TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Aplikace pro analýzu uživatelů sociální sítě Twitter

Bakalářská práce

Studijní program: B2646 Informační technologie
Studijní obor: Informační technologie

Autor práce: **Michal Červinka**
Vedoucí práce: Mgr. Jiří Vraný, Ph.D.
Ústav nových technologií a aplikované informatiky

Liberec 2022

TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Zadání bakalářské práce

Aplikace pro analýzu uživatelů sociální sítě Twitter

Jméno a příjmení: **Michal Červinka**
Osobní číslo: M18000069
Studijní program: B2646 Informační technologie
Studijní obor: Informační technologie
Zadávající katedra: Ústav nových technologií a aplikované informatiky
Akademický rok: **2021/2022**

Zásady pro vypracování:

1. Seznamte se s problematikou tvorby webových aplikací v prostředí Node.js a se základními metodami analýzy sociálních sítí a dále s metodami pro vizualizaci a datovou reprezentaci vztahů v sociálních sítích.
2. Na základě získaných znalostí navrhněte aplikaci pro analýzu uživatelů sociální sítě Twitter a jejich vztahů. Aplikace by měla pracovat na veřejně přístupných datech a přinášet uživatelům informace nad rámec základní analýzy poskytnuté samotnou platformou Twitter.
3. Vytvořený návrh prakticky implementujte, a funkčnost aplikace demonstrujte vypracováním ukázkové analýzy.

Rozsah grafických prací:	dle potřeby dokumentace
Rozsah pracovní zprávy:	30-40 stran
Forma zpracování práce:	tištěná/elektronická
Jazyk práce:	Čeština

Seznam odborné literatury:

[1] CHATTERJEE, Siddhartha a Michal KRYSTYANCZUK. Python social media analytics: analyze and visualize data from Twitter, YouTube, GitHub, and more. Birmingham: Packt, 2017. ISBN 978-1787121485.

[2] MDN JavaScript [online]. Mozilla foundation [cit. 2021-10-1]. Dostupné z: https://developer.mozilla.org/en-US/docs/Web/JavaScript

[3] Twitter API [online]. Twitter, 2021 [cit. 2021-9-15]. Dostupné z: https://developer.twitter.com/en/docs/twitter-api

Vedoucí práce:	Mgr. Jiří Vraný, Ph.D. Ústav nových technologií a aplikované informatiky
Datum zadání práce:	12. října 2021
Předpokládaný termín odevzdání:	16. května 2022

L.S.

prof. Ing. Zdeněk Plíva, Ph.D.
děkan

Ing. Josef Novák, Ph.D.
vedoucí ústavu

V Liberci dne 19. října 2021

Prohlášení

Prohlašuji, že svou bakalářskou práci jsem vypracoval samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Jsem si vědom toho, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má bakalářská práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědom následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

14. května 2022 Michal Červinka

Aplikace pro analýzu uživatelů sociální sítě Twitter

Abstrakt

Tato bakalářská práce se zabývá tvorbou analýz uživatelů na sociální síti Twitter a jejich vztahů. Nejprve je v práci popsána analýza sociálních sítí a její použití. Poté je popsán způsob práce s veřejně dostupnými daty o uživatelích na Twitteru a použité technologie pro tvorbu praktické části. Hlavním cílem práce je vytvořit webovou aplikaci, která tvoří analýzy přinášející jejím uživatelům užitečné informace, jež jsou nad rámec základních analýz poskytované od platformy Twitter. Druhá kapitola se zabývá uživatelským rozhraní této webové aplikace. Třetí kapitola se zabývá podrobným popisem webové aplikace a její strukturou. Všechny analýzy, jak fungují a jaké informace poskytují je zde také zmíněno. Poslední kapitola této práce se zabývá demonstrací funkčnosti této aplikace prostřednictvím vytvoření ukázkových analýz a demonstrací výsledků daných analýz, ze kterých odvozuje zajímavá a užitečná fakta.

Klíčová slova: twitter, twitter API, analýza sociálních sítí

Application for Social Network Twitter User Analysis

Abstract

This bachelor thesis deals with a creation of user analyzes and their relationships on Twitter. First, the social network analysis and its use is mentioned, then the way of working with publicly available data in order to create analyzes of Twitter users is described. The main goal of this thesis is to make a web application, which produces analyzes useful to its users, who can benefit from obtaining useful information out of them. These analyzes are beyond the analyzes available on Twitter. The second chapter describes this web application. The third chapter deals with a detailed description of the development process of this application. All analyzes, how they work and what information do they provide are also described in this chapter. The last chapter does a functionality demonstration of this web application using created sample analyzes. Results of these analyzes are shown together with a discussion of interesting and useful facts derived from these analyzes.

Keywords: twitter, twitter API, social network analysis

Poděkování

Rád bych poděkoval panu doktorovi Jiřímu Vranému za vedení mé bakalářské práce a za poskytnutí konzultací v průběhu akademického roku.

Obsah

Seznam zkratek

API	Application Programming Interface
EJS	Embedded Javascript
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer protocol
JSON	JavaScript Object Notation
SPA	Single Page Application
TUL	Technická Univerzita v Liberci
URL	Uniform Resource Locator
UX	User Experience

1 Úvod

Twitter byl založen v roce 2006 a od té doby se stal jednou z největších a nejpoužívanějších sociálních sítí na dnešním internetu. V současné době má tato platforma milióny aktivních uživatelů, kteří zde tráví svůj čas. Twitter funguje na principu tzv. tweetů, což jsou krátké příspěvky, které může napsat kdokoliv a ostatní uživatelé tyto příspěvky mohou vidět. Uživatel, jehož tweety mohou být pro ostatní lidi zajímavé, začne postupně získávat publikum v podobě followerů. Samotný Twitter však nenabízí žádné analýzy uživatelů, které by šli nějakým způsobem do hloubky, ačkoliv by mnohé takové analýzy mohli být pro uživatele užitečné. Co však Twitter nabízí je přístup k veřejně dostupným datům ve formě API, díky kterým jsme schopni si některé analýzy vytvořit sami. Cílem této práce je vytvořit nástroj sloužící k automatické tvorbě těchto analýz přinášející uživatelům užitečné informace, které jsou nad rámec informací dostupných na samotném Twitteru.
Analýzy tvořené v této práci vycházejí z konceptu analýza sociálních sítí, což je proces zkoumání specifických vztahů v nejrůznějších skupinách lidí. Tato analýza se v posledních letech začala hojně používat, kvůli velkému rozšíření sociálních sítí, na kterých je vztah, oproti skutečnému světu, jasně definován (např. followers/following na Twitteru).
Praktickým výstupem této práce je webová SPA naprogramovaná v prostředí Node.js, která je schopná vytvářet několik druhů výše zmíněných analýz pro konkrétní uživatele, či určité skupiny uživatelů na Twitteru. Tato aplikace analýzy vytváří automaticky na pokyn uživatele a pracuje s veřejně dostupnými daty od Twitteru. Práce se dále zabývá uživatelským rozhraním aplikace, popisem struktury vytvořeného serveru a následnou demonstrací funkčnosti aplikace vytvořením ukázkových analýz.

2 Rešerše

2.1 Analýza sociálních sítí

Analýza sociálních sítí je v dnešní době chápána jako proces zkoumání formálních i neformálních svazků lidí a specifické vztahy mezi nimi. Tyto svazky a jejich interní vztahy nazýváme sociální struktury. V dnešní době z důvodu velkého rozšíření internetových sociálních sítí je tato analýza velmi užitečná. Výstupem analýzy sociálních sítí je ve většině případů neorientovaný graf, jehož vrcholy znázorňují konkrétní osoby/skupiny osob a hrany mezi nimi znázorňují konkrétní vztah. Osoby (vrcholy) s podobnými zájmy či vztahy se po provedení této analýzy v grafu ocitnou blízko sebe (nacházejí se ve stejném klustru). Naopak osoby, které spolu nemají žádné, či pouze malý počet společných zájmů leží v grafu daleko od sebe. Pro podrobnější analýzu můžeme provést další operace jako např.: Rozdělení osob do skupin pomocí algoritmů založených na výpočtu modularity, výpočet váhy každého vrcholu pomocí centrality vlastního vektoru atd. Analýza sociálních sítí může být velmi užitečná např. při výzkumu fungování rodin nebo při vizualizaci vztahů v jednotlivých komunitách. V této práci se analýza sociálních sítí používá pro znázornění českých twitterových účtů s velkým počtem followerů a jejich vztahy v podobě počtu společných followerů (více informací v kapitole 4.4.4). Většina ostatních analýz v této práci také zapadá do tohoto tématu, jen jejich výstupem není neorientovaný graf, ale jiné způsoby zprostředkování informací (např. tabulky nebo mapy). [1]

2.2 Twitter

Twitter patří mezi největší a nejznámější sociální sítě dnešní doby. Jeho charakteristická vlastnost jsou tzv. tweety, což jsou krátké příspěvky primárně v textovém formátu (mohou obsahovat i obrázky a videa), které může každý registrovaný uživatel zveřejnit na svém profilu. Ostatní uživatelé mohou tyto tweety vidět a několika způsoby na ně reagovat. Tyto způsoby jsou: Like (vyjádření souhlasu či ocenění obsahu daného tweetu), Retweet (sdílení tweetu jiného uživatele na svém profilu) a přidání komentáře. Uživatel může dále aktivně sledovat jiné uživatele dle jeho výběru. Tweety těchto uživatelů jsou následně danému uživateli přednostně zobrazeny.

2.3 Twitter API

Twitter API je rozhraní od Twitteru, pomocí kterého mohou vývojáři získávat veřejně dostupné informace o všech aspektech sociální sítě. Mezi požadavky na toto API, které může vývojář uskutečnit je např. získání informací o konkrétním uživateli, koho jaký uživatel sleduje, kdo všechno sleduje daného uživatele, jaké tweety konkrétní uživatel zveřejnil, komentáře pod daným tweetem atd. Pro přístup k Twitter API je potřeba mít k dispozici unikátní klíč (tzv. token) vygenerovaný od samotné platformy Twitter. Bez klíče se nelze dostat ke většině informacím dostupných na Twitter API. Z tohoto API nemůžeme získat žádné informace, které nejsou dostupné normálním způsobem pomocí webové či mobilní aplikace, jde tedy pouze o usnadnění přístupu k informacím programovatelným způsobem. S Twitter API se komunikuje přes tzv. HTTP požadavky a získávané odpovědi jsou výhradně ve formátu JSON. Kromě získávání informací lze použít Twitter API i pro zveřejňování a úpravu tweetů a profilů, k tomu jsou však potřeba vyšší úrovně autorizace a autentizace. [2] Těmito možnostmi se však tato práce nezabývá.

2.3.1 Objekty v Twitter API

Twitter API nabízí většinu informací ve formě speciálních objektů. Těmito objekty jsou:

User objekty

Obsahují informace o specifickém uživateli na twitteru. V tomto objektu můžeme nalézt důležité informace o daném uživateli jako je např. jméno uživatele, popis účtu, lokace, počet followerů, počet přátel, kdy byl účet vytvořen, zda-li jde o ověřený účet atd. Všechny tyto atributy jsou veřejnosti nedostupné v případě, že uživatel, jehož daný user objekt představuje, má soukromý účet. V takovém případě si daný uživatel zvolil, že nechce, aby jeho informace byli dostupné veřejnosti a tím pádem se je nemůžeme dozvědět ani z Twitter API. Tento typ objektu se v této práci používá téměř ve všech případech.

Tweet objekty

Dalším objektem je tweet objekt, který představuje jeden konkrétní tweet. Zde se můžeme dozvědět obsah tweetu, kdo je autorem, zda-li jde o retweet, počet lajků, počet retweetů atd.

Geo a Entities objekty

Geo objekt obsahuje informaci o lokaci, kde byl konkrétní tweet zveřejněn na Twitteru (pokud ovšem autor tweetu povolil tuto informaci poskytnout). Entities objekt obsahuje metadata tweetů, jako jsou např. hashtagy nebo mentions (uživatelé, kteří byli označeni v daném tweetu). [3]

2.3.2 Rate Limits

Pro bezproblémové používání Twitter API je potřeba dodržovat tzv. rate limits. Jedná se o omezení počtu možných požadavků, které může jednotlivý uživatel API provést za určitý časový úsek. Toto omezení si nastavuje sám poskytovatel daného API (v tomto případě Twitter). Např. žádost o získání 100 user objektů (objekt nesoucí informace o specifickém uživateli) můžeme provést maximálně 300x za 15 minut, tj. maximálně 1 za 3 sekundy. Pokud bychom měli účet, který sleduje 100 000 uživatelů a chtěli bychom získat informace o každém z nich (ve formě user objektů), trvalo by nám to 50 minut. Důvodů, proč tomu tak je, je několik. V prvé řadě je toto API velmi používané. Nejen vývojáři, ale i automatické procesy mohou totiž využívat služby Twitter API např. pro provoz tzv. botů (účty spravované různými programy, namísto reálných osob). Z tohoto důvodu mohou být servery poskytující Twitter API velmi zaneprázdněny a bez implementace rate limitů by objem příchozích požadavků pravděpodobně nezvládly. Dalším důvodem je zvýšení peněžních příjmů pro společnost Twitter. Z tohoto důvodu Twitter nabízí placený přístup k lepším rate limitům. Kromě placeného přístupu existuje i možnost akademického přístupu, který má rovněž lepší podmínky pro použití. Oba tyto způsoby však zlepšují přístup pouze k informacím o tweetech a aplikace vytvořená v této práci se zabývá výhradně uživateli. Pro splnění cílů této práce proto nejsou exkluzivní rate limity potřeba. [4]

2.4 Relativní četnost

Rate limity mohou představovat velkou nevýhodu při získávání informací z Twitter API, znamenají totiž dlouhé časové úseky, než získáme všechna data potřebná k analýzám (zejména u analýz velkých twitterových účtů). Relativní četnost představuje způsob, jak můžeme do jisté míry obejít toto dlouhé čekání na výsledky rozborů. *Ing. Martina Litschmannová, Ph.D.* v *Úvodu do statistiky* [5] definuje relativní četnost následovně:

> Uvažujme nějaký náhodný jev A vyskytující se s pravděpodobností π a předpokládáme, že provádíme opakovaná nezávislá pozorování tohoto jevu. Označme $X_i = 1$, pokud jev A při i-tém pozorování nastal a $X_i = 0$, pokud nenastal. Pak $X_1, X_2, \ldots$ je náhodný výběr z alternativního rozdělení. Výběrový průměr $\overline{X}$ vypočítaný z prvních n pozorování označujeme v tomto případě jako relativní četnost

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.1}$$

2.4.1 Zákon velkých čísel

Relativní četnost vychází ze zákona velkých čísel. *Ing. Martina Litschmannová, Ph.D.* v *Úvodu do statistiky* [5] popisuje zákon velkých čísel následovně:

Vypočteme-li výběrový průměr (v tomto případě relativní četnost) z náhodného výběru o rozsahu rovném rozsahu populace, získáme střední hodnotu rozdělení, z něhož výběr pochází. Vypočteme-li výběrový průměr z náhodného výběru o rozsahu menším než je rozsah populace, nezískáme přesně střední hodnotu rozdělení, ale dostaneme číslo, které je skutečné střední hodnotě blízko.

$$\lim_{n\to\infty}\left(P\left|\overline{X}-\pi\right|\right)=0 \tag{2.2}$$

2.4.2 Příklad použití relativní četnosti

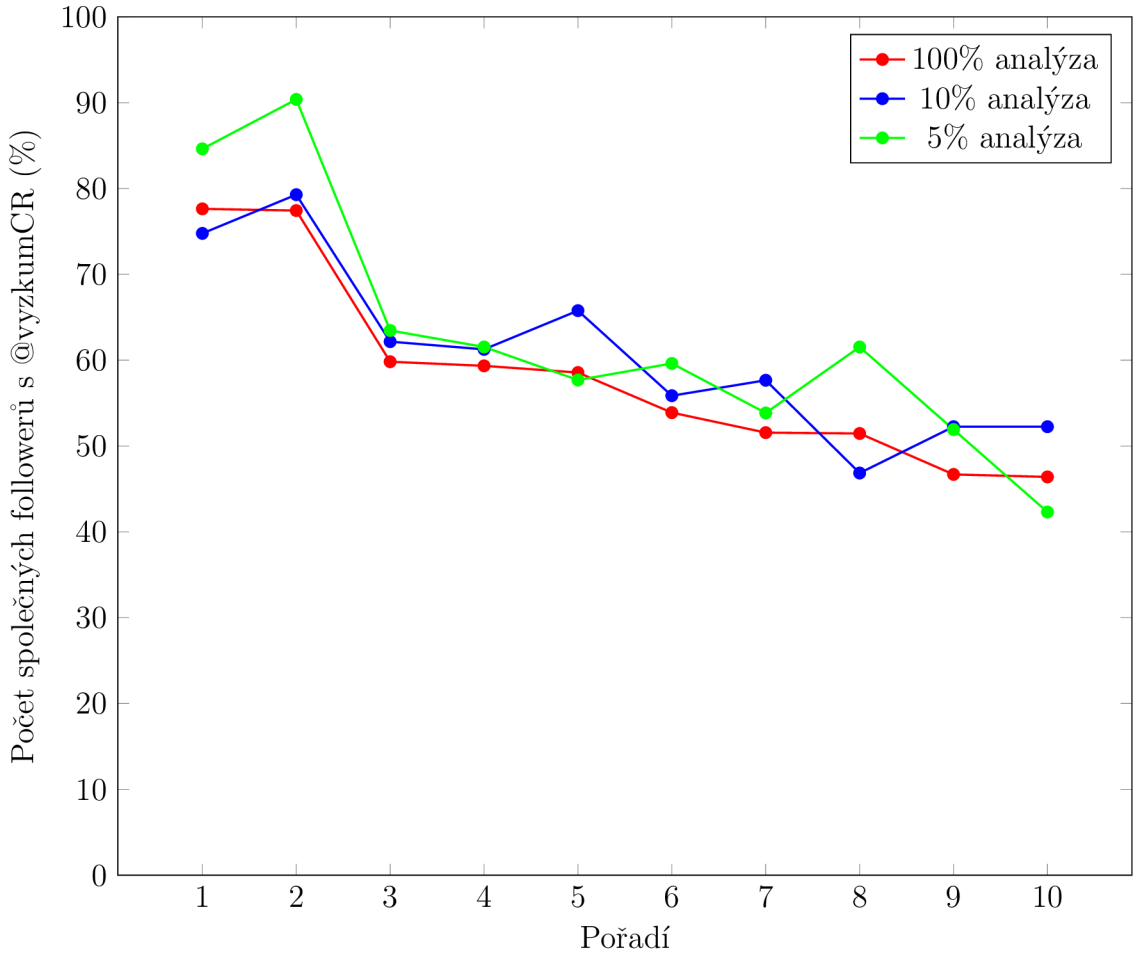
Při tvorbě analýz twitterových uživatelů, čímž se tato práce zabývá, se dá v některých případech použít relativní četnost pro zkrácení doby čekání na výsledky. Jedna z těchto analýz je analýza zájmů followerů jednotlivých uživatelů. Zde se můžeme dozvědět, kolik procent z followerů určitého uživatele na twitteru sleduje jiné uživatele (koho uživatelovi followeři sledují nejvíce, více informací o této analýze v kapitole 4.4.3). Pro demonstraci tohoto jevu byl vytvořen experiment, ve kterém byla několikrát provedena tato analýza pro twitterový účet @vuzkumCR, který měl v době provedení tohoto experimentu (Březen 2022) 1028 followerů. Nejprve byli pro analýzu staženi všichni followeři účtu, což trvalo 17 hodin a 8 minut. Poté bylo staženo pouze 10% ze všech followerů účtu, což zabralo pouze 1 hodinu a 43 minut. Tyto dvě kolekce followerů byly poté použity pro výše uvedenou analýzu a výsledky byly porovnány. Pokud si budou výsledky vzájemně podobné, znamená to, že relativní četnost je zde vhodná pro zkrácení dob čekání na výsledky.
V tabulce 2.1 můžeme vidět výsledky obou analýz (10 twitterových účtů, se kterými má účet @vyzkumCR nejvíce společných followerů). Z dat lze pozorovat, že výsledky obou analýz jsou si podobné. Průměrný rozdíl je 4,027%. Stejný proces byl poté proveden i pro 5% analýzu stejného twitterového účtu a výsledky jsou k nahlédnutí v tabulce 2.2. V tomto případě analýza trvala pouze 51 minut a výsledné hodnoty se od skutečných lišili průměrně o 5,407%. V grafu 2.4.2 lze pozorovat odchylky výsledků pro jednotlivé účty v každé z výše provedených analýz. Z grafu je patrné, že všechny data mají sestupnou tendenci ale patrné rozdíly jsou. Pokud však chceme, aby UX běžného uživatele této aplikace byl přijatelný, je tato úspora času, obzvláště u velkých twitterových účtů potřeba.

10% analýza				
Pořadí	Název účtu	100% analýza (%)	10% analýza (%)	Rozdíl (%)
1	vedavyzkum_cz	77,63	74,77	2,86
2	Akademie_ved_CR	77,43	79,28	1,85
3	okonovatech	59,82	62,16	2,34
4	Vedavyzkum	59,34	61,26	1,92
5	CT24zive	58,56	65,77	7,21
6	TACR_cz	53,89	55,86	1,97
7	DanielStach	51,56	57,66	6,10
8	Vedact24	51,46	46,85	4,61
9	DVTVcz	46,69	52,25	5,56
10	CzechTV	46,40	52,25	5,85

Tabulka 2.1: 10 twitterových účtů, se kterými má účet @vyzkumCR největší procento společných followerů. Analýza všech followerů a analýza 10% followerů

5% analýza				
Pořadí	Název účtu	100% analýza (%)	5% analýza (%)	Rozdíl (%)
1	vedavyzkum_cz	77,63	84,62	6,99
2	Akademie_ved_CR	77,43	90,38	12,95
3	okonovatech	59,82	63,46	3,64
4	Vedavyzkum	59,34	61,54	2,20
5	CT24zive	58,56	53,85	2,29
6	TACR_cz	53,89	57,69	0,87
7	DanielStach	51,56	59,62	5,73
8	Vedact24	51,46	61,54	10,08
9	DVTVcz	46,69	51,92	5,23
10	CzechTV	46,40	42,31	4,09

Tabulka 2.2: 10 twitterových účtů, se kterými má účet @vyzkumCR největší procento společných followerů. Analýza všech followerů a analýza 5% followerů



Obrázek 2.1: Graf znázorňující podobnost výsledků provedených analýz na účtu @vyzkumCR (možné nahlédnout do tabulky 2.1 a 2.2)

2.5 Použité technologie

2.5.1 Node.js

Node.js je open-source prostředí, které je primárně navržené pro tvorbu serverových částí webových aplikací. Je postaveno na programovacím jazyce JavaScript, který se před existencí Node.js dal používat pouze na straně klienta ve webovém prohlížeči. Za pomocí prostředí Node.js je však JavaScript možné použít i v serverové části aplikace. Node.js je díky své architektuře velmi výkonný a dobře škálovatelný a proto se často používá právě pro API servery a SPA, ačkoliv je toto prostředí možné použít i pro spoustu jiných projektů. Další výhodou Node.js je jeho multiplatformnost, tz. je možné spustit server naprogramovaný v tomto prostředí na zařízeních s různými operačními systémy a různými hardwarovými architekturami. Serverová část webové aplikace vytvořené v této práci je naprogramována v tomto prostředí. [6]

2.5.2 Python

Python je interpretovaný, vysokoúrovňový programovací jazyk, který je vhodný pro implementaci algoritmů zejména díky jeho snadné syntaxi. Pro Python existuje velké množství knihoven, tudíž je vhodný kandidát pro spoustu projektů. Nejnovější verze tohoto jazyku je Python 3, což je verze, která je použita pro tvorbu praktické části této práce. [7]

2.5.3 Gephi

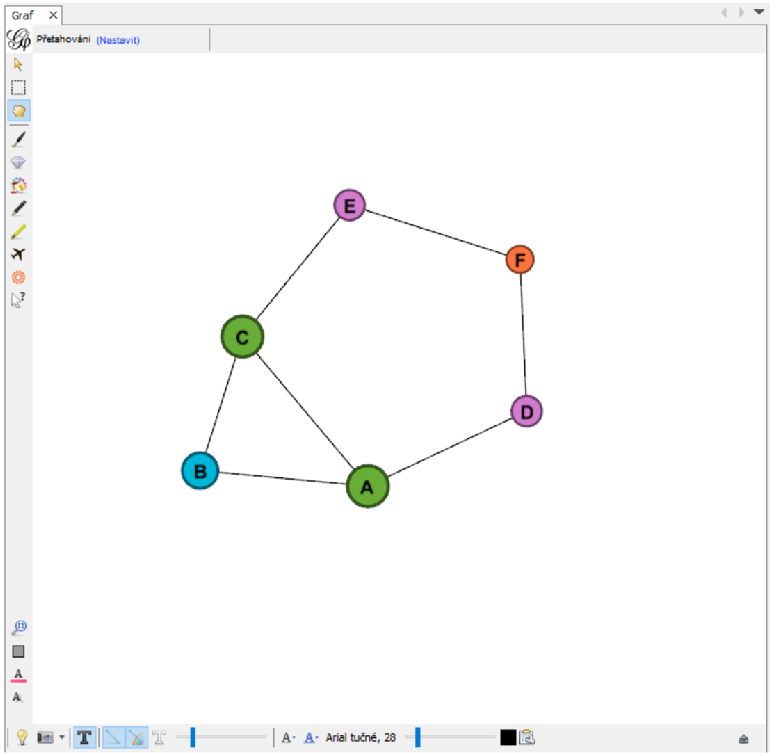
Gephi je open-source program pro vytváření, vizualizaci a úpravu grafů. Tento program je zadarmo a užitečný v projektech, kde se využívá analýza sociálních sítí. Software Gephi nabízí spoustu možností modifikace grafů jako např. výběr z několika algoritmů pro automatické rozložení uzlů v grafu, možnost změny velikosti uzlu a jeho barvy pomocí algoritmu našeho výběru, výpočty modularity, hustoty grafu, centrality vlastního vektoru atd. Pro potřeby této práce je tato aplikace naprosto dostačující [8]. V obrázku 2.2 je vidět jednoduchý graf vytvořený v tomto programu.

2.5.4 MongoDB

MongoDB je open-source databáze, která se řadí mezi tzv. NoSQL databáze (nerelační databáze), kde se data místo do tabulek ukládají do dokumentů připomínající formát JSON (ve skutečnosti jde o tzv. BSON, Binary JSON). Pro tuto práci byl zvolen právě tento typ databáze. [9]

2.5.5 Leaflet

Leaflet je open-source javascriptová knihovna umožňující vývojářům možnost přidání interaktivní mapy do jejich aplikací. Knihovna je relativně jednoduchá na implementaci a umožňuje spoustu úprav, které lze využít k vytvoření mapy na míru a pro



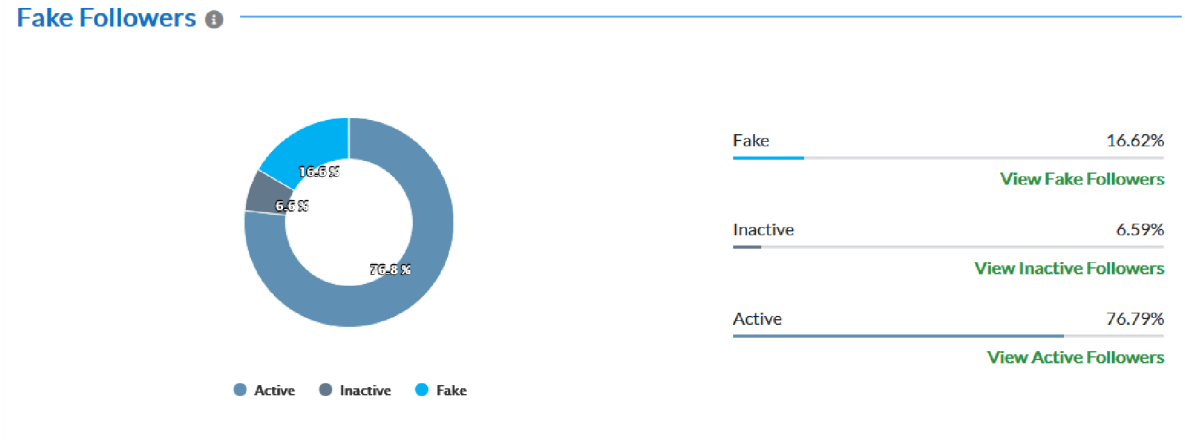
Obrázek 2.2: Jednoduchý neorientovaný graf vytvořený v programu Gephi

potřeby konkrétní aplikace. Pro tuto práci je tato knihovna více než dostatečná a proto je zde také využita. [10]

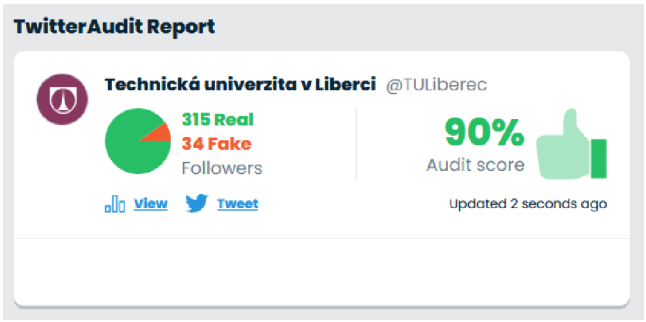
2.6 Podobné aplikace

2.6.1 Tweeps Map

Tweepsmap.com je webová aplikace sloužící k analýze followerů konkrétního uživatele, který se do aplikace přihlásí pod svým twitterovým účtem. Tweeps Map nabízí hned několik služeb, které se dělí na zadarmo a na placené. Mezi služby, které jsou zadarmo patří vyhledávání lokací, ve kterých se nacházejí uživatelovi followeři (státy, města) a vyhledávání lokací, ze kterých je daný uživatel často zmiňován ve tweetech ostatních uživatelů (data jsou následně uživateli poskytnuta ve formě interaktivní mapy). Po zaplacení měsíční částky se uživatelovi odemknou i ostatní služby, jako jsou: sledování historie účtu (růst/pokles počtu followerů atd.), histogramy ukazující demografii followerů (počet followerů v určité věkové skupině), sledování lidí, kteří uživatele přestali sledovat atd. Webová aplikace v této práci nabízí podobné služby a např. analýza velkých českých twitterových uživatelů (viz. kapitola 4.4.4) na Tweeps Map není. [11]



Obrázek 2.3: Odhad poměru falešných a neaktivních followerů twitterového účtu @TULiberec podle služby Follower Audit [12]



Obrázek 2.4: Odhad počtu falešných followerů twitterového účtu @TULiberec podle služby Twitter Audit [13]

2.6.2 Follower Audit

Followeraudit.com je webová aplikace, která se specializuje na detekci tzv. falešných účtů. Pro získání jakýchkoliv dat o jakémkoli uživateli se musíme do aplikace nejprve přihlásit pod svým twitterovým účtem (aplikace potřebuje uživatelův účet pro vyhledávání dat z Twitter API). Výsledky analýzy followerů jsou poté uživatelovi zobrazeny ve formě grafů. Kromě odhadu počtu falešných followerů zde můžeme dostat i jiné informace jako např. počet followerů se soukromým účtem, či počet ověřených followerů. Pokud používáme bezplatnou verzi této služby, aplikace nám udělá analýzu pouze 5000 nejnovějších followerů. V obrázku 2.3 můžeme vidět výsledek analýzy falešných a neaktivních followerů twitterového účtu TUL (@TULiberec). Mělo by jít o odhad všech followerů, jelikož účet @TULiberec má méně followerů než 5000. Podle odhadu služby Follower Audit má účet @TULiberec 16.62% falešných followerů (hodnota později použita pro porovnání s výsledkem aplikace vyvinuté pro účely této práce). [12]

2.6.3 Twitter Audit

Twitteraudit.com je webová služba, která obdobně jako Follower Audit nabízí analýzu falešných followerů zadaného uživatele na twitteru. Oproti Follower Audit nabízí tato služba méně informací o uživateli, zato však nabízí všechny analýzy vykonané v minulosti veřejnosti. Např. pokud se v aplikaci uživatel autorizuje svým twitter účtem a provede analýzu určitého uživatele, tato analýza je následně dostupná pro všechny budoucí návštěvníky (i pro ty neautorizované). Bezplatná verze nám opět nabízí analýzu pouze 5000 nejnovějších followerů. Výsledek analýzy od této služby pro účet @TULiberec je k nahlédnutí v obrázku 2.4. Výsledek této analýzy nám říká, že @TULiberec má pouze 10% falešných followerů (méně než odhad od Followers Audit). [13]

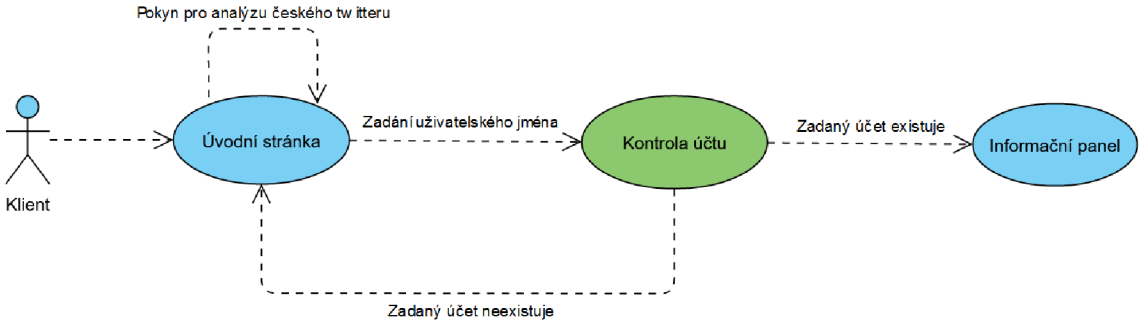
3 Uživatelské rozhraní aplikace

3.1 Úvodní stránka

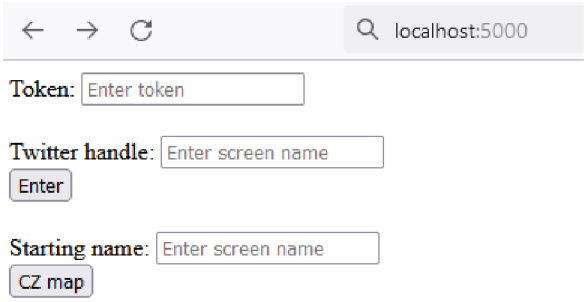
První stránka, kterou uživatel ve vytvořené webové aplikaci uvidí je úvodní stránka (zobrazena v obrázku 3.2). URL adresa je shodná s adresou serveru. Na této stránce se nacházejí jen 2 jednoduché formuláře. První formulář uživatel použije, pokud si chce nechat vytvořit analýzy konkrétního twitterového účtu. V takovém případě napíše do textového pole s popiskem Twitter handle jméno twitterového účtu, kterého chce zkoumat, stiskne tlačítko Enter a je přesměrován na informační panel o konkrétním uživateli. Druhý formulář uživatel použije, pokud si chce nechat vytvořit novou analýzu velkých českých twitterových uživatelů a jejich společných followerů. V takovém případě vyplní textové pole názvem startovního účtu (více informací v kapitole 4.4.4) a po stisknutí tlačítka CZ map se patřičná analýza spustí. Na úvodní stránce se nachází také textové pole pro zadání tokenu (více informací v kapitole 4.3). Toto pole musí uživatel vyplnit v případě, že chce vytvářet nové analýzy. Pokud si přeje pouze zobrazit výsledky starších analýz, může nechat pole prázdné. Pokud je zadán chybný token, uživatel je přesměrován zpět na úvodní stránku.

3.2 Informační panel o konkrétním uživateli

Po zadání uživatelského jména konkrétního twitterového účtu, kterého chceme zanalyzovat, dojde ke kontrole, zda daný účet na twitteru opravdu existuje. Pokud je kontrola úspěšná a na Twitteru se opravdu vyskytuje účet s daným jménem, který není nastavený jako soukromý, uživatel je přesměrován na informační panel, kde je možné zobrazit si výsledky analýz zadaného twitterového účtu. V opačném případě je uživatel přesměrován zpět na úvodní stránku. V tomto informačním panelu má uživatel u každého druhu analýzy 2 možnosti: Zobrazit si výsledky poslední vykonané analýzy u daného twitterového účtu (pokud již byla někdy v minulosti u tohoto účtu daná analýza vykonána) anebo učinit pokyn pro provedení nové analýzy. URL adresa tohoto panelu je /dashboard zapsaný za adresou serveru. Uživatel se však na tuto stránku nemůže dostat obyčejným zadáním URL adresy, šlo by totiž o neautorizovaný přístup. Aplikace by v tomto případě nevěděla, s jakým twitterovým účtem má pracovat a tudíž je uživatel přesměrován zpět na úvodní stránku. Přístup k informačnímu panelu dostane uživatel až tehdy, kdy na úvodní stránce zadá jméno zkoumaného twitterového účtu (poté je ovšem již možné se do panelu dostat i zadáním příslušné URL adresy). Po načtení informačního panelu jsou na



Obrázek 3.1: Use case diagram aplikace

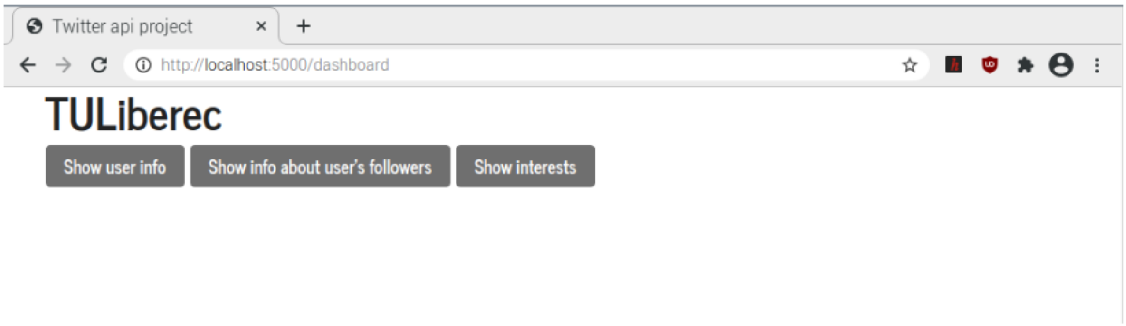


Obrázek 3.2: Úvodní stránka webové aplikace

pár sekund tlačítka pro zobrazení analýz nefunkční (možno vidět v obrázku 3.3). Důvodem je, že aplikace počítá odhad dob trvání jednotlivých analýz pro daný twitterový účet v případě, že se uživatel rozhodne dát pokyn pro jejich vykonání (tyto doby jsou poté u každé analýzy zobrazeny). Po cca 3 sekundách se již přístup k tlačítkům povolí a uživatel může začít s aplikací pracovat. V případě, že uživatel nezadal svůj token na úvodní stránce, informační panel zobrazí pouze výsledky starých analýz, možnost vytvořit nové je nedostupná.

3.2.1 Základní informace o uživateli

První analýza, kterou lze v informačním panelu zobrazit je analýza samotného twitterového uživatele (tlačítko Show user info). Tato analýza nabízí základní informace o zvoleném účtu. Mezi tyto informace patří: Počet followerů, počet přátel, počet tweetů, jméno, lokace účtu uvedená od uživatele, zda je daný účet "verified"(ověření od samotné platformy Twitter, zda je daný účet autentický), popis účtu, kdy byl účet vytvořen, průměrný počet lajků na tweetech daného účtu, průměrný počet retweetů na daném účtu, jazyk, ve kterém daný účet tweetuje, průměrná tweetovací frekvence (jak často daný účet v průměru tweetuje) a odhad, zda je účet pravý, či falešný/neaktivní. Tyto data nemají moc velkou hodnotu vzhledem k tomu, že většinu těchto informací lze zjistit jednoduše pouhým navštívením twitterového profilu daného účtu. V obrázku 3.4 lze vidět zobrazená analýza základních informací



Obrázek 3.3: Informační panel pro twitterový účet @TULiberec, ihned po načtení (tlačítka pro analýzy ještě nefunkční)

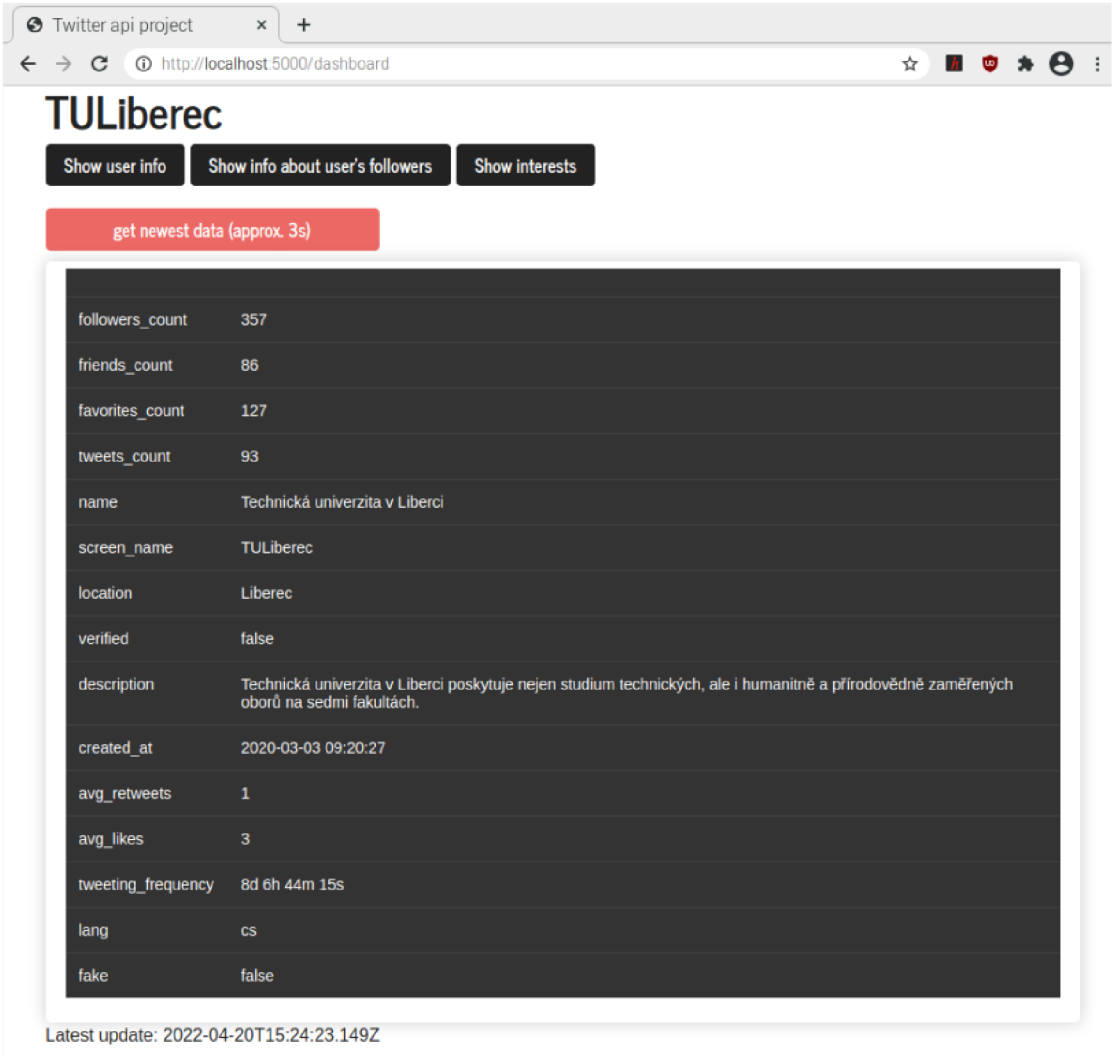
twitterového účtu @TULiberec vykonána 20. dubna 2022 (více informací v kapitole 5.2.1).

3.2.2 Informace o followerech

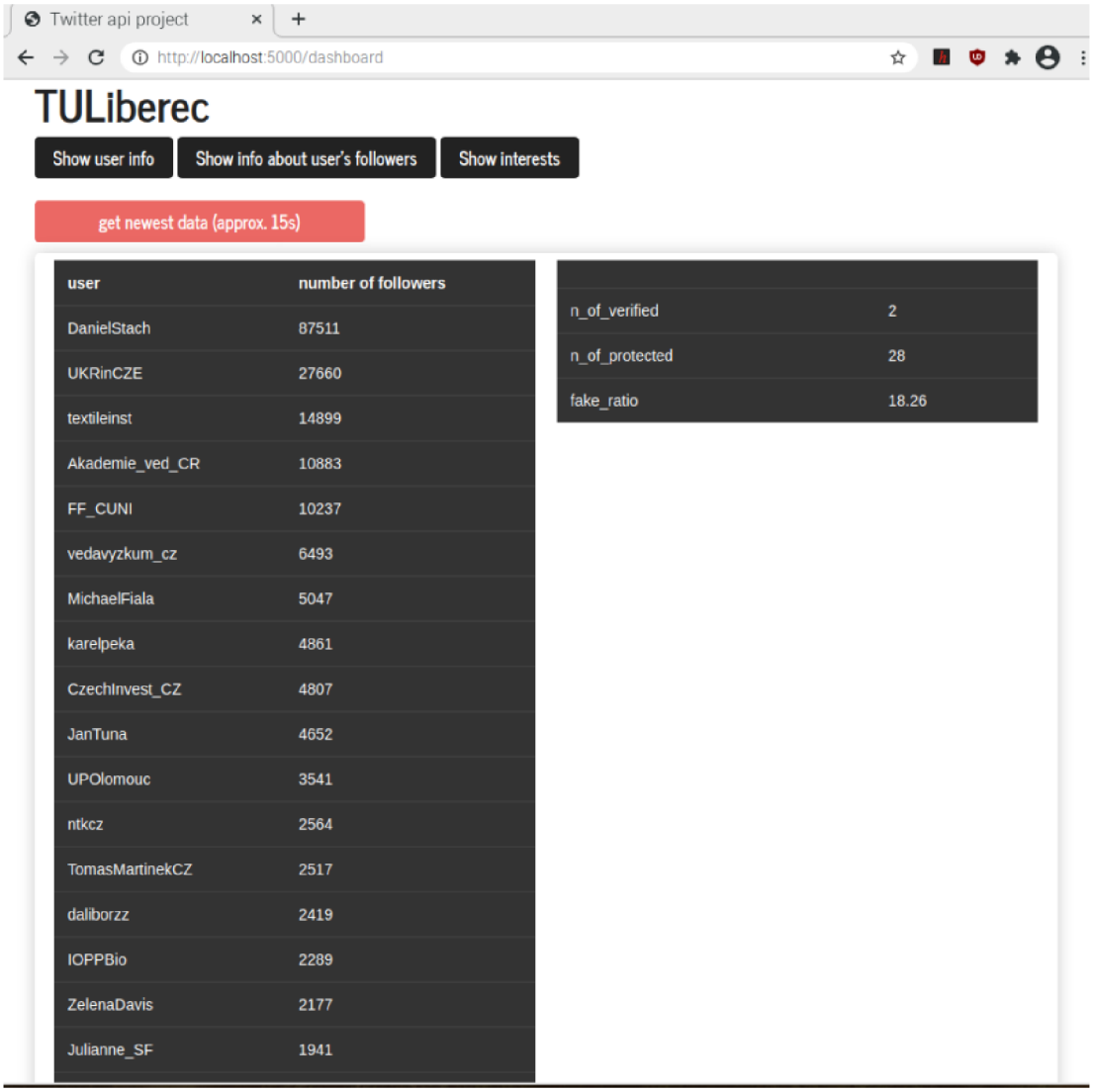
Druhá analýza dostupná na informačním panelu je analýza followerů zkoumaného twitterového účtu. Tato analýza již poskytuje informace, které je od samotného twitteru normální uživatelskou cestou nemožné sehnat. Těmito informacemi jsou: List 30 největších twitterových uživatelů, kteří sledují zkoumaný účet, seznam, kolik followerů je z jaké země (v podobě interaktivní mapy), hrubý odhad podílu followerů, kteří jsou buďto falešní nebo neaktivní, počet verified účtů a počet soukromých účtů. V obrázku 3.5 je vidět, jak aplikace tuto analýzu zobrazuje uživateli. V obrázku jsou vidět také výsledky analýzy twitterového účtu @TULiberec (více informací v kapitole 5.2.2). V obrázku 3.6 lze nahlédnout do interaktivní mapy geografie followerů účtu @TULiberec (kolik followerů je z jaké země, více informací v kapitole 5.2.2) (vytvořená za použití open-source knihovny Leaflet [10]). Analýza viditelná v těchto obrázcích byla vykonána 26. dubna 2022.

3.2.3 Zájmy followerů

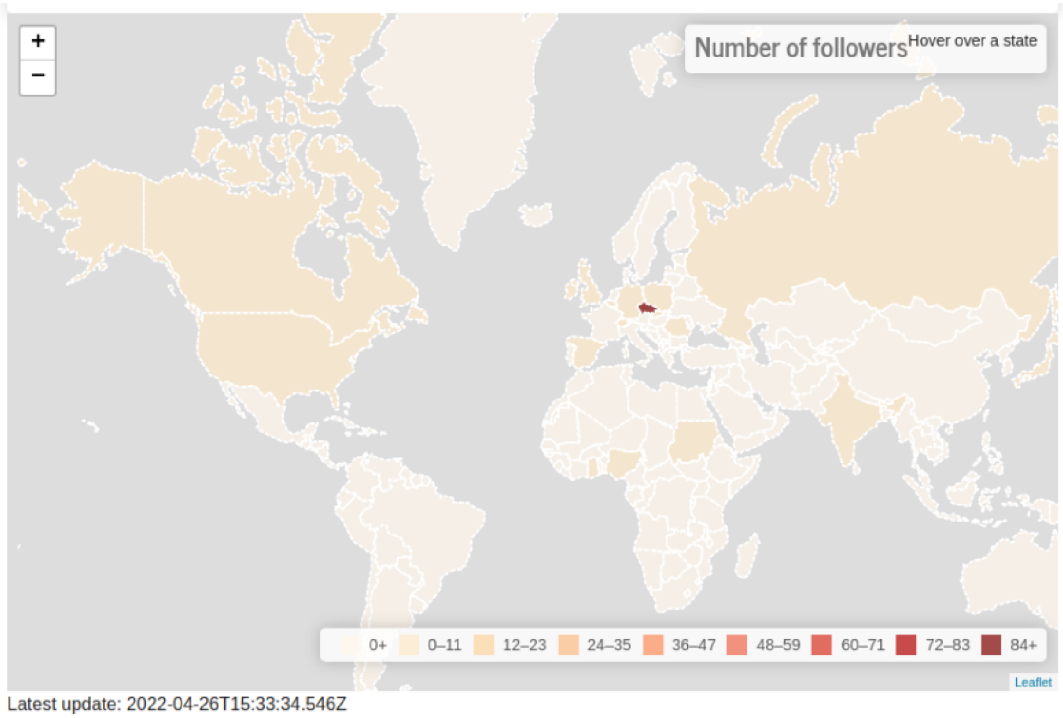
Třetí analýza dostupná na informačním panelu je analýza zájmů followerů zkoumaného účtu. Výsledek této analýzy nám umožňuje zjistit seznam až 50-ti twitterových účtů, se kterými má zkoumaný účet nejvíce společných followerů. Např. pokud je v seznamu účet X se 40% společných followerů, znamená to, že 40% lidí, kteří sledují zkoumaný twitterový účet, sledují také účet X. Tyto informace jsou velmi vhodné pro uživatele, který se zajímá o různé zájmy, které followeři zkoumaného twitterového účtu mají. V obrázku 3.7 lze vidět, jak aplikace zobrazuje seznam twitterových účtů, se kterými má zkoumaný účet nejvíce společných followerů. V tomto případě jde o analýzu účtu @TULiberec ze dne 16. dubna 2022 (více informací v kapitole 5.2.3).



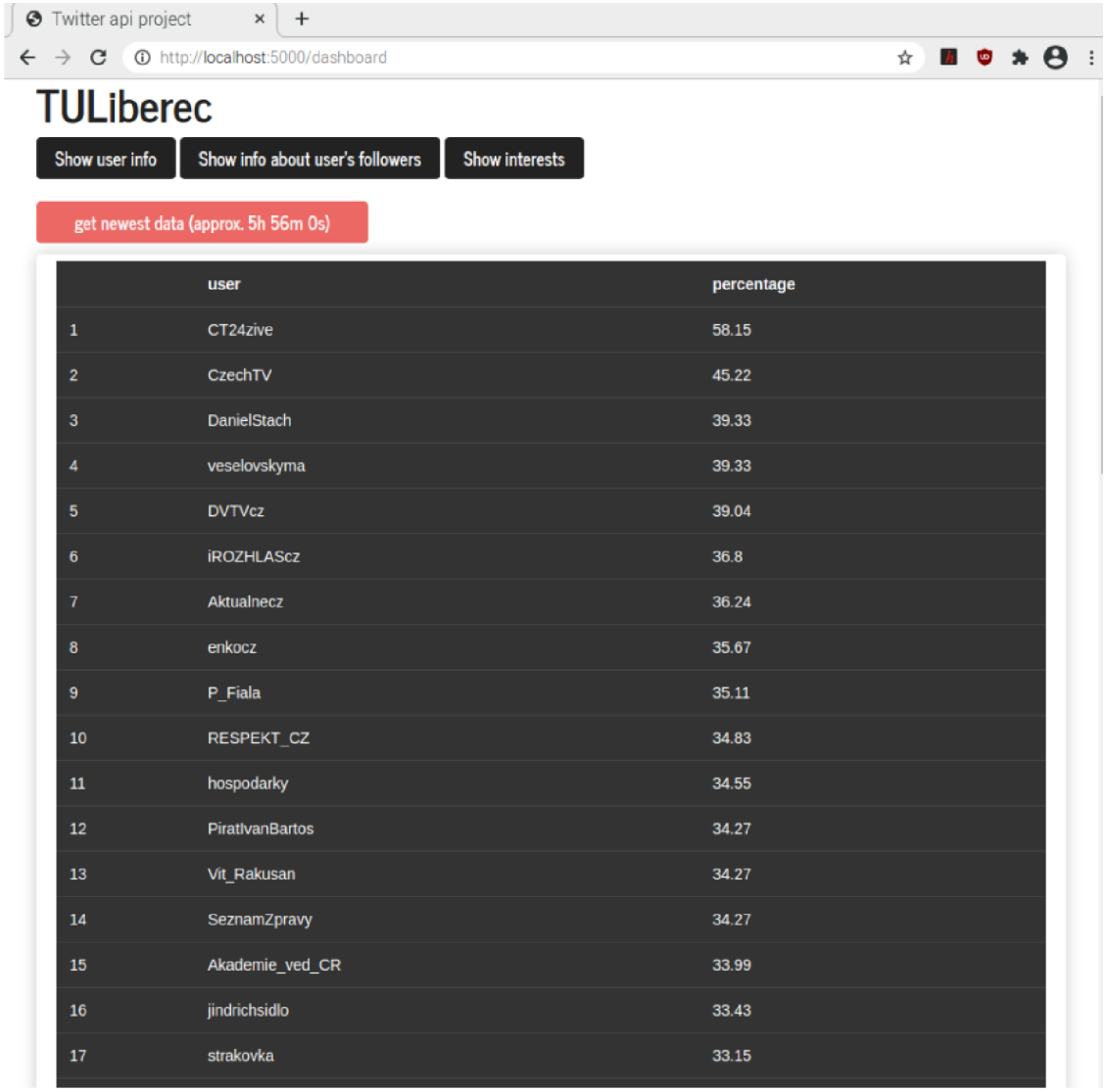
Obrázek 3.4: Informační panel pro twitterový účet @TULiberec, zobrazená analýza základních informací o účtu @TULiberec



Obrázek 3.5: Informační panel pro twitterový účet @TULiberec, zobrazená analýza informací o followerech účtu @TULiberec



Obrázek 3.6: Informační panel pro twitterový účet @TULiberec, zobrazená interaktivní mapa informující uživatele o geografii followerů účtu @TULiberec



Obrázek 3.7: Informační panel pro twitterový účet @TULiberec, zobrazená analýza společných followerů (zájmů) účtu @TULiberec s jinými účty

4 Popis aplikace

4.1 Přístup k Twitter API

Jak již bylo v této práci zmíněno, ke správné funkčnosti této aplikace potřebujeme mít přístup k Twitter API. Toto API však většinu svých informací neposkytuje anonymně a abychom mohli s API plně komunikovat, potřebujeme získat od Twitteru povolení. K této autorizaci je zapotřebí nejprve mít Twitter účet a poté požádat o vytvoření vývojářského účtu svázaného s tímto normálním účtem. Tato žádost se vytváří prostřednictvím dotazníku od Twitteru, kde musí dotázaný (uživatel, žádající o vývojářský účet) odpovědět na pár otázek (jméno, příjmení, za jakým účelem hodlá uživatel používat přístup k Twitter API). Poté stačí čekat pár dnů, než bude uživatelova žádost posouzena a schválena (existují případy, kdy žádost schválena nebyla, ale v drtivé většině případů se tak nestane). Vývojářský účet potřebný pro tvorbu této aplikace byl schválen 7 dní po podání žádosti. Po schválení žádosti si může uživatel vygenerovat tzv. token (unikátní textový klíč určený pro autorizaci uživatele při požadavcích na Twitter API). Tento token je nutné připojit ke každému požadavku směřující na Twitter API z důvodu autentizace autorizovaného uživatele. Pro dotazy na Twitter API se používají tzv. HTTP požadavky typu GET (požadavek od klienta na určitý server a následné očekávání potřebné odpovědi). Kromě uživatelova tokenu je součástí požadavku i specifická adresa vedoucí ke koncovému bodu, který nám může předat chtěné informace a různé parametry definující, jaké informace chceme od Twitter API získat. Na obrázku 4.1 je vidět ukázkový požadavek na Twitter API. Jde o žádost o user objekt (informace o uživateli) twitterového účtu @TULiberec (oficiální twitterový účet TUL). V proměnné token je uložen autorizační klíč. [2]

```
url = 'https://api.twitter.com/1.1/users/show.json?screen_name=tuliberec'
headers = {'Authorization': 'Bearer ' + token}
response = requests.get(url, headers = headers).json()
```

Obrázek 4.1: Příklad uskutečnění HTTP requestu na Twitter API v programovacím jazyce Python 3

4.2 Struktura serveru

4.2.1 Webový server

Webový server je část aplikace, která má primárně za úkol komunikaci s uživatelem. Je naprogramován v Node.js a funguje jako takový mezičlánek mezi klientskou částí aplikace a s backendem, která je určená pro tvorbu jednotlivých analýz. webový server funguje na protokolu HTTP a na principu API tz. přijímá od klientů žádosti a náležitě na ně odpovídá.

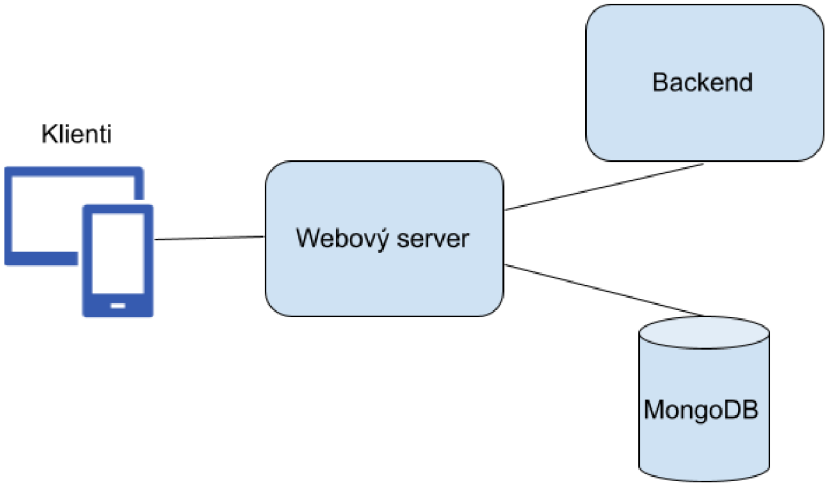
Moduly

Seznam modulů, které jsou pro vývoj webových aplikací v Node.js k dispozici a tato práce je používá:

- Express.js (Framework pro zjednodušení tvorby webových serverů v Node.js)
- Express ejs layouts (Přidává možnost pracovat s EJS soubory, které vývojářům umožňují možnost dynamické úpravy HTML souborů před tím, než jsou vyrenderovány uživateli na straně klienta)
- Express session (HTTP je bezestavový protokol, tento modul vývojářům umožní ukládat parametry o konkrétním klientovi, který provádí více žádostí na server)
- Mongoose (Tento modul umožňuje propojení serveru s MongoDB databází, více informací v kapitole 4.2.3)
- Socket.io (Umožňuje oboustrannou komunikaci mezi serverem a klientem, vhodné pro SPA [14])

Routes

Routes, neboli cesty, jsou v problematice vývoje webových aplikací chápány jako předepsání funkcionality pro určité URL adresy. V případě vyvíjené aplikace pro účely této práce není potřeba těchto cest příliš mnoho. Aplikace má totiž jen 2 aktivní stránky pro uživatele. Použitelné cesty v aplikaci jsou: Po zadání požadavku typu GET bez zadání koncového bodu (pouhé "/") je uživatel jednoduše přesměrován na úvodní stránku aplikace. Požadavek /dashboard typu POST přijímá jako parametr



Obrázek 4.2: Schéma aplikace

název twitterového účtu, který chceme v aplikaci zanalyzovat. Aplikace zkontroluje, zda daný účet na twitteru existuje a pokud ano, uživatel je přesměrován na informační panel a v aplikaci je pro tohoto klienta nastavený daný twitterový účet jako aktivně zkoumaný účet. Pokud daný účet neexistuje, uživatel je přesměrován zpět na úvodní stránku. Poté je možné učinit požadavek se stejným názvem jako požadavek předchozí (/dashboard), nicméně v tomto případě jde o požadavek typu GET. Po zavolání tohoto koncového bodu je klient přesměrován na informační panel, pokud již pro jeho instanci byl nastaven existující twitterový účet, který má být analyzován. V opačném případě je uživatel opět přesměrován na úvodní stránku. Poslední HTTP požadavek, který lze poslat na webový server je /cz typu POST. Tento požadavek klient použije v případě, že chce začít novou analýzu velkých twitterových uživatelů na českém twitteru. Tento požadavek přijímá 1 argument a tím je název počátečního twitterového účtu (více informací v kapitole 4.4.4). Toto jsou všechny HTTP požadavky, se kterými může klient komunikovat se serverem, není to však veškerá komunikace, která je v aplikaci možná, pro další požadavky se akorát používá jiný protokol, který je popsán v následující kapitole.

Socket.io

Modul Socket.io se používá pro komunikaci s klientem pomocí protokolu WebSocket, který umožňuje oboustrannou komunikaci mezi klientem a serverem v reálném čase (tzn. klient není omezen např. potřebou opakovaného načítání stejné stránky). Tento protokol je díky své vlastnosti velmi vhodný pro SPA. Z tohoto důvodu se tento modul používá např. při tvorbě chatovacích aplikací. Socket.io funguje na principu posluchač - vyvolávač. Posluchač neustále čeká na požadavek a ve chvíli, kdy požadavek přijde, je provedena náležitá, předem definovaná operace. Vyvolávač poté

posílá požadavky/odpovědi jednoho definovaného typu, který je na druhé straně zachycen příslušným posluchačem. Ukázková komunikace může fungovat např. Klient, pomocí určitého vyvolávače, pošle na server požadavek, který jeden ze serverových posluchačů zachytí. Poté je provedena příslušná operace a pokud je potřeba, určitý serverový vyvolávač pošle odpověď zpět ke klientovi. V případě aplikace, kterou se tato práce zabývá existuje hned několik posluchačů/vyvolávačů na straně serveru i klienta. Na straně serveru jsou posluchači, kteří reagují na situace kdy: Klient chce zjistit, jak dlouho budou trvat jednotlivé analýzy pro konkrétní twitterový účet, chce zobrazit poslední vytvořenou analýzu konkrétního typu pro konkrétní twitterový účet a nebo chce si nechat vytvořit novou konkrétní analýzu pro konkrétní twitterový účet. Všichni tyto posluchači mají své příslušné vyvolávače, kteří pošlou klientovi náležitou odpověď. Na straně klienta jsou v podstatě posluchači a vyvolávači ekvivalentní svým protějškům na serveru (vyvolávači, co volají výše popsané serverové posluchače a posluchači, čekající na odpovědi ze serveru). [14]

Model

V modelové části webové aplikace se nachází předpis objektu twitterového účtu s atributy obsahující výsledky jednotlivých analýz daného účtu (za předpokladu, že byli někdy v minulosti již alespoň jednou pro daný účet provedeny). Samotné objekty, které model předepisuje, jsou uloženy v databázi MongoDB (více informací v kapitole 4.2.3).

Volání backendu

Webový server v Node.js má primárně za úkol komunikovat s klienty. Není tedy vhodné, aby na stejném serveru běželi i všechny výpočty nutné k vypracování analýz, které aplikace nabízí, vzhledem k tomu, že by to mohlo server značně zpomalovat a tím pádem i negativně ovlivnit UX. Z tohoto důvodu byl pro účely této práce vyvinut ještě backendový server. Tento server dostává od webového Node.js serveru požadavky, na které reaguje tím, že začne vytvářet např. nové analýzy a poté získané výsledky pošle zpět jako odpověď (více informací v kapitole 4.2.2). S tímto backendem dokáže webový server komunikovat za použití předem nadefinovaných požadavků společně s příslušnými parametry, např. požadavek na zjištění 30-ti největších followerů twitterového účtu TUL s parametrem nesoucí název účtu, v tomto případě "TULiberec". Pro komunikaci mezi těmito dvěma servery se používá protokol TCP. Základní nastavení je vytvořené pro situaci, kdy je jak webový, tak i backendový server spuštěný na stejném zařízení. Pokud uživatel, který instaluje tuto aplikaci, z nějakého důvodu preferuje mít tyto dva servery na dvou různých zařízeních/sítích, může tak samozřejmě učinit a aplikace bude fungovat, jen musí ve zdrojovém kódu na patřičném místě změnit adresu backendového serveru a učinit patřičná opatření k umožnění komunikace (port forwarding, firewall atp.).

4.2.2 Backend

Backend (analytická část aplikace), jak už bylo v rychlosti zmíněno v předchozí kapitole, je server, který má na starosti všechny velké výpočty, tvorbu analýz a komunikaci s Twitter API. Je naprogramován čistě v jazyce Python 3 a skládá se z několika skriptů, kteří mezi sebou komunikují.

Hlavní skript

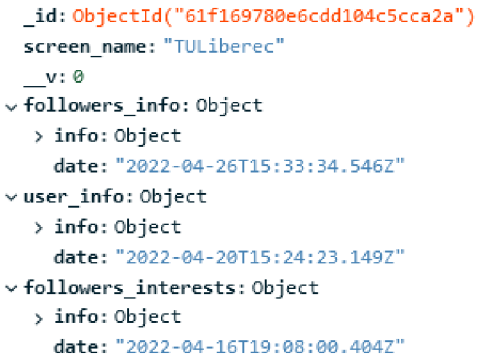
Hlavní skript má název server.py a zajišťuje komunikaci s ostatními články aplikace. I přesto, že je aplikace navržená tak, že pouze webový server komunikuje s backendem, tento skript v podstatě vůbec neřeší, od koho daný požadavek přišel, pouze ho vykoná a patřičně na něj odpoví. Server.py je jediný skript, který je potřeba spustit na patřičném zařízení k tomu, aby backendový server fungoval. Tento skript se nikdy sám od sebe nevypne (pokud nedojde k neočekávané chybě) a v ideálním případě by ho měl případný správce aplikace nechat zapnutý na dobu neurčitou. Všechny ostatní skripty už si čte hlavní skript sám a není potřeba s nimi nic dělat. Hlavní skript funguje na principu neustálého čekání na případný požadavek s parametry a následné vykonání patřičné akce a odeslání odpovědi. Mezi tyto požadavky patří: Zjištění, zda daný twitterový účet existuje, výpočet doby trvání jednotlivých analýz pro daný účet, zjištění základních informací o účtu, zjištění zajímavých informací o followerech daného účtu, provedení analýzy zájmů followerů daného účtu a pokyn pro vytvoření souborů potřebných k vytvoření grafu velkých českých twitterových uživatelů a jejich společných followerů. Ke každému z těchto požadavků patří příslušný skript, který hlavní skript může spustit.

4.2.3 MongoDB

MongoDB je typ databáze, který je využit pro účely této aplikace. Na cloudovém uložišti od společnosti MongoDB [9] je vytvořen tzv. databázový cluster (jeden cluster pro jeden účet je zadarmo), ve kterém je jedna databáze, která slouží pro ukládání výsledků všech vykonaných analýz pro všechny twitterové uživatele, společně s datem, kdy byla daná analýza vykonána. Pokud je vykonána analýza, která byla již v minulosti uložena, výsledky se v databázi přepíší na ty novější. Pokud je v aplikaci zkoumán nový uživatel, vytvoří se nový dokument v databázi, obsahující informace o daném uživateli. V obrázku 4.3 lze nahlédnout do struktury objektu nesoucí informace o analýzách twitterového účtu @TULiberec, který je uložen v databázi.

4.3 Načtení tokenu

Důležitou součástí aplikace je tzv. token, který je klíčový pro přístup k Twitter API. Bez tohoto tokenu přístup k Twitter API není možný. Pro plné používání aplikace musí uživatel daný token vlastnit (více informací v kapitole 4.1). Pokud již uživatel token má, musí ho v aplikaci uvést, aby mohl být použit při získávání dat



Obrázek 4.3: Ukázka objektu uloženého v databázi MongoDB, nesoucí informace o analýzách twitterového účtu @TULiberec

potřebných pro analýzy. Pro tento účel slouží textový řádek na úvodní stránce, do kterého je třeba napsat požadovaný token.

4.4 Popis tvorby analýz

Analýzy jsou počítány, jak už bylo v této práci zmíněno, v backendovém serveru a jsou naprogramované čistě v jazyce Python 3. Pro každou analýzu existuje samostatný Python skript, který je v případě potřeby volán hlavním skriptem.

4.4.1 Informace o uživateli

První analýza zjišťuje základní informace o daném twitterovém uživateli, jehož jméno je uvedeno v parametru. Postup této analýzy je následující. Podle jména účtu si algoritmus stáhne z Twitter API objekt uživatele obsahující informace. Poté zkontroluje, zda není účet nastaven jako soukromý (pokud ano, nelze z něj získat data, protože si to majitel twitterového účtu nepřeje). Dalším krokem je vytažení zajímavých informací ze získaného objektu a drobný výpočet některých údajů (např. průměrná tweetovací frekvence). Součástí algoritmu je i odhad, zda je daný uživatel falešný, či nikoliv. Získaná data jsou následně předána hlavnímu skriptu, který je odešle.

4.4.2 Informace o followerech

Tato analýza poskytuje informace o followerech twitterového účtu, který je uveden v parametru. Mezi tyto informace patří: zjištění největších followerů daného účtu, geografie followerů a odhad podílu falešných a neaktivních followerů. Algoritmus si nejprve zjistí informace o uživateli, kterého právě zkoumá, poté se rozdělí na 2 vlákna. První vlákno stahuje z Twitter API identifikátory followerů (unikátní číselné označení uživatelů na twitteru), přičemž pro jedno vykonané volání získá až

5000 těchto identifikátorů. Algoritmus provádí přesně jedno volání za minutu, aby dodržel rate limits nastavené v Twitter API. Tento proces se opakuje až do chvíle, kdy jsou staženy identifikátory všech followerů daného účtu. Druhé vlákno poté obdrží identifikátory získané z vlákna prvního a použije je pro stažení user objektů všech followerů. Proces může stahovat až 100 followerů každé 3 sekundy. Po stažení všech user objektů jsou poté data analyzována. Algoritmus je rozdělen do 2 vláken, protože provádějí na sobě nezávislé operace a urychlí se tím celý proces analýzy.

Největší followeři

Součástí této analýzy je zisk seznamu až 30-ti největších followerů zkoumaného účtu (followeři s největším počtem vlastních followerů). Tato část analýzy je důvod proč v tomto případě nelze použít relativní četnost pro zkrácení doby trvání této analýzy. Kdyby byla použita a z Twitter API by byla stažena jen část followerů, nemůžeme s jistotou nalézt všechny největší followery, či je uspořádat do správného pořadí odpovídající realitě.

Geografie followerů

Každý uživatel na Twitteru má možnost uvést na svém profilu svojí lokaci. Pokud tak učiní, jsme schopni z jejich user objektu tuto lokaci zjistit a následně s ní pracovat. Tento atribut je využit pro analýzu geografie followerů, která nám říká, z jakých zemí jsou followeři daného účtu a kolik. K analýze lokací byly využity JSON soubory dostupné na odkazu [15] a [16]. Tyto soubory obsahují kompletní seznam všech států a velkých měst v těchto státech. Pomocí těchto dokumentů lze zjistit z jaké země pravděpodobně pochází daný follower v případě, že má jako svojí lokaci uvedený stát, či město v nějakém státě. V tomto případě je však potřeba počítat s faktem, že velká část uživatelů na Twitteru nemají uvedenou svojí lokaci, nebo mají jako svojí lokaci něco neidentifikovatelného. Takoví uživatelé se do geografie nezapočítají a tím pádem je reálný výsledek této části analýzy pravděpodobně trochu odlišný. K zobrazení geografie followerů je poté použita interaktivní mapa (více informací v kapitole 3.2.2).

Odhad falešných a neaktivních followerů

Poslední součástí této analýzy je odhad podílu falešných a neaktivních followerů vůči zbytku. Je důležité zmínit, že problematika detekce falešných účtů na jakékoliv sociální platformě je velice složitá a komplexní. Existuje velký počet způsobů, jak tuto detekci provádět, přičemž každý má své výhody, nevýhody, časovou náročnost a ve finále i výsledky. Vzhledem k neustálému vývoji falešných účtů a i způsobů pro jejich detekci nemůžeme téměř nikdy s jistotou označit konkrétní účet jako falešný. Vědecký článek *Fame for sale: Efficient detection of fake Twitter followers* [17] se touto problematikou zabývá velice podrobně. Nicméně analýza tvořená v této aplikaci se zabývá touto problematikou pouze povrchně, v podstatě jde jen o kontrolu zda mají jednotliví followeři profilový obrázek, popis účtu, lokaci, stáří účtu, zda-li aktivně tweetují nebo jestli nesledují veliký počet uživatelů v porovnání

s počtem jejich followerů. Pokud dostatečný počet těchto kontrol neprojde, follower je označen za falešný.

4.4.3 Zájmy followerů

Poslední analýzou pro konkrétní twitterové uživatele je analýza zájmů followerů. V podstatě jde o zjištění, kolik procent z followerů daného účtu sleduje ostatní twitterové účty. Výsledkem je seznam až 50-ti účtů, se kterými má zkoumaný účet vysoký počet společných followerů. Pokud je v seznamu např. účet X se 35% společných followerů, znamená to, že právě 35% lidí, kteří sledují zkoumaný účet, sledují také účet X. Algoritmus této analýzy je rozdělen do 2 vláken. První vlákno získává identifikátory všech followerů zkoumaného účtu (5000 identifikátorů za minutu, stejně jako v předchozí analýze). Druhé vlákno poté pro každého followera (pokud není definováno jinak) stáhne identifikátory všech účtů, který daný follower sleduje (pokud nejde o soukromý účet, v takovém případě je follower ignorován). Pouze 1 uživatel za 1 minutu lze z Twitter API získat. Po stažení všech dat se všechny identifikátory porovnají a ty, co se ve stažených datech nacházejí v největším množství jsou právě ty identifikátory, které hledáme. Tyto identifikátory jsou použity pro stažení korespondujících user objektů a ty jsou poté použity jako výsledek analýzy. Tato analýza je časově nejnáročnější, lze tu však použít relativní četnost pro zkrácení čekací doby. Podle relativní četnosti nemusíme kontrolovat všechny followery ale pouze část (např. 10%). Výsledek by se neměl nijak zásadně lišit od skutečnosti (více informací v kapitole 2.4). V případě této aplikace je algoritmus navržen tak, aby pro zkoumané účty do 1440 followerů byly zkoumáni všichni followeři a pro uživatele nad 1440 followerů jen takové procento followerů, aby analýza netrvala déle než 1 den (např. pro uživatele se 14 400 followery bude provedena pouze 10% analýza). Algoritmus počítá s tzv. nezávislým výběrem followerů v případě použití relativní četnosti (např. nemůže dojít k výběru pouze nových followerů atp.).

4.4.4 Postup analýzy velkých českých twitterových uživatelů a jejich společných followerů

Tato analýza, na rozdíl od analýz výše představených, již není pro jednoho konkrétního uživatele, nýbrž pro celou českou část Twitteru. Výstupem této analýzy je neorientovaný graf, jehož vrcholy představují jednotlivé české velké uživatele a hrany mezi nimi představují jejich společné followery. Tento graf má takovou vlastnost, že uživatelé, kteří mají relativně stejné publikum jsou umístěni blízko sebe, zatímco uživatelé s malým počtem společných followerů jsou od sebe v grafu daleko.
Algoritmus této analýzy přijímá jeden vstupní parametr a tím je název jednoho českého uživatele. Zde je na místě upozornit, jak tohoto vstupního uživatele vhodně zvolit. Musí jít o českého uživatele a nejlépe s co největším počtem followerů. Nejlepší volba je v současné době (duben 2022) účet @CT24zive. Tento účet je poté použit při startu analýzy. Algoritmus stáhne user objekty všech followerů daného účtu a pro každého provede několik kontrol. Nejprve zkontroluje, zda daný follower dosahuje na předem nastavený minimální počet followerů (20 000). Pokud je první

kontrola úspěšná, je třeba ještě zkontrolovat, zda jde o českého uživatele, či nikoliv. Pro tento účel skript stáhne 200 posledních tweetů od kontrolovaného uživatele a zkontroluje, jakým jazykem jsou tweety napsané. Pokud alespoň polovina z nich je napsaná v českém jazyce, kontrola je úspěšná a uživatel je označen jako velký český twitterový uživatel. Nově objevení velcí uživatelé jsou uloženy do seznamu pro budoucí scrape. Z tohoto seznamu se vytáhne nový uživatel ve chvíli, kdy byl právě scrapovaný uživatel celý zkontrolován. Důležité je také zmínit, že pokud byl těmito dvěma kontrolami nalezen uživatel, který již byl scrapovaný, nebo je přichystáný v seznamu pro budoucí scrape, je již ignorován. Tímto způsobem jsou postupně prohledáni všichni nalezení uživatelé.
Ve chvíli, kdy je seznam pro budoucí scrape prázdný, znamená to, že již nebyl nalezen žádný nový velký uživatel a algoritmus přejde do druhé fáze, počítání společných followerů. V této fázi se pro každou kombinaci nalezených uživatelů a jejich followerů provede sčítání společných followerů. Opět lze využít relativní četnost a není tudíž potřeba kontrolovat všechny followery, ale pouze část (úspora času). Seznam všech kombinací 2 uživatelů a jejich počet společných followerů je uložen do paměti a algoritmus přechází do fáze třetí. V této fázi už jen pouze dojde k vytvoření vrcholů a hran pro budoucí graf. Pokud mají 2 různí uživatelé alespoň 20 000 společných followerů, je mezi nimi vytvořena hrana. Poté je algoritmus u konce a výstupem jsou dva soubory typu csv (nodes.csv obsahující vrcholy a edges.csv obsahující hrany).

Vytvoření grafu

Pro vytvoření a následné zobrazení grafu je potřeba použít program Gephi (popis programu v kapitole 2.5.3). V programu je nutné nejprve nahrát získané soubory z analýzy do laboratoře dat. Poté zvolit vhodný algoritmus rozložení, nechat si spočítat centralitu vlastních vektorů a aplikovat jí na velikosti vrcholů (vlivnější uživatelé se stanou většími), nechat si spočítat modularitu a aplikovat ji na graf (graf se rozdělí do barevně odlišených skupin, přičemž každá skupina je do jisté míry nezávislá na ostatních). Podrobnější informace o postupu lze vyčíst v článku [18].

4.4.5 Počítání dob

Tato analýza počítá jak dlouho bude zhruba trvat tvorba jednotlivých analýz pro předem definovaného twitterového uživatele uvedeného v parametru. Tento skript dostane v parametru název účtu, kterého má zkoumat, podle jména si z Twitter API stáhne informace o daném účtu, zjistí, kolik má daný účet followerů a podle toho spočítá, jak dlouho bude jednotlivá analýza trvat. Tyto údaje se poté zobrazují uživateli, kterému jsou užitečné, protože díky ním ví, jak dlouho musí na dané výsledky čekat.
Pro zisk informací o uživateli nezáleží, kolik má daný uživatel followerů, proto analýza trvá pro všechny uživatele stejně dlouho a to zhruba 3 sekundy. V případě analýzy informací o followerech účtu (tj. největší followeři, geografie followerů a odhad falešných followerů) však už na počtu followerů záleží, čas, potřebný pro

vytvoření této analýzy se dá spočítat následovně (n značí počet followerů):

$$t(s) = \left(\frac{n}{100} + 1\right) \cdot 3 \tag{4.1}$$

Analýza zájmů followerů konkrétního twitterového uživatele je časově nejnáročnější. Pokud by k tvorbě této analýzy nebyla použita relativní četnost, potřebný čas by se dal vypočítat následovně:

$$t(s) = 60 \cdot n \tag{4.2}$$

Jak se při této analýze využívá relativní četnosti je popsáno v kapitole 4.4.3. Co se týče analýzy velkých českých twitterových uživatelů a jejich společných followerů, tak zde již neexistuje možnost, jak předem zjistit dobu trvání, jelikož nevíme kolik českých uživatelů vlastně najdeme, kolik budou mít followerů atp.

5 Ukázková analýza

5.1 Výběr účtu

Pro demonstraci funkčnosti této aplikace byly vytvořeny ukázkové analýzy. Pro výsledky analýzy velkých uživatelů na českém twitteru a jejich společných followerů stačí pouze pokyn k vytvoření analýzy. Pro výsledky analýz konkrétního twitterového uživatele je však nejprve potřeba zvolit ukázkový twitterový účet, na kterém budou dané analýzy provedeny. V tomto případě nenajdeme lepšího kandidáta, než twitterový účet TUL (@TULiberec), jelikož tato práce právě na této univerzitě vznikala. Tento účet měl v době vytvoření analýz (duben 2022) celkem 356 followerů (důležitý fakt pro určení čekacích dob pro jednotlivé analýzy). Všechny analýzy v následujících kapitolách byly provedeny ve vytvořené aplikaci, či za její pomoci.

5.2 Výsledky analýz účtu Technické Univerzity v Liberci

5.2.1 Základní informace o účtu

První analýza, jejíž výsledky jsou dostupné v tabulce 5.1, obsahuje základní informace o účtu Technické Univerzity v Liberci. Těmito informacemi jsou: počet followerů, počet přátel, počet tweetů, jméno, lokace účtu uvedená od uživatele, zda je daný účet "verified"(ověření od samotné platformy Twitter, zda je daný účet autentický), popis účtu, kdy byl účet vytvořen, průměrný počet lajků na tweetech daného účtu, průměrný počet retweetů na daném účtu, jazyk, ve kterém daný účet tweetuje, průměrná tweetovací frekvence (jak často daný účet v průměru tweetuje) a odhad, zda je účet pravý, či falešný/neaktivní. Informace získané z této analýzy nejsou zatím moc zajímavé, protože většinu z nich (vyjma posledních dvou uvedených) můžeme získat jednoduše např. obyčejným navštívením twitterového profilu TUL. Tato analýza trvala pouhé 3 sekundy.

5.2.2 Informace o followerech daného účtu

Druhá analýza účtu @TULiberec nám nabízí již poněkud zajímavější údaje. Tato analýza nám přináší spoustu informací o followerech daného účtu. Mezi tyto získané informace patří: List 30 největších twitterových uživatelů, kteří sledují zkoumaný

Základní informace	
Atribut	Hodnota
Počet followerů	356
Počet přátel	86
Počet lajků	126
Počet tweetů	93
Jméno	Technická univerzita v Liberci
Lokace	Liberec
Verified	False (ne)
Popis účtu	Technická univerzita v Liberci poskytuje nejen...
Datum vzniku	3. 3. 2020
Průměrný počet lajků	3
Průměrný počet retweetů	1
Jazyk	cs
Průměrná tweetovací frekvence	8d 6h 44m 15s
Falešný účet	False (ne)

Tabulka 5.1: Základní informace o účtu @TULiberec získané z aplikace

účet, seznam, kolik followerů je z jaké země (v podobě interaktivní mapy), hrubý odhad podílu followerů, kteří jsou buďto falešní nebo neaktivní, počet verified účtů a počet soukromých účtů. Pro zjištění, jak dlouho bude tato analýza trvat, již potřebujeme znát počet followerů zkoumaného účtu. Z první analýzy víme, že v případě účtu TUL je počet followerů 356 (duben 2022). V tomto případě analýza účtu @TULiberec trvala 15 sekund (Podrobnější informace o výpočtu čekací doby naleznete v kapitole 4.4.5).

Největší followeři

V tabulce 5.2 lze nahlédnout do výsledků provedené analýzy. V této tabulce je uvedeno 10 největších twitterových uživatelů (tj. s největším počtem followerů), kteří sledují účet TUL. Ve skutečnosti vytvořená aplikace nabízí až 30 největších uživatelů, nicméně do tabulky jich bylo uvedeno pouze 10 z důvodu malé relevantnosti uživatelů umístěných na nižších příčkách. Na prvním místě se umístil Daniel Stach (@DanielStach, moderátor z České Televize) s 86 189 followery. Jelikož je tento uživatel první v tabulce, znamená to, že právě tento uživatel je největší účet na twitteru, který sleduje účet @TULiberec. Na druhém místě je poté twitterový účet Ukrajinské ambasády v Praze (@UKRinCZE), který v první polovině roku 2022 zažil velký nárůst followerů (samozřejmým důvodem je aktuální válka na Ukrajině a podpora Ukrajiny napříč českou populací).

Geografie followerů

V obrázku 5.1 je možno vidět interaktivní mapu uživatelů, kterou vytvořená aplikace nabízí, jakožto součást analýzy followerů. Z mapy můžeme pohodlným způsobem získat informace o geografii followerů účtu @TULiberec. Čím více je daná země v

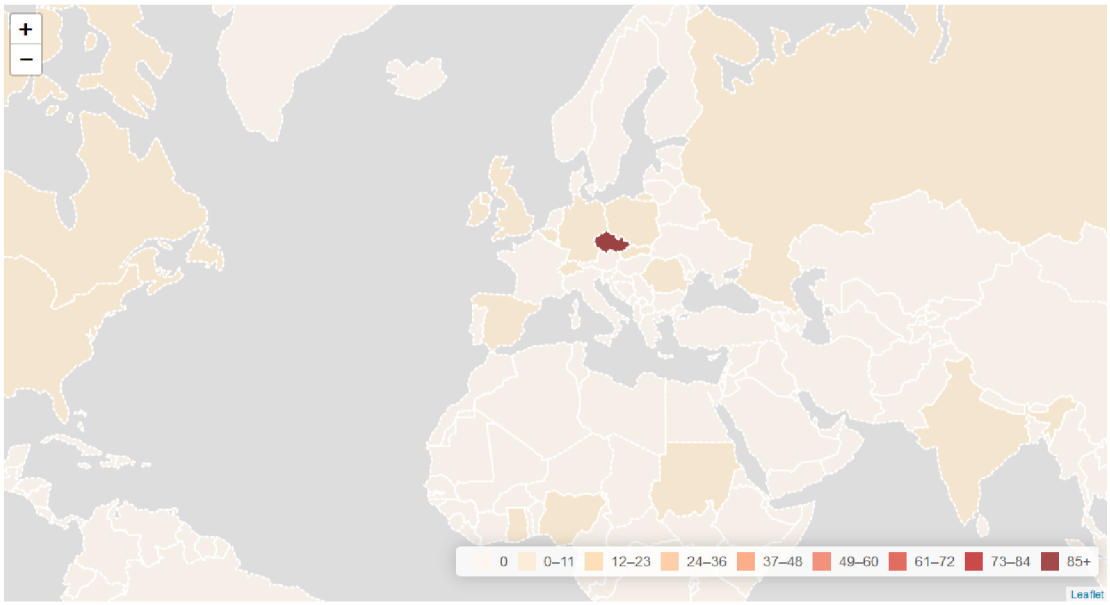
Největší uživatelé sledující @TULiberec		
Pořadí	Uživatel	Počet followerů
1	DanielStach	86 189
2	UKRinCZE	27 120
3	textileinst	14 890
4	Akademie_ved_CR	10 815
5	FF_CUNI	10 230
6	vedavyzkum_cz	6 470
7	MichaelFiala	5 038
8	karelpeka	4 839
9	CzechInvest_CZ	4 796
10	JanTuna	4 648

Tabulka 5.2: 10 největších twitterových uživatelů, kteří sledují účet @TULiberec. Informace získané z aplikace

mapě zabarvená do červena, tím více followerů daného účtu je právě z této země. Vpravo dole je přítomna legenda, která nás informuje o krajních počtech followerů, které připadají k určitému odstínu. Tyto hodnoty jsou pro každý zkoumaný účet jiné (záleží na celkovém počtu followerů). V případě účtu @TULiberec, který zkoumáme v naší ukázkové analýze, je patrné, že drtivá většina uživatelů, co sledují tento účet, jsou z České Republiky (122 followerů). Vzhledem k tomu, že TUL je česká univerzita a tweety píše tento účet v českém jazyce, výsledek je smysluplný. Ostatní země buďto nemají žádného followera @TULiberec, a nebo pouze v řádu jednotek (např. 1 slovenský follower, 1 polský follower, 3 němečtí followeři). Důležité je také zmínit, že u spoustu followerů jednoduše nelze poznat, odkud jsou (v případě @TULiberec je to 151 followerů), vytvořená aplikace se totiž může o lokaci specifického followera dozvědět pouze v případě, kdy ji on sám do svého profilu uvedl.

Ostatní informace

Tabulka 5.3 poté ukazuje další relevantní informace o followerech účtu TUL, které vytvořená aplikace umí získávat. Konkrétně zde můžeme vidět, že 2 uživatelé, co sledují účet @TULiberec jsou verified (prokázaná autentičnost uživatele od samotné platformy Twitter), 28 z followerů mají svůj účet nastavený na soukromý (neveřejný) a také je zde vidět výsledek analýzy podílu falešných/neaktivních followerů oproti těm skutečným a aktivním. Pro účet @TULiberec je to 18,26%, což je o 1,64% více než analýza od FollowerAudit (kapitola 2.6.2) a o 8,26% více než analýza od TwitterAudit (kapitola 2.6.3). Odhad falešných/neaktivních followerů od vytvořené aplikace se liší od odhadů konkurence. Jelikož nevíme, jak fungují algoritmy pro detekci falešných followerů u zmíněné konkurence, nemůžeme s jistotou říci, který odhad se nejvíce blíží skutečné hodnotě. K tomu musíme také počítat s faktem, že nikdy nemůžeme se 100% jistotou u žádného účtu říci, zda je falešný nebo ne. I přesto by však bylo dobré nahlížet na výsledek z vytvořené aplikace, v tomto případě, jako pouze na orientační.



Obrázek 5.1: Interaktivní mapa ukazující geografii followerů účtu @TULiberec

Informace o followerech účtu @TULiberec	
Atribut	Hodnota
Počet verified followerů	2
Počet soukromých followerů	28
Fake ratio	18,26%

Tabulka 5.3: Relevantní informace o followerech účtu @TULiberec získané z aplikace

5.2.3 Informace o zájmech followerů daného účtu

Třetí analýza twitterového účtu TUL nám umožňuje nahlédnout do zájmů followerů daného účtu. Výsledkem této analýzy je seznam až 50-ti twitterových účtů, se kterými má účet @TULiberec nejvíce společných followerů. Pro zjištění, jak dlouho bude tato analýza trvat, již potřebujeme znát počet followerů zkoumaného účtu. Z první analýzy víme, že v případě účtu TUL je počet followerů 356 (duben 2022). V tomto případě analýza účtu @TULiberec bude trvat 5 hodin a 56 minut (Podrobnější informace o výpočtu čekací doby naleznete v kapitole 4.4.5)
V tabulce 5.4 můžete nalézt výsledky analýzy účtu TUL. Tabulka obsahuje 15 uživatelů, kteří mají velký podíl společných followerů s @TULiberec. Na prvním místě v tabulce je účet ČT24 (@CT24zive, zpravodajský účet České televize) s 58,15% sledovanosti, tzn. 58,15% uživatelů, kteří sledují účet @TULiberec, sledují také účet @CT24zive. Dále je také patrné, že jelikož je tento účet na prvním místě v tabulce, která je setříděná od největší sledovanosti, můžeme říci, že právě s tímto účtem má TUL nejvíce společných followerů na Twitteru. Z tabulky je dále patrné, že uživatelé, sledující @TULiberec, sledují převážně české novináře a zpravodajské služby, vzhledem k tomu, že tyto 2 kategorie twitterových účtů v tabulce převládají. Jediní uživatelé, kteří se v tabulce nacházejí, ale nespadají do těchto

Koho sledují followeři @TULiberec		
Pořadí	Uživatel	Sledovanost (%)
1	CT24zive	58,15
2	CzechTV	45,22
3	DanielStach	39,33
4	veselovskyma	39,33
5	DVTVcz	39,04
6	iROZHLAScz	36,80
7	Aktualnecz	36,24
8	enkocz	35,67
9	P_Fiala	35,11
10	RESPEKT_CZ	34,83
11	hospodarky	34,55
12	PiratIvanBartos	34,27
13	Vit_Rakusan	34,27
14	SeznamZpravy	34,27
15	Akademie_ved_CR	33,99

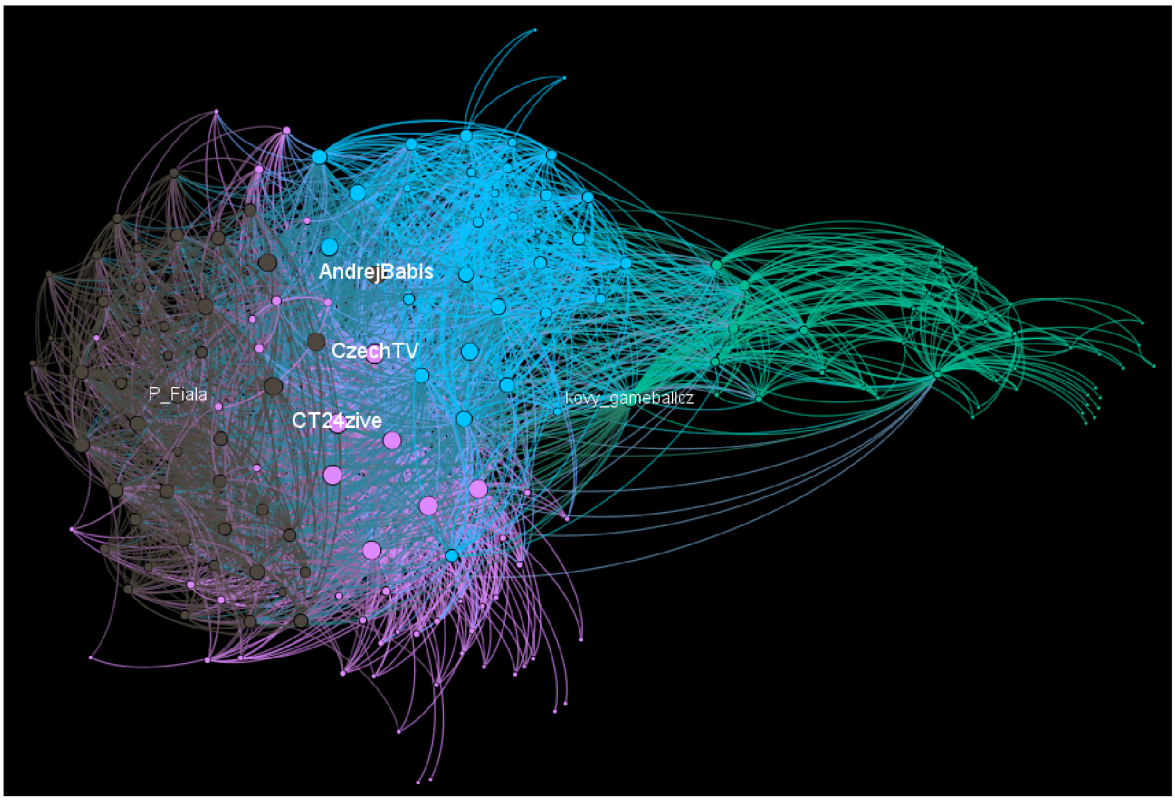
Tabulka 5.4: Kolik procent lidí, kteří sledují @TULiberec, sleduje ostatní twitterové účty

kategorií je Petr Fiala (@P_Fiala, současný předseda vlády České Republiky) na 9. místě, Ivan Bartoš (@PiratIvanBartos, předseda České pirátské strany) na 12. místě, Vít Rakušan (@Vit_Rakusan, předseda hnutí STAN) na 13. místě a Akademie věd České Republiky (@Akademie_ved_CR) na 15. místě.

5.3 Rozbor analýzy českého Twitteru

Tato analýza již nezkoumá jednoho konkrétního twitterového uživatele, jak to bylo u analýz předchozích, nýbrž zkoumá všechny velké české uživatele na twitteru a jejich společné followery. V této analýze se nejvíce používá analýza sociálních sítí a výsledkem je neorientovaný graf. Vrcholy u tohoto grafu znázorňují twitterové uživatele, kteří mají alespoň 20 000 followerů a jejich tweety jsou převážně v českém jazyce. Pokud počet společných followerů u dvou různých uživatelů v grafu přesáhne určitou nastavenou minimální hodnotu, objeví se mezi nimi hrana, která je spojuje a znázorňuje právě jejich společné followery. Výsledkem je poté graf, ve kterém se uživatelé s relativně stejným publikem objeví blízko sebe, zatímco uživatelé, kteří nemají žádné, nebo jen poměrně málo společných followerů, se objeví v grafu od sebe daleko. Vytvořená aplikace poskytne pouze data o vrcholech a hranách výsledného grafu. Tyto data je poté vhodné importovat do programu Gephi, ve kterém je možné graf vizualizovat a patřičně upravit např. použitím vhodného algoritmu pro rozložení grafu, výpočtem vlastních vektorů, kterým můžeme spočítat centralitu jednotlivých vrcholů a následně ji použít na určení velikosti vrcholu, výpočtem modularity, který můžeme v programu Gephi provést a následně aplikovat na graf. Výsledkem je poté

rozdělení vrcholů v grafu do skupin, přičemž každá ze skupin je na ostatní skupiny do jisté míry nezávislá. Tyto skupiny je vhodné v grafu barevně rozdělit.
V obrázku 5.2 je vidět výsledek této analýzy z července roku 2021. V tomto případě jsou v grafu celkem 4 skupiny, které jsou barevně rozlišeny. Hnědí uživatelé jsou politici a zelení uživatelé jsou různí influenceři. Skupiny modrých a růžových uživatelů jsou si navzájem nejpodobnější, v obou případech jde o různé novináře a zpravodajské služby. Jediný rozdíl mezi nimi je, že růžová skupina se zdá býti více politicky zaměřená, nicméně není to tak u všech. Je tedy možné argumentovat, že modrá a růžová skupina by mohla být spojena do jedné skupiny. Tyto skupiny byli však nalezeny automaticky pomocí výpočtu modularity, a tak můžeme vidět, že to není bezchybný způsob. Je to však způsob, který značně ušetří čas i práci v porovnání s manuálním zjišťováním, do jaké skupiny patří jednotliví uživatelé v grafu. V obrázku je vyznačeno pár zajímavých twitterových uživatelů (manuálně přidáno po vzniku výsledného grafu). Jsou to v první řadě @CT24 (zpravodajský účet České televize), @CzechTV (oficiální účet České televize) a @AndrejBabis (v době vzniku grafu předseda vlády České Republiky). To jsou uživatelé s poměrně velkým počtem followerů a tím pádem je do jisté míry sledují lidé ze všech zájmových skupin. Proto se v grafu nacházejí relativně uprostřed. Vlevo je dále vidět vrchol, představující účet @P_Fiala (v době vzniku grafu předseda strany ODS), který je uprostřed hnědých vrcholů (politiků). Z toho vyplývá, že Petr Fiala má na twitteru publikum, které se převážně zajímá o politiku. Nakonec vlevo se nachází účet @kovy_gameballcz (populární český youtuber), který je modularitou správně označen jako zelený (influencer), nicméně oproti ostatním influencerům je v grafu blíže ostatním skupinám. Ostatní influenceři jsou v grafu nejvíce izolovaní a z toho vyplývá fakt, že influenceři mají na českém twitteru nejizolovanější publikum (člověk, který sleduje na twitteru české influencery pravděpodobně nesleduje zbytek české scény na twitteru). Youtuber Kovy očividně funguje jako takový mezičlánek mezi influencery a zbylými skupinami.



Obrázek 5.2: Výsledek analýzy velkých českých twitterových uživatelů a jejich společných followerů v podobě neorientovaného grafu

6 Závěr

Tato bakalářská práce se zabývá vytvořením webové aplikace sloužící ke tvorbě automatických analýz uživatelů na sociální síti Twitter. Tyto analýzy pracují s veřejně dostupnými daty a jejich výsledky poskytují informace, které jsou nad rámec základních analýz poskytované od samotné platformy Twitter a mohou být několika způsoby užitečné pro uživatele aplikace. Nejprve je popsána analýza sociálních sítí, což je proces zkoumání vztahů ve skupinách lidí a jak je konkrétně použita pro tvorbu analýz v této práci. Poté se práce zabývá tzv. Twitter API, které umožňuje přístup k veřejně dostupným informacím o twitterových uživatelích potřebné k analýzám. Podobné služby a jak se liší od webové aplikace navržené v této práci jsou zde také zmíněny.
Další kapitola se věnuje popisu uživatelského rozhraní vytvořené webové aplikace, jak funguje a jaké analýzy svým uživatelům nabízí. První analýzou je zisk základních informací o konkrétním uživateli na Twitteru (jméno, počet followerů, popis účtu atp.). Druhá analýza poskytuje informace o followerech konkrétního twitterového účtu. Mezi tyto informace patří: List nejsledovanějších uživatelů, kteří sledují daný účet, ukázka, kolik followerů je z dané země (v podobě interaktivní mapy), kolik followerů mají soukromý účet, ověřený účet a hrubý odhad podílu followerů, kteří jsou buďto falešní nebo neaktivní. Třetí analýza nabízí pohled do zájmů followerů daného účtu. Výsledkem této analýzy je seznam twitterových účtů, se kterými má zkoumaný účet nejvíce společných followerů. Z tohoto seznamu je patrné, kolik procent lidí, kteří sledují zkoumaný účet, sledují ostatní účty. Poslední analýzou je analýza velkých českých twitterových uživatelů a jejich společných followerů. Výstupem této analýzy je neorientovaný graf, jehož vrcholy jsou velcí čeští uživatelé a hrany mezi nimi znázorňují určitý dosažený počet společných followerů mezi nimi. Uživatelé, kteří mají podobné publikum se v grafu umístí blízko sebe zatímco uživatelé s malým počtem společných followerů budou od sebe daleko.
Třetí kapitola popisuje aplikaci po technické stránce, jak je naprogramovaná, její struktura, jakým způsobem přistupuje k Twitter API atp. Popis algoritmů, které se používají k analýze uživatelů a jak fungují je zde také zmíněno. Největší překážkou při vývoji aplikace byly tzv. rate limits na Twitter API (omezení počtu požadavků, které lze provést za určitý časový úsek). Některé analýzy kvůli tomuto omezení trvají velmi dlouho. Tento problém byl však minimalizován pomocí relativní četnosti a ukládání již existujících výsledků analýz do databáze (umožnění pozdějšího přístupu k výsledkům).
Poslední kapitola se nejprve věnuje výběru vhodného twitterového uživatele pro vytvoření ukázkových analýz. Byl zvolen účet @TULiberec (oficiální účet TUL) a

všechny analýzy určené pro konkrétní uživatele byli pro tento účet vytvořeny. Poté se tato aplikace zabývá výsledky daných analýz. Některé získané informace z analýz odhalily zajímavé poznatky o tomto účtu jako např. největší twitterový uživatel, který sleduje @TULiberec je Daniel Stach (moderátor z České Televize) nebo lidé, kteří sledují tento účet, sledují hlavně zpravodajské služby. Jiné objevené informace spíše prokazují správnou funkčnost aplikace, např. bylo zjištěno, že drtivá většina followerů účtu @TULiberec jsou češi. Tento výsledek je očekávaný, jelikož TUL je česká univerzita a tweety na tomto účtu jsou v drtivé většině v českém jazyce. Neorientovaný graf s výsledky zmíněné analýzy společných followerů velkých českých uživatelů je zde také ukázán a poskytuje zajímavý pohled do stavu současné scény na českém Twitteru. Mezi hlavní poznatky, které analýza objevila na českém twitteru jsou skupiny s velkým počtem společných followerů (politici, novináři/zpravodajské služby a influenceři), jací uživatelé mají vliv na hodně zájmových skupin a jací jsou naopak více izolovaní.
Tato bakalářská práce splňuje své zadání, je tu však prostor pro možné rozšíření např. o sofistikovanější detekci falešných účtů, jež v dnešní době představují velký problém.

Seznam obrázků

Seznam tabulek

Literatura

[1] BUŠTÍKOVÁ, Lenka. Analýza sociálních sítí. *Sociologický Časopis / Czech Sociological Review* [online]. 1999, **35**(2), 1-2 [cit. 2022-05-02]. Dostupné z: http://www.jstor.org/stable/41131461

[2] Getting Started. *Twitter Developer Platform* [online]. San Francisco: Twitter, 2022 [cit. 2022-05-02]. Dostupné z: https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api

[3] Data dictionary. *Twitter Developer Platform* [online]. San Francisco: Twitter, 2022 [cit. 2022-05-02]. Dostupné z: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model

[4] Rate limits. *Twitter Developer Platform* [online]. San Francisco: Twitter, 2022 [cit. 2022-05-02]. Dostupné z: https://developer.twitter.com/en/docs/twitter-api/rate-limits

[5] LITSCHMANNOVÁ, Martina. *Úvod do statistiky (interaktivní učební text)* [online]. Ostrava, 2012 [cit. 2022-05-02]. Dostupné z: https://mi21.vsb.cz/sites/mi21.vsb.cz/files/unit/interaktivni_uvod_do_statistiky.pdf. Učební text. Technická univerzita Ostrava.

[6] MÁCA, Jindřich. Úvod do Node.js. *Itnetwork* [online]. Praha: Itnetwork, 2018, 2018 [cit. 2022-05-02]. Dostupné z: https://www.itnetwork.cz/javascript/nodejs/uvod-do-nodejs/

[7] What is Python? Executive Summary. *Python* [online]. Python [cit. 2022-05-02]. Dostupné z: https://www.python.org/doc/essays/blurb/

[8] About Gephi. *Gephi* [online]. France: Gephi.org, 2008 [cit. 2022-05-02]. Dostupné z: https://gephi.org/about/

[9] BOTELHO, Bridget. MongoDB. *TechTarget* [online]. TechTarget, 2020, 2020 [cit. 2022-05-02]. Dostupné z: https://www.techtarget.com/searchdatamanagement/definition/MongoDB

[10] *Leaflet* [online]. Kyiv: Leaflet, 2010 [cit. 2022-05-02]. Dostupné z: https://leafletjs.com/

[11] *Tweepsmap* [online]. Toronto: Tweepsmap, 2022 [cit. 2022-05-02]. Dostupné z: https://tweepsmap.com/

[12] *Followeraudit* [online]. Algodom Media LLP, 2022 [cit. 2022-05-02]. Dostupné z: https://www.followeraudit.com/

[13] *Twitteraudit* [online]. Twitteraudit, 2012 [cit. 2022-05-02]. Dostupné z: https://www.twitteraudit.com/

[14] The WebSocket API (WebSockets). *Developer Mozilla* [online]. Mountain View, USA: Mozilla, 2022 [cit. 2022-05-02]. Dostupné z: https://developer.mozilla.org/en-US/docs/Web/API/WebSockets_API

[15] BELLANGER, Félix. Countries.json. *Github* [online]. San Francisco: Github, 2012 [cit. 2022-05-02]. Dostupné z: https://gist.github.com/keeguon/2310008

[16] DUFOUR, Johan. Cities.json. *Github* [online]. San Francisco: Github, 2018 [cit. 2022-05-02]. Dostupné z: https://github.com/lutangar/cities.json

[17] CRESCI, Stefano. Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems* [online]. 2015, 2015, **80** [cit. 2022-05-02]. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0167923615001803

[18] LAMBERSON, PJ. Collecting and Visualizing Twitter Network Data with NodeXl and Gephi. *Social Dynamics* [online]. Social Dynamics, 2012 [cit. 2022-05-02]. Dostupné z: http://social-dynamics.org/twitter-network-data/