V případě sociální sítě Twitter bylo využito programovacího jazyka a prostředí R.

Možnost autorizace do API rozhraní jednotlivých sítí umožňovala knihovna OAuth. V kapitole 4.3 je nastíněn proces autorizace pro každou z vybraných sociálních sítí a potřebné klíče pro přihlášení. Pomocí funkce OAuthFactory pak bylo využito tohoto procesu autorizace u zmiňované sociální sítě Twitter. Součástí práce je příloha 9, která obsahuje popsaný skript jak získat data prostřednictvím Streaming API a knihovny StreamR. Výsledkem této studie byl získání dvaceti JSON souborů obsahující tweety s tématem Krize na Ukrajině v období 10. – 30. 3. 2014. Tyto soubory obsahovaly nejen samotné texty jednotlivých tweetů, ale také všechny dostupné veřejné informace, které se daly získat. Výsledkem je velké množství dat. Za vizualizované období 16. – 30. 3. bylo získáno celkem 262 493 tweetů (Příloha 1). Velké množství dat však znamená také velký objem dat. Pro cílovou vizualizaci tak data byly minifikována o zbytečné mezery i o nevizualizované argumenty. Jak uvádí tabulka v příloze 1, počet příspěvků, u kterých sám uživatel zvolí možnost přidání polohy je opravdu mizivá. V případě analýzy těchto dat se počet pohyboval okolo 0,008 %. Nepotvrdil se tak výrok Karimi (2014) ohledně 1 % - 3 %. Důvodem však může být i omezení přes API Twitteru. V případě tweetů, které se podařilo zgeokódovat se hodnota pohybovala okolo 0,5 %.

Z této studie tak plyne, že získávat data z Twitteru prostřednictvím Streaming API má smysl. Práce je sice velmi časově náročná, ale získané výsledky mají velkou vypovídající hodnotu.

Prostorová data, která se prostřednictvím jeho API dají vydolovat jsou dvojího druhu. Prvním případem jsou to prostorové informace o přátelích uživatele s autorizací. Data jsou však pouze textovým atributem, proto je třeba je dále geokódovat.

Druhým případem je informace o zemi autora veřejného příspěvku. Díky této vlastnosti tak máme možnost získat představu o umístění autora příspěvku. Sociální síť Facebook je ale založena více na sdílení informací mezi známými přáteli. Většina příspěvků proto není veřejná a vytvářet prostorové analýzy prostřednictvím takto získaných dat tak není vhodné.

Sociální síť LinkedIn je z těchto vybraných sítí nejmladší. To je také důvodem k menší propracovanosti API na rozdíl od výše zmiňovaných sítí. Vzhledem k charakteru informací v tomto prostředí, zde také ani prostorová data nehrají velkou roli. Získ geodat prostřednictvím API je proto v této síti nevhodný.

## 10 DISKUZE

Ačkoli jsou sociální sítě populárním tématem dnešní doby a prostorová data jsou součástí téměř každé z nich, není toto téma v českém jazyce příliš rozebíráno. Výjimkou jsou diplomové práce studentů převážně inženýrských oborů, kteří se však problematikou nezabývají z geografického ani kartografického hlediska. Proto pro teoretickou část práce muselo být čerpáno převážně ze zahraničních zdrojů.

Sociálních sítí je však velká spousta a každá je svým charakterem a zaměřením hodně odlišná. Proto teoretická část s tímto tématem musela popsat informace hodně zešíroka. Pro lepší orientaci v problematice byly proto vybrány tři sociální sítě – Twitter, Facebook a LinkedIn. Motivace výběru zrovna těchto sítí byla popsána v kapitole 2.1.

Prvním a důležitým krokem tak bylo seznámit se podrobně s těmito sítěmi, jejich filosofií a prostředím. Následným krokem byla aplikace teoretických znalostí při výběru způsobu možností získání dat z těchto sítí. U všech sítí se ukázalo nejlepším řešením získat informace prostřednictvím jejich API. U sítí Twitter a LinkedIn je to dokonce jediná možnost jak zdarma získat data podle jejich licenčních podmínek.

Data miningových metod existuje velké množství. Jedním ze zmiňovaných komerčních nástrojů je Mozenda, který nabízí grafické rozhraní pro vybudování tzv. agentů. Tímto způsobem se dolují data, která má uživatel možnost vidět přímo ve svém prohlížeči. Ačkoli grafické prostředí je příjemné, licenční podmínky demo verze jsou omezené pouze na dvacet dní. V případě této práce, kdy bylo třeba se zaměřit na získání určitých specifických dat, tak toto časové období na plné seznámení s nástrojem nestačilo.

Podobným nezdarem dopadl také pokus získat data prostřednictvím open-source nástroje WEKA. V době zpracování této práce dokonce probíhal webový kurz pro seznámení se s tímto prostředím. Nedostatek zkušeností v oblasti dolování dat v kombinaci s náročností nástroje bylo však hlavní příčinou neúspěchu získání dat.

Nejjednodušším řešením tak bylo využití API sociálních sítí. Sběr dat byl poměrně komplikovaný. Na začátku byla motivace u sociální sítě Twitter sbírat geolokovaná data se zaměřením na zimní olympijské hry pořádané v roce 2014 v Rusku. Tweety měly být zaměřeny nejen na globální pojem olympijských her, ale podle průběhu her také na úspěšné sportovce, případně další témata, o kterých by se hojně mluvilo. Pro tento případ byla nastudována část API GET, vyzkoušena na vzorových datech a následně připravena na získání dat během samotného konání akce. Bohužel byla zjištěna skutečnost, že počet geolokovaných tweetů se pohybuje okolo 1 % ze všech příspěvků a vzhledem k tomu, že sociální síť Twitter je nejvíce oblíbená v USA, i drtivá většina těchto příspěvků odtud pocházela. Druhou velkou překážkou byla výrazná změna licenčních podmínek Twitteru v lednu 2014, díky které uživatel prostřednictvím GET API získává pouze zprávy populárních uživatelů a nikoliv výběr ze všech tweetů, jak tomu bylo dříve.

Proto bylo přistoupeno k náhradnímu řešení a to využití Streaming API. Vzhledem k budoucí myšlence tweetů geokódovat, byly příspěvky získávány bez ohledu na

geolokační atribut. Pro následnou vizualizaci tak bylo získáno ohromné množství příspěvků. Velké množství příspěvků však znamenalo velký objem dat. Získané soubory dosahovaly velikostí v řádech stovek MB. Díky velkému množství záznamů a velkému objemu dat proto geokódování probíhalo téměř celý měsíc. Tento způsob zisku prostorových dat se však ukázal jako dobrým krokem, protože bylo získáno velké množství záznamů s geolokací a vznikla tak možnost pro jejich další zpracování a zobrazení. Při procesu vizualizace bylo využito prostředí Leaflet. Díky dobré dokumentaci jednotlivých knihoven Leaflet a Crosslet i JavaScriptu bylo vytvořeno mapové prostředí, do kterého by bylo možné získaná data načíst. I zde se však ukázal jako problém velký objem dat. Ačkoli byly soubory GeoJSON očištěny o všechny nevizualizované atributy, načíst do mapy geokódované soubory za všechny zkoumané dny a zachovat tak interaktivnost mapy, bylo shledáno jako nemožné. Proto jsou ve výsledné mapě vizualizována geolokovaná data za období dvou týdnů a geokódovaná data za období jeden týden.

Ačkoli má sociální síť Facebook geolokační možnost, tato funkce není v tomto prostředí implementována jako samostatná služba, ale je pouze doplňujícím prvkem při přidávání příspěvků. Zisk těchto dat je tak velmi obtížný, a proto byla případová studie u této sociální sítě zaměřena na síť uživatelů (přátele), kteří hrají v tomto prostředí největší roli.

Sociální síť LinkedIn je světově jedinečná ve svém zaměření. Vzhledem k množství a charakteru informací, které zde uživatelé zveřejňují, je ohromnou databází specifických dat. Podstatou této sociální sítě však nejsou informace o poloze a díky relativně mladému API prostředí této sítě je možnost zisku většího objemu dat se zaměřením na prostorové informace téměř mizivá.

Data mining je výbornou možností k zisku informací vztažených k více podnětům zároveň. Vzhledem k velké variaci sociálních sítí je však nemožné zvolit vhodný obecný postup. Složitým krokem je i výběr dolovacích prostředků pro jednotlivé sociální sítě, vzhledem k velké variabilitě dat, která sítě obsahují. Pro správné zvolení je potřeba hluboké analýzy dané sociální sítě a všech dat v ní a dostatečné promyšlení konkrétních cílů před samotným výběrem dolovacích způsobů. Zájemce o dolování dat tak předejde řadě problémů, které zisk takovýchto informací přináší.

## 11 ZÁVĚR

Tato bakalářská práce se věnuje sociálním sítím a geoinformacím v nich obsaženým. Cílem práce bylo kromě zjištění typů prostorových dat, také nastudování možností data miningu (dolování dat) z prostředí sociálních sítí a následná vizualizace těchto dat. V praktické části práce byly zvoleny tři sociální sítě vhodné pro data mining prostorových dat a z nich získaná data byla vizualizována.

Bakalářská práce se tak věnovala třem světově rozšířeným sociálním sítím – Twitteru, Facebooku a LinkedIn a prostorovým datům, která je z nich možné získat. Postupně tak byla krátce popsána historie a filosofie těchto sítí, jejich rozhraní, jejich prostorové vlastnosti a možnosti získání těchto dat. Zvláštní důraz byl kladen i na popis API rozhraní jednotlivých sítí.

Z vybraných sociálních sítí byly vyextrahovány prostorové informace v podobách JSON souborů. V případě, že sociální síť neumožňovala přímé geotagování příspěvků, tak bylo nutné použít další textové atributy buď zprávy na sociální síti nebo z profilu uživatele. Tyto atributy byly následně geokódovány pomocí OSM Nominatim případně Google Geocoding API proto, aby bylo možné je následně zobrazit do webové mapy. Získaná data neobsahovala pouze prostorové prvky, ale také další doplňkové informace. Pro zobrazení těchto prvků jako interaktivní webové mapy byly využity především JavaScriptové knihovny Leaflet, které kromě zobrazení polohy umožňují znázornit i další informace.

U sociální sítě Twitter se praktická část práce zaměřila na oba obsažené prostorové prvky (automatická geolokace tweetů a prostorový atribut lokace uživatele). Původním záměrem bylo získat a hodnotit data s obsaženou geolokací u tweetů týkající se zimních olympijských her pořádané v roce 2014 v Rusku. Vysvětlení neúspěchu je popsáno v kapitole 7. Výsledkem dolování dat ze sociální sítě Twitter je tak případová studie zaměřená na situaci na Ukrajině v březnu roku 2014. Výsledkem jsou tedy vizualizovaná geolokovaná data uložená ve formátech GeoJSON. Dolování dat prostřednictvím části Streaming API u Twitteru se ukázalo jako vhodně zvoleným způsobem. Vydolovaná data totiž obsahovala na rozdíl od části Search API také informace o poloze autorů tweetů a po proběhlém geokódování byla tato data také vizualizována.

Sociální síť Facebook je založena na vytvoření sítě přátel a sdělování informací mezi nimi. Proto jednou z částí praktické práce bylo i zaměření se na prostorová data přátel jednotlivého uživatele. Vzhledem k API podmínkám není možné získat v této sociální síti souřadnice veřejných příspěvků s geolokací. Proto se autorka práce pokusila získat prostorová data o poloze veřejných příspěvků prostřednictvím atributu locale, který obsahuje informace o zemi, z které autoři pochází. Tyto znalosti byly aplikovány na klíčová slova obsahující příjmení vybraných prezidentů světa.

Ačkoli sociální síť LinkedIn je jedinečná svého druhu ve světě, způsob data miningu prostorových prvků prostřednictvím API a přístupových informací autorky práce se neukázal jako vhodný.

Bakalářská práce poskytuje několik hlavních výsledků. Nejdříve teoreticky popisuje sociální síť, možnosti získávání prostorových dat z prostředí sociálních sítí a následně na praktických ukázkách v prostředí Twitteru a Facebooku ukazuje, jak vybraná data extrahovat. Dalším neméně důležitým výsledkem jsou pak ukázky vizualizace ve formě interaktivních webových map, které data přehledně zobrazují a umožňují jejich další zkoumání.



# POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

## POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

### **Knížní zdroje**

HAN, J. a Mike KAMBER, 2000. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

KARIMI, Hassan A. *Big Data: Techniques and Technologies in Geoinformatics*. Boca Raton: CRC Press, 2014, 205 - 232. ISBN 978-1466586512.

MEMON, Nasrullah, Jennifer Jie XU, David L. HICKS a Hsinchun CHEN. *Data Mining for Social Network Data*. New York, USA: Springer New York Dordrecht Heidelberg London, 2010. ISBN 978-1-4419-6287-4.

MILLER, Harvey J. a Jiawei HAN. *Geographic Data Mining and Knowledge Discovery*. Salt Lake City: CRC Press, 2009. ISBN 978-1420073973.

MOON, Francis C. *Social Networks in the History of Innovation and Invention*. Ithaca, NY, USA: Springer Dordrecht Heidelberg New York London, 2014. ISBN 978-94-007-7527-5.

PEHLIVANIAN, Ara a Don NGUYEN. *JavaScript Okamžitě*. Brno: Albatros Media a.s., 2014. ISBN 978-80-251-4163-2.

RUSSELL, Matthew A. *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*. Sebastopol, CA: O'Reilly Media, 2011. ISBN 978-1-449-38834-8.

RUSSELL, Matthew A. *Mining the Social Web, Second Edition*. Sebastopol, CA: O'Reilly Media, 2014. ISBN 978-1-449-36761-9.

YAU, Nathan. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. USA: Wiley, 2011. ISBN 978-0470944882.

## Internetové zdroje

*About Python: Python.org* [online]. 2001-2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://www.python.org/about/>>.

API Mapy.cz: Ukázky práce s API. *API Mapy.cz* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://api.mapy.cz/view?page=geocoding>>.

ArcGIS - Chrome - Twitter. *ArcGIS* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.arcgis.com/home/item.html?id=870be7b510a646d6ae0137551aa9242d>>

ArcGIS - Social Media. *ArcGIS* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.arcgis.com/home/item.html?id=18230986538047b4b761715f0b5dd913>>.

BOYD, Danah M. a Nicole B. ELLISON. Social Network Sites: Definition, History and Scholarship. *Social Network Sites: Definition, History and Scholarship: Journal of Computer Mediated Communication* [online]. 2007 [cit. 2014-05-11]. Dostupné z WWW: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/pdf>> .

CARLSON, Nicholas. The Real History Of Twitter. *How Twitter Was Founded: Business Insider* [online]. 2011 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.businessinsider.com/how-twitter-was-founded-2011-4?page=2>>.

*Couchsurfing* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://www.couchsurfing.org/>>.

CRAN - Package twitteR. *The Comprehensive R Archive Network* [online]. [cit. 2014-05-11]. Dostupné z WWW: <<http://cran.r-project.org/web/packages/twitteR/index.html>>.

*Developer Network: seznam aplikací: LinkedIn* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://www.linkedin.com/secure/developer>>.

Developer Rules of the Road: Twtter Developers. *Twitter Developers* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://dev.twitter.com/terms/api-terms>>.

Esri/social-media-map-template-js: GitHub. *GitHub: Build software better, together* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://github.com/Esri/social-media-map-template-js>>.

*Facebook* [online]. 2014 [cit. 2014-05-11]. Dostupné z: <https://www.facebook.com/>

*Facebook Developers* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW:  
<<https://developers.facebook.com/>>.

*Facebook Reports Fourth Quarter and Full Year 2013: Facebook* [online]. 2014 [cit. 2014-05-11]. Dostupné z  
WWW:<<http://investor.fb.com/releasedetail.cfm?ReleaseID=821954>>.

FAYYAD, Usama, Gregory PIATETSKY-SHAPIRO a Padhraic SMYTH. *From Data Mining to Knowledge Discovery in Databases*. [online]. 1996 [cit. 2014-05-11].  
Dostupné z WWW:<<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>>.

*FOLDOC - Computing dictionary* [online]. 2014 [cit. 2014-05-11]. Dostupné z  
WWW: <<http://foldoc.org/>>.

*Foursquare* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW:  
<<https://foursquare.com/>> .

*GeoJSON* [online]. 16.8.2008 [cit. 2014-05-11]. Dostupné z  
WWW: <<http://geojson.org/>>.

*GeoNames* [online]. [cit. 2014-05-11]. Dostupné z WWW: <<http://www.geonames.org/>>.

*GeoSocialApp Aggregator* [online]. 2012 [cit. 2014-05-11]. Dostupné z  
WWW: <<http://gsaa.vlastovka.eu/>> .

Graph API Explorer. *Facebook Developers* [online]. 2014 [cit. 2014-05-11]. Dostupné z  
WWW: <<https://developers.facebook.com/tools/explorer/>>.

GREGOROVÍČ, Tomáš. *Extrakce informací ze sociálních medií*. Brno, 2010. Dostupné z  
WWW: <[https://is.muni.cz/th/139855/fi\\_m/dp.pdf](https://is.muni.cz/th/139855/fi_m/dp.pdf)>. Diplomová práce. Masarykova  
Univerzita.

HYETHORNTHWAITE, Caroline. Social networks and internet connectivity effects.  
[online]. 2005 [cit. 2014-05-11]. Dostupné z WWW:  
<[http://www.itu.dk/stud/speciale/specialeprojekt/Litteratur/Haythornthwaite\\_2005%20-%20Social%20networks%20and%20internet%20connectivity%20effects.pdf](http://www.itu.dk/stud/speciale/specialeprojekt/Litteratur/Haythornthwaite_2005%20-%20Social%20networks%20and%20internet%20connectivity%20effects.pdf)> ,

JavaScript API Developer Console. *LinkedIn Developer Network* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://developer.linkedinlabs.com/jsapi-console/#examples/login/simple.html>> ,

JARCOVJÁK, David. *Testování aplikace Foursquare pro sledování pohybu osob v Olomouci*. Olomouc, 2013. Dostupné z WWW: <[http://www.geoinformatics.upol.cz/dprace/bakalarske/jarcovjak13/download/jarcovjak\\_text.pdf](http://www.geoinformatics.upol.cz/dprace/bakalarske/jarcovjak13/download/jarcovjak_text.pdf)>. Bakalářská práce. Univerzita Palackého v Olomouci.

JSON [online]. 1999 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.json.org/>>

Klabosení: vyhledávání v archivu českého a slovenského Twitteru [online]. [cit. 2014-05-11]. Dostupné z WWW: <<http://www.klaboseni.cz/>> ,

Leaflet: a JavaScript library for mobile-friendly maps [online]. 2010-2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://leafletjs.com/>> ,

LinkedIn [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://www.linkedin.com/>>.

LinkedIn - About. *LinkedIn: Press Center* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://press.linkedin.com/about>>

MACHULDA, Ondřej. *Agregátor geolokačních sociálních sítí*. Praha, 2012. Dostupné z WWW: <[http://o.vlastovka.eu/download/machuon1\\_2012bach.pdf](http://o.vlastovka.eu/download/machuon1_2012bach.pdf)>. Bakalářská práce. České vysoké učení technické v Praze. Vedoucí práce Ing. Ondřej Macek.

No surprises: More people tweet more. Visualizing twitter counts during election day: R news a tutorials. *R-bloggers: R news a tutorials from the web* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.r-bloggers.com/no-surprises-more-people-tweet-more-visualizing-twitter-counts-during-election-day/>>.

O nás: *LinkedIn* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://www.linkedin.com/about-us>>.

OpenStreetMap Nominatim: Search [online]. [cit. 2014-05-11]. Dostupné z WWW: <<http://nominatim.openstreetmap.org/>>.

*SemioCast: Brazil becomes 2nd country on Twitter, Japan 3rd* [online]. 2012 [cit. 2014-05-11]. Dostupné z WWW: <[http://semioCast.com/en/publications/2012\\_01\\_31\\_Brazil\\_becomes\\_2nd\\_country\\_on\\_Twitter\\_supersedes\\_Japan](http://semioCast.com/en/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_supersedes_Japan)> .

Social Map. *Geographics.cz* [online]. 2010 [cit. 2014-05-11]. Dostupné z WWW: <<http://geographics.cz/socialMap/>>.

Sociální síť, seznam a přehled sociálních sítí. *Dvoran.com* [online]. [cit. 2014-08-17]. Dostupné z WWW: <<http://www.dvoran.com/socialni-site/marketing/prehled>>.

Statacumen.com: SC1\_16\_Maps. *Statacumen.com* [online]. 2008-2014 [cit. 2014-05-11]. Dostupné z WWW: <[http://statacumen.com/teach/SC1/SC1\\_16\\_Maps.pdf](http://statacumen.com/teach/SC1/SC1_16_Maps.pdf)>.

*The IPython Notebook: IPython* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<http://ipython.org/notebook.html>> .

*The R Project for Statistical Computing* [online]. 2009 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.r-project.org/>>.

*Throttle Limits: LinkedIn Developer Network* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://developer.linkedin.com/documents/throttle-limits>> .

TŮMOVÁ, Lucie. *Příležitosti a limity ve využívání sociálních sítí při náboru a výběru zaměstnanců*. Brno, 2013 [cit. 2014-05-11]. Diplomová práce. Masarykova Univerzita. Dostupné z WWW: < Dostupné z WWW: <<https://www.linkedin.com/about-us>>.

*Twitter* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://twitter.com/>>.

*Twitter Developers* [online]. 2014 [cit. 2014-05-11]. Dostupné z WWW: <<https://dev.twitter.com/>>.

Twitter Libraries: Twitter Developers. *Twitter developers* [online]. [cit. 2014-05-11]. Dostupné z WWW: <<https://dev.twitter.com/docs/twitter-libraries>>.

US Wildfire Activity. *ArcGIS* [online]. 2013 [cit. 2014-05-11]. Dostupné z WWW: <<http://www.arcgis.com/apps/OnePane/azuretwitter/index.html>>.

ZIV, Nina D. a Bala MULLOTH. An Exploration on Mobile Social Networking: Dodgebal as a Case in Point. [online]. 2006 [cit. 2014-05-11]. Dostupné z WWW: <[http://archive.poly.edu/management/\\_doc/nina/socialnetworking21.pdf](http://archive.poly.edu/management/_doc/nina/socialnetworking21.pdf)>.

## **SUMMARY**

Nowadays, social networks are amongst the most popular topics, which is caused mainly due to the rapid development and spreading of web and web based technologies. Thanks to social networks, millions of users are able to communicate with their friends and people with similar interests. They can easily share messages, multimedia, web links and also their location. The important fact in the evolution of social networks' geosocial aspects is the usage of mobile devices. These devices allow geolocation using GPS, WiFi or just a cell information gained directly from the mobile network.

This bachelor thesis describes the social networks and the types of geographical data that are contained within them. One of main goals was to find and to try suitable methods for mining of geographical information from selected social networks and subsequently visualize the data. The first step was to study the available literature in order to choose suitable social networks for data mining. Particular social networks - Twitter, Facebook and LinkedIn - were selected for the purpose of a detailed description and case studies.

Twitter was evaluated as a great choice for data mining and analysis of spatial data. The main reason is a possibility of direct geolocation (geotagging) of tweets. The practical case study was applied to the beginning of the Ukrainian crisis during March 2014. Data gained from Twitter stream using Twitter API were visualized in the form of the interactive web map based on the JavaScript library called Leaflet. Unfortunately, only 1% of all records were geotagged. The rest of records were geocoded using the attribute from users' profiles. After geocoding, there were much more results suitable for the geovisualization.

Data mining techniques were applied also to spatial attributes of Facebook and LinkedIn.. There is no geotagging option in LinkedIn network and although Facebook records eventually contained the spatial attribute, so the way of data storage is not suitable for data mining similar to Twitter. But map visualizations of Facebook user profile was made.

This bachelor thesis provides several useful results. Firstly it describes the process of (geo)data mining from several world spread social networks, then geovisualizations and interactive web maps were created.

# SEZNAM PŘÍLOH

## Vázané přílohy:

- Příloha 1 Tabulka tweetů získaných během procesu data miningu
- Příloha 2 Tabulka tweetů získaných během procesu data miningu
- Příloha 3 programový kód zisku informací ze sociální sítě LinkedIn prostřednictvím jazyka Python

## Volné přílohy

- Příloha 4 Poster
- Příloha 5 DVD

## Přílohy na DVD - ROM

- Příloha 6 R skript pro funkci geokódování finder
- Příloha 7 R skript pro dolování dat o přátelích ze sociální sítě Facebook
- Příloha 8 R skript pro dolování veřejných příspěvků ze sociální sítě Facebook
- Příloha 9 R skript pro dolování dat ze sociální sítě Twitter
- Příloha 10 data získaná ze sociální sítě Facebook
- Příloha 11 data získaná ze sociální sítě Twitter
- Příloha 12 záznam metadat ve formátu XML
- Příloha 13 tabulka informací o počtech získaných tweetů v zemích světa
- Příloha 14 Webová mapa znázorňující získané informace z Twitteru
- Příloha 15 Webová mapa znázorňující informace přátel z Facebooku
- Příloha 16 Webová mapa znázorňující veřejné příspěvky z Facebooku

## Popis struktury DVD

Adresáře:

- mapy
- metadata
- programove\_kody
- WEB
- data.rar
- Ttxt\_prace
- readme.txt



## Příloha 1

Tab. 1 Přehled získaných tweetů týkající se Ukrajinské krize

Datum sběru tweetů	Počet získaných tweetů	Počet tweetů s geolokací	Tweety s geolokací	Počet geokódovaných tweetů	Tweety s atributem location
16.3.2014	25822	269	0,01042	13660	0,52901
17.3.2014	9376	94	0,01003	4690	0,50021
18.3.2014	17868	150	0,00839	7642	0,42769
19.3.2014	13705	85	0,00620	6230	0,45458
20.3.2014	41241	313	0,00759	13449	0,32611
21.3.2014	12546	114	0,00909	5581	0,44484
22.3.2014	15847	125	0,00789	6971	0,43989
23.3.2014	11074	136	0,01228	5331	0,48140
24.3.2014	19297	238	0,01233	8535	0,44230
25.3.2014	25444	132	0,00519	9253	0,36366
26.3.2014	15230	125	0,00821	7446	0,48890
27.3.2014	17107	136	0,00795	7554	0,44157
28.3.2014	7666	59	0,00770	3159	0,41208
29.3.2014	12869	83	0,00645	5533	0,42995
30.3.2014	17401	103	0,00592	7151	0,41095
<b>CELKEM</b>	<b>262493</b>	<b>2162</b>	<b>0,00824</b>	<b>112185</b>	<b>0,42738</b>

## Příloha 2

Tab. 2 Přehled velikosti souborů obsahující tweety

<b>Datum sběru tweetů</b>	<b>Získaný soubor JSON [kB]</b>	<b>Soubor GeoJSON po geokódování [kB]</b>	<b>Soubor GeoJSON určený pro vizualizaci [kB]</b>
16.3.2014	115291	12302	5369
17.3.2014	94817	4199	1679
18.3.2014	78182	6944	2625
19.3.2014	59298	5704	2381
20.3.2014	192346	12168	4689
21.3.2014	54861	5115	2078
22.3.2014	73119	6255	2369
23.3.2014	49643	4832	2014
24.3.2014	80658	7770	2997
25.3.2014	115939	8228	3213
26.3.2014	66946	6844	2618
27.3.2014	97135	6912	2662
28.3.2014	34745	2906	1116
29.3.2014	64163	4965	1913
30.3.2014	78380	6567	2514
<b>CELKEM</b>	<b>1255523</b>	<b>101711</b>	<b>40237</b>

### Příloha 3

Programový kód zisku informací ze sociální sítě LinkedIn prostřednictvím jazyka Python:

```
#import knihoven
from linkedin import linkedin
from prettytable import PrettyTable
import json

# zisk klíčů z https://www.linkedin.com/secure/developer
CONSUMER_KEY = "XXXXX"
CONSUMER_SECRET = "XXXX"
USER_TOKEN = "XXXXXXXX"
USER_SECRET = "XXXXXXXX"

# vytvoření autorizačního protokolu OAuth
autorizace = linkedin.LinkedInDeveloperAuthentication(CONSUMER_KEY,
CONSUMER_SECRET, USER_TOKEN, USER_SECRET, RETURN_URL,
permissions=linkedin.PERMISSIONS.enums.values())

# přidání autorizace do aplikace
app = linkedin.LinkedInApplication(autorizace)

# příklad zisku informací o svém profilu
app.get_profile()

# zisk sítě kontaktů uživatele s atributy "Name " a "Location "
pt = PrettyTable(field_names=[ "Name", "Location"])
pt.align = "l "

[ pt.add_row((c["firstName"] + ' ' + c["lastName"], c["location"] [
"name"])))
    for c in connections["values "]
        if c.has_key("location")]
print pt

#vytvoření json souboru s daty
print json.dumps(pt)
```

