Západočeská univerzita v Plzni

Filozofická fakulta

DIPLOMOVÁ PRÁCE

Automatická detekce trollů na českém Twitteru

Bc. Denisa Míková

Plzeň, 2021

Západočeská univerzita v Plzni

Filozofická fakulta
Katedra filozofie

DIPLOMOVÁ PRÁCE

Automatická detekce trollů na českém Twitteru

Autor práce: Bc. Denisa Míková
Studijní obor: Teorie a filozofie komunikace

Vedoucí práce: Doc. Ing. Pavel Ircing, Ph.D.
Pracoviště vedoucího práce: Katedra kybernetiky ZČU
Plzeň, 2021

Prohlašuji, že já, Denisa Míková, jsem tuto diplomovou práci zpracovala samostatně a použila jsem pouze uvedené prameny a literaturu.

V Plzni, 2021 ……..

Bc. Denisa Míková

Tímto bych ráda poděkovala panu Doc. Ing. Pavlovi Ircingovi, Ph.D., za odborné vedení mé diplomové práce, trpělivost, ochotu, vstřícnost a za rady, které mi byly v průběhu kompletování dány.

Abstrakt

Každý den se na sociálních sítích vytvoří tisíce nových profilů. Lidé na nich sdílí svůj každodenní život, inspirují se, rozšiřují si obzory nebo udržují kontakt se svými blízkými. Každým dnem sociální sítě nabývají na popularitě. Není proto divu, že se najdou i tací, kteří nacházejí zálibu v obtěžování ostatních a záměrně na internet přicházejí s cílem zkazit ostatním uživatelům zážitek z užívání sítě. V internetovém žargonu hovoříme o trollech. V této práci je definován fenomén trollingu, charakteristické rysy chování a metody užívané pro jeho detekci. Taktéž je část práce věnována celosvětově rozšířené sociální sítí Twitter. Objasňuje strukturu Twitteru a vysvětluje metody získávání dat. Hovoří o tom, proč je dobré monitorovat společnost pomocí této sociální sítě a věnuje se politické debatě, která se na Twitteru odehrává. Přesah práce spočívá v testování metod automatické detekce trollů v rámci českého Twitteru, přesněji na účtech českých politiků.

Abstract

There are thousands of new social media accounts created on a daily basis. People share their everyday life, get inspired, widen their perspective or maintain contact with their loved ones. Every day, we can observe the growing popularity of social media. Therefore, there is no wonder we can find online users whose primary goal is to spoil the environment for others just for their own amusement. We describe those with a term *troll*. In this thesis, we define the phenomenon of trolling, we describe characteristic features of a troll and his behavior and methods used for their detection. Part of this thesis is devoted to the description of the favourite social media - Twitter - and its structure. We name and explain data mining methods and we clarify the importance of monitoring the society via Twitter. We also describe the political communication which takes place on Twitter. What makes this thesis unique is the testing of methods used for automatic troll detection in a scope of Czech Twitter, specifically Czech political accounts.

Klíčová slova

Twitter, Troll, detekce, automatická detekce trollů, umělá inteligence, politika

Keywords

Twitter, Troll, detection, automatic troll detection, artificial intelligence, politics

Obsah

TEORETICKÁ ČÁST **9**

Úvod **9**

Telekomunikační technologie **10**

Sociální média **11**

2.1. Vývoj sociálních médií 11

Twitter **12**

3.1. Vznik a vývoj 12

3.1.1. Twitter I 13

3.1.2. Twitter II 14

3.2. Struktura Twitteru 15

3.2.1. Odpovědi 16

3.2.2. Retweeting 17

3.2.3. Hashtagy 18

3.3. Sběr dat z Twitteru 18

3.4. Získávání informací z dat Twitteru 19

3.4.1. Odlišení Twitteru od ostatních zdrojů informací 19

3.4.2. Techniky získávání informací z dat Twitteru 20

3.5. Soudobá společnost na Twitteru 21

3.6. Twitter a politika 22

3.6.1. Vývoj politické komunikace 23

3.6.2. Politická komunikace na internetu 24

3.6.3. Obsah politické komunikace online 25

3.6.4. Čeští politici na Twitteru 26

3.7. Typy uživatelů na Twitteru 28

Troll **30**

4.1. Odlišení trolla od ostatních uživatelů 31

4.2. Komunikační vlákna s trollem 33

4.3. Napáchané škody 34

4.4. Jak se s trollem vypořádat? 35

4.5. Detekovací metody 36

4.5.1. Analýza příspěvků 37

4.5.2. Analýza komunikačních vláken 38

4.5.3. Chování uživatelů 38

4.5.4. Vztahy v komunitě 40

PRAKTICKÁ ČÁST **41**

Metoda výzkumu **42**
5.1. Google Colaboratory 42
5.2. Tři fáze výzkumu 43
5.2.1. První fáze 43
5.2.2. Druhá fáze 45
5.2.3. Třetí fáze 48

Výsledky **51**
6.1. Hodnocení příspěvků 51
6.1.1. Politické účty 51
6.2. Hodnocení trollů 55
6.2.1. Uživatelé, co trolly nejsou 56
6.2.2. Uživatelé, co možná trolly jsou 57
6.2.3. Uživatelé, co trolly jsou 59
6.3. Shrnutí 62

Závěr **65**

Příloha **67**

Bibliografie **68**

TEORETICKÁ ČÁST

Úvod

Rychlý vývoj nových komunikačních technologií se zasadil o změnu v každodenní mezilidské komunikaci. Zatímco dříve bylo setkávání lidí závislé na společném časoprostoru, dnes má na způsob komunikace radikální dopad vysoká rychlost přenosů informací. Tato změna proniká nejen do osobní, politické a ekonomické sféry života.

S nástupem nových médií, především pak internetu, se úměrně rozrůstá jednání, které může způsobit druhým psychickou újmu. Kromě skutečně vážných problémů, jakými mohou být zneužívání nezletilých nebo kyberšikana, se často setkáváme s dalšími nepříjemnými fenomény, které narušují ostatním uživatelům internetový prostor a činí ho nepříjemným. Jedním z takových potížistů může být tzv. internetový troll, jehož zákeřné aktivity v online prostoru často nejsou zcela očividné a proto jsou těžko rozpoznatelné.

Ačkoliv se povědomí veřejnosti o trollingu každým dnem rozšiřuje, rozhodně nelze tvrdit, že je každý uživatel internetu, zejména pak sociálních sítí, seznámen s úskalími, které s sebou trolling nese a není poté schopen na trolla adekvátně reagovat, setká-li se s ním. Konfrontace nejčastěji vyústí v tzv. *flaming*. Tímto termínem označujeme ostře vyhrocenou hádku mezi uživateli sociálních sítí, která často obsahuje vulgarismy, provokaci nebo argumentační fauly a často tak nesouvisí s původním tématem. Právě dosažení flamingu je nejčastější motivací a záměrem trolla.

Kvůli jisté podlosti trolla a jeho skrytým úmyslům, se bohužel stále dost dobře nedaří plně automatizovat detekovací systémy. To může být způsobeno neustálenou definicí termínu samotného. Je bezpečné ale tvrdit, že troll se v konverzačním vlákně nejlépe pozná na základě toho, *jak* reaguje, nikoliv na základě toho, *co* říká. Díky poznatkům umělé inteligence byly navrženy různé techniky, které se starají o automatickou detekci nejen trollů, ale i dalšího antisociálního chování v online prostředí.

V této diplomové práci se budeme zaměřovat na zkoumání prostředí českého Twitteru. Twitter je rozšířené sociální médium, ze kterého lze poměrně snadno

extrahovat různá data, tweety a odpovědi jednotlivých uživatelů. Práce v první části popisuje Twitter. Vysvětluje jeho strukturu, věnuje se metodám sběru dat a získávání informací z dostupných dat. Popisuje Twitter jako platformu pro sdílení aktualit a zpráv, což z něj činí platformu hojně využívanou politiky k profesním účelům. Zaměřuje se na českou politickou reprezentaci a na její aktivitu na Twitteru.

Právě politické prostředí je vhodným místem ke zkoumání fenoménu trollingu. Práce kriticky hodnotí tento druh chování v online prostředí a to z různých úhlů pohledu. Věnuje se popisu trollů a jejich vlastností, včetně výčtu a deskripce možných metod pro jejich detekci.

Diplomová práce si klade za cíl prozkoumat různé možnosti automatické detekce trollů především pomocí strojového zpracování přirozeného jazyka a síťové analýzy. Druhá část se proto věnuje praktickému zkoumání dostupných metod detekce, přičemž testuje především jejich účinnost. Tato práce netestuje software, který by byl schopný automatické detekce trollů, jak by z názvu mohlo vyplynout, nýbrž prakticky testuje popsané metody detekce, které by bylo vhodné zahrnout při budoucí konstrukci potenciálního softwaru pro automatickou detekci trollů na Twitteru.

1. Telekomunikační technologie

Před vývojem informačních technologií lidé neměli příliš způsobů komunikace na dálku. Kromě kouřových signálů později komunikovali klasickou poštou, ale do vynálezu telegrafu se dlouhá staletí v komunikaci nekonal žádný pokrok. Poměrně za krátkou dobu (od 19. století) se vývoj telekomunikačních technologií natolik zrychlil, že fatálně zasahuje do životů lidí a významně je ovlivňuje.

V 2. polovině 20. století byla spuštěna vůbec první počítačová síť ARPANET, která položila základy tomu, co dnes vnímáme jako internet. Poté byl v Ženevě představen World Wide Web (www), který měl usnadnit komunikaci mezi badateli propojením počítačů. Postupně společnost dospěla až k Webu 2.0, kde se uživatel přímo účastnil tvorby obsahu webu. V současné době hovoříme o Webu 3.0, jenž je charakteristický sdíleným obsahem. Internet tedy od samého začátku plní komunikační funkci, stejně jako sociální média, jejichž obsah téměř výlučně tvoří právě jejich uživatelé.

2. Sociální média

Termín “sociální média” pod sebou shlukuje online aplikace, které slouží především k soukromé i profesní komunikaci. Dříve se pro sociální média užívalo označení “sociální sítě”, od jeho užívání se upustilo převážně kvůli ambivalentnosti termínu. Termín sociální sítě není nový, jak by se mohlo minimálně dnešní generaci mladých lidí zdát. Sociální sítě lidé tvoří od nepaměti, nicméně význam termínu se později přesunul směrem k online prostředí. Zatímco dříve termín označoval nejrůznější druhy mezilidských kontaktů a vztahů, dnes si pod termínem představíme jednotlivé internetové stránky jako Facebook, Instagram, LinkedIn, Pinterest, Clubhouse, TikTok nebo Twitter.

Mohlo by se zdát, že využívat každé jednotlivé sociální médium z tak velkého množství je zbytečné, ale opak je pravdou. Všechna média, ať už byla na začátku jakákoliv, se postupem času změnila, aby nejlépe vyhovovala potřebám svých uživatelů. A byli to také uživatelé, kteří sociální média změnili k obrazu svému. Dnes je zakořeněno v povědomí veřejnosti, že naprosto každé sociální médium slouží jinému účelu, jinak se používá a sdružuje jinou komunitu lidí.

Důvodů, proč byla dříve sociální média označená termínem sociální sítě, je mnoho. Obecně vzato se mezi “starými” a “novými” sociálními sítěmi vyskytuje mnoho paralel, přičemž rozdílné využívání několika sociálních sítí patří mezi řadu z nich. Dalším uvedeným příkladem by mohl být jistý sociální kapitál, vlivné známosti, zkrácení dlouhých chvil nebo udržování kontaktů s přáteli. Lidé se pomocí sociálních sítí inspirují, motivují i edukují. Na druhou stranu rozdíl mezi “starými” a “novými” sociálními sítěmi nalezneme například v různé etiketě. Online etiketa obecně bývá nejen více uvolněná, než střet lidí v reálném životě, ale mnohdy v internetovém prostředí přehlížíme a odpouštíme takové chování, které bychom jindy považovali za jasnou známku hlubokého disrespektu.

2.1. Vývoj sociálních médií

Sociální média již přerostla svůj původní záměr, kterým bylo navazování kontaktů s přáteli a udržování styku mezi lidmi z různých částí světa. Nyní jsou slangově “socky” oblíbeným marketingovým nástrojem. Firmy si zakládají své vlastní

účty, nabízí své produkty a komunikují se zákazníky. To vše navíc s minimálními náklady a bezprostředně.

Nově pozorujeme vzrůstající trend náboru zaměstnanců skrze sociální média. Výběr budoucích zaměstnanců se pomalu přesouvá ze známých internetových portálů (jobs.cz) na osobnější platformy, které se stávají výhodnější alternativou. Firma je schopna ušetřit velké množství kapitálu a personalista je schopen z osobních profilů uchazečů vyčíst často více, než by dokázal při klasickém osobním pohovoru. Taková změna využití oproti původnímu účelu, za jakým byla sociální média vytvořena, je nejvíce markantní pro Facebook a Twitter.(Böhmová a Pavlíček, 2013).

Osobní profily se stávají doplňkem k životopisu. Dnes je již běžnou praxí, že si zaměstnavatel ověří internetovou stopu kandidáta o pracovní pozici a to napříč všemi věkovými skupinami. Představa výskytu pouze mladé generace v sociálních médiích se stala mylnou. Profily jsou profesní vizitkou člověka. Například Twitter je známým polem působení politiků, kteří skrze jednotlivé příspěvky komentují dění společnosti.

3. Twitter

3.1. Vznik a vývoj

Jak již bylo zmíněno výše, Twitter je jedním z nejstarších a nejpopulárnějších sociálních médií na trhu. Vznikl v roce 2006 v San Franciscu a jeho zakladatelem se stal Jack Dorsey. První verze Twitteru (Twitter I) měla sloužit lidem s městským životním stylem, kteří se skrze něj informovali o aktualitách a novinkách ze svého života. Dorsey na Flickru načtrtl platformu a označil ji jako Twttr, což byl krátký pětimístný kód, který by spolupracoval s SMS. Pro SMS zprávy bylo vyhrazeno 160 znaků, po jejichž překročení se zpráva rozdělila na dvě. Pro jednotlivě tweety, což jsou zprávy posílané skrze Twitter, bylo množství povolených znaků stejné. 20 znaků bylo rezervováno pro jméno uživatele a zbylých 140 pro samotný text zprávy.

Dorsey v rozhovoru pro *Los Angeles Times,* vydaném v roce 2009, definoval Twitter jako nové médium a veřejný komunikační systém, který stojí na psaní zpráv. Tento systém je nezávislý na zařízení a užívané platformě, což zajišťuje jeho

užitečnost pro širokou veřejnost. Kromě sdělování zpráv se Twitter začal užívat také jako nástroj pro sledování známých a přátel.

Zajímavý je původ názvu systému. Bylo žádoucí, ale název označoval a zachytil "*The physical sensation that you're buzzing your friend's pocket.",* tedy fyzický pocit bzučení v kamarádově kapse. *Twitter* označuje jednak ptačí zpěv (cvrlikání, štěbetání, švitoření), jednak "krátký nával informací, které na sebe nenavazují" (Weller a kol., 2014). Odtud je odvozeno také logo Twitteru, modrý ptáček, který má označovat sled na sobě nezávislých informací, tedy původní účel Twitteru I. Postupem času přichází posun k Twitteru II, jehož hlavním účelem není sledovat přátele, ale sledovat aktuality, důležité zprávy a události. (Weller a kol., 2014).

3.1.1. Twitter I

Dle různých studií se původní význam Twitteru, tedy nástroje pro sdílení informací mezi přáteli, ztratil. Tento význam vychází z původní etymologie názvu a jako další jediný spolehlivý zdroj můžeme brát slova zakladatele Jacka Dorseyho. V roce 2009 BBC uvedla, že dle jejich studie byla většina sdílených tweetů pouhé blábolení bez informační hodnoty. Tweety, které obsahují například informaci o tom, co jsem snědl k snídani, se během dvoutýdenního pozorování vyskytly ve 40% zkoumaných případů. Tweety s přidanou hodnotou tvořily naopak pouhých 9% příspěvků.

Twitter ostatně byl navržen právě proto, aby na něm lidé mohli chatovat o každodenních věcech. Právě jeho uživatelé ho začali postupem času měnit a formovat. Vymysleli například užití symbol @ při označování jiného uživatele v konverzačním vlákně, čímž vytvořili prostor pro dialog. Také uživatelé vymysleli užívání hashtagu #, který označuje klíčová slova v příspěvku nebo můžeme označovat (tagovat) jednotlivé tweety. Hashtagy spojují a zastřešují různé texty a to i od uživatelů, kteří se vzájemně nesledují. Poprvé byl použit hashtag #sandiegofire v roce 2007.

Díky tweetům o jídle a další nezávazné konverzaci se posunulo i vědecké zkoumání webu. Zatímco Web 1.0 byl čistě informativní, na Webu 2.0 můžeme zkoumat nejrůznější sociokulturní jevy. Ačkoliv se zdá konverzace o snídani nepodstatnou, lze sledovat, proč zrovna tuto informaci mají lidé potřebu sdílet a co

tím vyjadřují. Z výzkumů plyne, že lidé na Twitteru nekomunikovali za účelem dozvědět se nové informace, nýbrž protože vyhledávali lidský přístup, empatii a přišli si tak navzájem bližší. Samotný obsah sociálních médií nebyl důležitý, důležité bylo zůstat v kontaktu.

Twitter mohl být do značné míry označen za mělké sociální médium. Důležitou roli zde totiž hraje aktuální časoprostor a zvýrazňování oblíbeného. Příspěvky jsou řazeny v obráceném chronologickém pořadí a zeď se neustále obnovuje. Dříve bylo tweety možné na zdi (feedu) dohledat pouze po dobu 20 dní. Doba se následně snížila na 15 a poté na 7 dní. V minulosti nebylo na Twitteru možné vyhledávat, takže pokud se uživatel chtěl vrátit po časové ose, znamenalo to dlouhé scrollování. (Weller a kol., 2014)

Twitter I se od ostatních nově vznikajících médií lišil. Narozdíl od ostatních sociálních médií není nutné, aby se lidé vzájemně sledovali jako se například na Facebooku přidávají do přátel. Pokud tedy uživatele zajímá obsah jiného uživatele, může ho sledovat a tím získat přístup k jeho příspěvkům bez ohledu na to, zda ho druhý uživatel také sleduje či nikoliv. Místo jisté sociální bubliny proto uživatelé Twitteru budují spíše vlastní publikum, jelikož tendence vzájemného sledování je skutečně malá. Lidé na základě kvality svých příspěvků ztrácí nebo nabírají sledující (followery). Kvalitním, zábavným a pestrým obsahem pak mohou dominovat tomuto mikroblogovacímu světu.

3.1.2. Twitter II

V Listopadu 2009 se změnila otázka, kterou Twitter pokládal svým uživatelům z “*Co děláte?*” na “*Co se právě děje?*”. Tato změna charakterizovala jak vnitřní posun Twitteru, tak napověděla uživatelům i výzkumníkům, že mohou sdílet i jiné, než tzv. já-tweety, které tvořily do této doby 80% příspěvků. Jinak formulovaná našeptávací otázka ve statusovém okénku nabádá uživatele, že je možné platformu využívat pro oznamování novinek a zpráv nejen ze soukromého, ba naopak i z veřejného života. Kromě použití hashtagu ve spojení s požárem v San Diegu se na Twitteru lidé spojili také při sdílení postů o teroristických útocích, různých zemětřesení, nebo také při vytvoření kampaní (např. MeToo) a protestů (např. Volby v Íránu)

Z Twitteru se stává platforma, která mnohdy jako jediná informuje o událostech a tak nahrazuje klasická média při zprostředkování zpráv. (Weller a kol.,

2014). Kvalita a přesnost takových zpráv je negativně ovlivněna neprofesionalitou “novinářů”, kterými se zjevně stávají naprosto všichni uživatelé. Z místa, kde lidé bezpředmětně blábolili, se stává nouzový komunikační kanál (Weller a kol., 2014), který je využíván při velkých událostech, nejen politických kampaní, revolucích, atd. K této proměně dopomohlo taktéž zdvojnásobení povoleného počtu znaků v příspěvcích. Od roku 2017 se uživatelé nemusí nadále omezovat počtem 140 znaků a své informace mohou formulovat do příspěvků o 280 znacích.

Mimo to je nyní médium využíváno pro sběr a studii dat. Twitter má celou řadu dostupných analytických nástrojů. Je například možné zohlednit počet lajků, odpovědí a retweetů při určování důležitosti tweetů nebo pomocí hashtagů rozdělit tweety do skupin. Odpovědi konkrétním uživatelům poukazují na síť a vztah mezi sledovaným a sledujícím uživatelem.

Z předchozích odstavců vyplývá, že dnes se Twitter užívá různě. Je k dispozici na webu i jako mobilní aplikace, lidé ho využívají ke sběru dat nebo jako zpravodajskou relaci, ke kampaním, strategiím nebo jako formu deníku, kde sdílí s přáteli aktuality ze svého života. Jednotlivé tweety již nejsou omezeny počtem 140 znaků a uživatelé mohou sdílet i obrázky, videa a svoji GPS lokaci. Účty si zakládají nejen fyzické osoby, ale také firmy, skupiny nebo politické strany, které Twitter využívají k marketingové komunikaci. V roce 2020 bylo na Twitteru 186 milionu aktivních účtů, což je cca o 34% více než rok předchozí (2019). Zkoumat se tedy dá z velkého vzorku nejen komunikační dovednost uživatelů, ale také lingvistické, psychologické, politické a ekonomické fenomény společně s informatikou. Větší dataset informací lze získat pomocí *Application Programming Interface (API)*. (Weller a kol., 2014).

3.2. Struktura Twitteru

Původní myšlenka a struktura Twitteru byla velice jednoduchá. Postupem času se vnímání Twitteru proměnilo a s ním i jeho struktura. Lidé se na Twitteru začali seskupovat nejen z osobních důvodů, ale také kvůli sociálnímu a politickému dění. Pátrali zde po událostech a názorech na ně. Původní jednoduché nastavení této sociální sítě proto přestalo uživatelům stačit a začali vytvářet a používat další rozšíření. Některé nové prvky se uchytily, jiné ne. V momentě, kdy se nový prvek dostatečně uchytil a rozšířil mezi většinu uživatelů, ho vývojáři Twitteru zahrnuli do

kódu, přičemž Twitter záměrně umožnil uživatelům třetí strany přístup ke kódu a datům skrze API.

Intervence uživatelů dala vzniknout názoru, že za vytvořením současné podoby Twitteru stojí nejen jeho zakladatelé, ale i samotní uživatelé, kteří jako první implementovali @označení, retweetnutí nebo #hashtagy. Twitter není zdaleka jediná technologie, která se v průběhu času změnila od původního záměru vývojářů. Naopak, postup, kdy se technologie mění tak, aby odpovídala potřebám svých uživatelů, je běžný. Celý Web 2.0 a následně 3.0, stejně tak i tradiční telekomunikační média jako televize a rádio se také vyvíjely v závislosti na tom, co od nich lidé očekávali. (Weller a kol., 2014)

Je třeba podotknout, že s přívlastněním si některých funkcí ztrácí Twitter možnost variability a osobního přizpůsobení. Zahrnutí retweetu přímo do samotného kódu znamená, že různá označení, která byla dříve užívána (RT, via, by, MT) se sjednotí a mizí jistá míra personalizace uživatelského rozhraní. Pro vývojáře Twitteru není možné dostatečně uspokojit všechny své uživatele, pokud se jejich vidina platformy jedna od druhé liší.

3.2.1. Odpovědi

Po dlouhou dobu nebyla v mikrobologovacím světe Twitteru žádná možnost, jak rozvíjet konverzaci. Bylo sice možné na jednotlivé tweety odpovídat, ale komentáře nebyly žádným adekvátním způsobem uspořádány a proto v nich často nastával zmatek. Čtenář neměl šanci se vyznat v logické chronologii konverzačního vlákna. Uživatelé Twitteru si proto osvojili metodu, která fungovala v blogovacím světě a začali označovat komentář, na který reagovali. Tato inovace se zvláště ve svých začátcích nesetkala s kladnými ohlasy, někteří uživatelé si na zdaleka ne dokonalou novinku stěžovali. Twitter nicméně své rozhraní měnil často měnil. (Weller a kol., 2014)

K označení uživatelů se začal užívat symbol @, za kterým následuje dané uživatelské jméno. Označit uživatele je možné kdekoliv v tweetu a uživatel tak dává najevo, na koho mluví. Pokud se označení umístí na samý začátek tweetu, pak se jedná o odpověď. Odpovědi pomáhají uživatelům vyznat se v konverzačních vláknech. V roce 2009 začali uživatelé vidět odpovědi pouze těch lidí, které sledovali. Twitter tuto změnu obhájil tím, že uživatelé nemají zájem číst dlouhá

konverzační vlákna, přestože tímto krokem Twitter prakticky přišel o svou jedinečnou funkci spojovat různé lidi a umožnit svým uživatelům nalézt další inspirativní nebo zajímavé lidi. Uživatelé zanedlouho přišli na to, že jedná-li se o označení a nikoliv o odpověď, tzn. pokud je symbol @ s uživatelským jménem umístěn kdekoliv, jen ne na začátku tweetu, zmíněný problém nenastává. Stačilo proto před symbol @ umístit jakýkoliv znak kromě mezery. Nejméně rušivým elementem na začátku tweetu se ukázala být tečka. (Weller a kol., 2014)

3.2.2. Retweeting

Zatímco označení a odpovědi jsou prvky převzaté ze světa blogů a dalších konverzačních médií, samotné retweetování je unikátní pro Twitter. Retweetování vychází z předpokladu, že si uživatelé prostředí platformy vytvářejí sami. Jejich vnímání Twitteru je především závislé na uživatelích, které sledují. Aby se zpráva dostala do širšího povědomí, začali uživatelé zprávy někoho jiného přepisovat na svoje zdi. Tweety měly zpravidla podobu následující: začínaly RT (re-tweet), pokračovaly @označením původního autora zprávy a pokračovaly samotným textem přesdíleného tweetu. Někdy zkratku na začátku nahradilo MT (modified tweet), pokud uživatel původní tweet nějakým způsobem zkrátil nebo změnil. Pověšením retweetnuté zprávy na svojí zeď se k příspěvku dostali i uživatelé, kteří původního autora nesledují.

Twitter se poučil z nezdaru při implementování @odpovědí a při zahrnování prvku retweetingu postupoval opatrněji. Po kliknutí na šipky ve tvaru čtverce se původní tweet objeví na zdi uživatele přesně tak, jak ho původní autor napsal. Tím mizí původní forma retweetu. Tweet je jakoby zkopírován s původním jménem i profilovou fotografií autora a je pouze vložen na zeď někoho jiného, za předpokladu, že původní autor nemá soukromý profil. S retweetováním mizí také komentáře původního příspěvku, takže se tvoří nové komunikační vlákno. Tato funkce byla mnohými přivítána s nadšením, jiným se nelíbilo, že mizí zasazení tweetu do kontextu a retweetnutý příspěvek nelze uchránit před smazáním. Funkce je nicméně trvale součástí kódu Twitteru a je hojně využívána hlavně novými uživateli. (Weller a kol., 2014)

3.2.3. Hashtagy

Pro hashtagy užíváme symbolu #, za který se píše klíčové slovo nebo fráze celého příspěvku. Označení klíčových slov začalo dávat smysl ve chvíli, kdy začalo být možné na Twitteru vyhledávat, jelikož se uživatelé chtěli věnovat jistému tématu. Kromě klíčových slov mohou hashtagy označovat také ironii, pobídnutí ke konverzaci nebo okrajový komentář k tweetu. Když se Twitter v očích designerů měnil ze sociálního média na masové médium, změnil se i pohled na užívané hashtagy, které už neseskupují uživatele dle jejich zájmu a dle probíraného témata, ale označují trendy témata. Oproti jiným prvkům se hashtagy dlouhou dobu vyhýbaly procesu integrace do platformy. Nyní lze na hashtagy kliknout a tím otevřít databázi všech tweetů, které daný # použily. (Weller a kol., 2014)

3.3. Sběr dat z Twitteru

Kromě toho, že je to všestranná platforma užívaná ke komunikaci po celém světě, je Twitter také skvělým zdrojem různých informací. Výzkumníci různých oborů čerpají data pro svoje zkoumání právě z této oblíbené platformy. Může se jednat o jednoduché informace o určitém uživateli, ale také o rozsáhlé analýzy. Pro účely některých studií stačí vybrat pár uživatelů nebo malé komunity, pro jiné studie je zapotřebí sbírat komplexní data po dlouhou dobu.

Data se vzhledem k účelu jejich následujícího použití sbírají různými způsoby. Jedná se o sběr, analýzy a vizualizace, které jsou podloženy webovým rozhraním, nebo je možné těžit data přímo z Twitter API. Po získání dat je rovněž důležitá jejich následná správná interpretace.

Kromě pár výjimek se veškeré současné výzkumy provádějí pomocí těžení dat z API (*Application Programming Interface*) trojího typu: Streaming API, REST API a Search API, přičemž Search API je kombinací dvou předešlých. Na jaké bázi Twitter API funguje podrobně popisuje webová stránka https://developer.twitter.com/en/docs/twitter-api.

3.4. Získávání informací z dat Twitteru

Hledání relevantní informace z obrovského množství dat může být výzva. Za relevantní informace považujeme takové, které se umístí na předních příčkách po vyhledání a odpovídají na položenou otázku. Tradičně obrovské vyhledávače (Google) pracují s textovými informacemi dokumentů na základě několika požadavků na text. Například každý dokument obsahuje alespoň několik stovek slov nebo je obsah gramaticky správně napsaný. (Weller a kol., 2014)

Mikroblogovací svět Twitteru se přirozeně liší od velkých textových souborů. Vezmeme-li v potaz, že každý tweet nebo komentář může být relevantní při vyhledávání a že Twitteru je kladeno přes miliarda otázek denně, pak je naprosto jasně patrná potřeba nalézt metodu vyhledávání relevantní informace.

3.4.1. Odlišení Twitteru od ostatních zdrojů informací

Ačkoliv dobře zavedené metody vyhledávání dobře fungují jinde, na Twitteru jsou ty samé metody nepoužitelné. Svět Twitteru se liší třeba počtem znaků. Zatímco velké vyhledávače považují za relevantní texty takové, které obsahují několik stovek slov, každý tweet obsahuje maximálně 280 znaků. Uživatelské chování na sociálních sítích obecně se také mění. Vyhledávání lidí na Twitteru je značně rozdílné od etablovaného vyhledávání na Googlu. Mění se chování uživatele, postovací styl i vyhledávací styl.

Vzhledem k omezenému počtu znaků všech tweetů bez výjimky se na Twitter obrací mnohem více uživatelů, kteří hledají novinky a aktuální zprávy, než na jiné platformy založené na textu jako třeba osobní blogy. 85% všech tweetů je nějakým způsobem spojeno s novinkami, což z Twitteru činí významnou zpravodajskou relaci. Navzdory tomu je však Twitter zcela rozdílný oproti tradičním médiím, jelikož jeho uživatelé fungují jako sociální senzory, které poskytují nejaktuálnější informace z první ruky, čímž předávají informace o společenském dění dalším uživatelům.

Uživatelé volí různé taktiky při psaní příspěvků na Twitter tak, aby v 280 znacích byli schopni pokrýt vše, co mají na mysli, včetně emocí. Hojně se užívají zkratky a zkracování slov, občas uživatelé vynechávají samohlásky a v cizích jazycích vynechávají členy. Pro vyjádření emocí se užívají emoji nebo prodloužení slov opakovanou samohláskou. Hashtagy slouží jako klíčová slova, díky nimž se

člověk zapojí do konkrétní diskuze. Dále se užívají označování a odpovědi a při vkládání odkazů se zkracují URL (např. odstraněním domény z URL).

Již bylo zmíněno, že vyhledávání na Twitteru se liší od vyhledáváním na webu. Teevan a kol. v roce 2011 zjistili, že při vyhledávání informací na Twitteru uživatelé pracují s průměrně 1,64 slovy, zatímco pokud hledají na webu, průměr použitých slov se pohybuje kolem 3,08 slov. To může být zapříčiněno omezeným počtem znaků každého tweetu, což ústí v to, že se uživatelé snaží o co možná nejkratší klíčová slova. Uživatelé na Twitteru vyhledávají mnohem více ostatní uživatele než jak je tomu ve webovém prostředí. Z výzkumu vyplývá také to, že na Twitteru se velice často opakuje vyhledávání jednoho a toho samého tématu stejným uživatelem a to za účelem sledování novinek a nového vývoje situace, která ho zajímá.

3.4.2. Techniky získávání informací z dat Twitteru

Vzhledem k omezenému počtu znaků v příspěvcích a vzhledem ke speciálnímu stylu sdílení, není získávání informací možné pouze na základě zkoumání textu tweetu. Proto je na místě zabývat se těžením důležitých dat. K řešení problému se využívá například systém NER (*Named Entity Recognition*). Je to systém, který je schopen identifikovat vlastní jména, názvy nebo funkce. V ideálním případě dokáže vyhledat a propojit synonyma. Pokud by třeba český uživatel měl za cíl nalézt všechny tweety, které se týkají “českého prezidenta”, s pomocí NERu by se mu během jednoho hledání podařilo nalézt také tweety, které obsahují slova jako “Miloš Zeman” nebo “prezident ČR”.

Dalším způsobem, jak vytáhnout data z tweetů, je prozkoumání vloženého odkazu do textu tweetů. V tweetech se často objevují odkazy na URL stránky nejrůznějších zpravodajských a blogovacích služeb. Tyto tweety se ukázaly být jako nejrelevantnější v průzkumech, jelikož z jejich obsahu a obsahu přiložených textových dokumentů lze vytěžit nejvíce užitečných dat. (Tao a kol., 2012)

Kvalita obsahu jednotlivých tweetů se samozřejmě mění v závislosti na počtu relevantních informací v něm obsažených. Proto nestačí příspěvky označit za významné pouze na základě míry shody položené otázky s obsahem příspěvku. Přesnější je hodnotit informace pomocí kombinace několika faktorů. Vhodné je se zaměřit na podobnost mezi tweetem a dotazem, na kvalitu obsahu tweetu, na jazyk tweetu a na důvěryhodnost uživatele. Dostupné jsou různé modely zacházení s

tweety dle relevantnosti. Learning to rank (LTR) model například sestupně řadí tweety po uvážení výše zmíněných kritérií, jiné třídící procesy tweety neřadí za sebe, ale tvoří soubory relevantních tweetů. Cílem je zkrátka vytvořit užitečný a rozmanitý set prvků stejně jako přesně odhadnout důležitost parametrů. (Weller a kol., 2014)

3.5. Soudobá společnost na Twitteru

Studování a dokumentování soudobé společnosti získalo s výzkumem Twitteru nový rozměr. Zvláště vezmeme-li v potaz přeměnu Twitteru, kdy se ze sociální platformy, na kterou lidé sdíleli svoje životy, stala platforma na téměř zpravodajské úrovni. Zároveň Twitter využívají také společnosti jako marketingový nástroj a ke komunikaci se svými potenciálními zákazníky. Díky limitu 280 znaků na jeden příspěvek dochází k osekání informace pouze na její podstatu a zbavujeme se tak květnaté snahy o zaujetí. Další nespornou výhodou Twitteru je jeho dostupnost na všech mobilních zařízení, což podněcuje lidi k jeho bezprostřednímu užívání během celého dne. Uživatelé rychle reagují na události ve svém okolí, o kterých informují své sledující. ale také jsou interagují s ostatními uživateli přímo na platformě. Takové chování má za vedlejší účinek dostatek informací o soudobé společnosti, které jsou bohaté na detaily.

Skrze tweety mají výzkumníci skvělý vhled do toho, jak lidé komunikují jako jedinci, uvnitř sociální skupiny nebo dokonce uvnitř velkého společenství. Zatímco dříve byly sociální studie vždy napsány z pohledu jednotlivce, dnes máme díky Twitteru a jemu podobným platformám komplexní a detailní sbor informací o denním fungování mezilidské komunikace. To může dalším generacím pomoci při snaze pochopit systém, jakým se šířily informace, pochopit, jak se vyvíjely a šířily pocity a názory a nebo pochopit, jaký vliv na společnost a život mají online platformy.

Data dostupná na Twitteru je nutno z dlouhodobého hlediska extrahovat a chránit, aby je bylo možné zachovat a užívat po desítky let. V současné době se sice situace nevyvíjí směrem k zániku Twitteru, nicméně nic nezaručí, že Twitter, společně se všemi daty, v dlouhodobém horizontu nezanikne a tím společnost nenávratně přijde o všechny informace. Měli bychom se snažit o zachování obsahu, provedení a sociálního kontextu tweetu. O archivování tweetů se pokouší Knihovna Kongresu ve Washingtonu, nicméně užívání korpusu má své limity a právě velice důležitý sociální kontext není v korpusu zahrnut.

Sociální kontext tweetů naznačují jiné parametry, než právě jejich obsah, který je limitován na 280 znaků a proto v něm nenalezneme žádné doplňující informace nebo vysvětlení. To během pár let způsobí, že obsah tweetu nic nevypovídá. Je nutné brát v úvahu také autora, odpovědi a retweetování, což nám lépe dokreslí sociální status uživatele. Hashtagy určují téma tweetu a jeho zařazení. Pokud tweety obsahují odkaz na jinou URL, je vhodné pracovat s informacemi, které nalezneme na přiložené stránce. Uživatelé jsou aktivní a každý den na Twitter přispívají obrovským množstvím obsahu, proto je nutné k archivaci použít efektivní mechanismus. Při vyhledávání dat z korpusu je možné pracovat jednak s plným textem, který byl ke tweetu přiřazen a který ho lépe zasazuje do kontextu, jednak s meta-daty, která obsahují informace stažené přímo z Twitteru (uživatelské jméno, datum, lokace) a informace vytažené z obsahu (téma, události, pocity).

Užívání sociálních médii rapidně stoupá a proto se stávají obsáhlým zdrojem informací. Jejich struktura nabízí možnost dozvědět se fakta a názory stejně jako zjistit sociální vazby a interakce. Dobře vytěžená a vizualizovaná data ze sociálních vazeb a interakcí jsou zásadní pro studium fenoménu trollingu v prostředí Twitteru. Internetový troll negativně interaguje s ostatními uživateli a právě interakce z něj trolla činí.

3.6. Twitter a politika

Již bylo mnohokrát uvedeno, že z Twitteru se stalo příhodné místo k šíření novinek a událostí nejen z osobního života jeho uživatelů. Twitter je hojně využíván také k adresování různých politických událostí. Lidé zde sdílí své názory a postoje směrem k politikům, kteří mají možnost okamžitě reagovat a také vyjádřit svůj postoj. Twitter tak nabízí mnoho úhlů pohledu na politické debaty a díky hashtagům je možné tyto debaty strukturovat. Na Twitteru se přirozeně řeší stejná politická témata, která jsou odprezentována masovými médii, nicméně rozdělení pozornosti, která je daným tématům věnována, se podstatně liší. Uživatelé si sami zvolí, kterému tématu a jak dlouho se mu chtějí věnovat. Twitter není využíván pouze k oznámení politických pohnutek, uživatelé rychle politické dění interpretují a spojují ho s dalšími tématy. Lidé často diskutují politické názory uvnitř svého sociálního kruhu, čímž je ukotvují a pojí je se současným politickým děním, osobní zkušeností světovým pohledem. Vysvětlují, rozřazují a interpretují to, co právě slyšeli nebo viděli. Twitter

tudíž poskytuje unikátní informace a komentář daleko za hranicí samotné události a obohacuje tak tradiční média (Weller a kol., 2014).

V České republice se sociální sítě v politickém kontextu začaly řešit až s prvními prezidentskými volbami. Hovořilo se o potenciálu ovlivnit výsledky skrze digitální marketing a online kampaně. Nejedná se tedy pouze o neformální cestu, jakou mohou politikové hovořit s potenciálními voliči, ale o vědomou volbu politických stran (Macková, 2017).

3.6.1. Vývoj politické komunikace

Zájmem politické komunikace je vztah politiků a médií, přičemž komunikace skrz nová média se v posledních letech těší velké pozornosti výzkumníků a to zejména kvůli své naprosté odlišnosti od tradičních médií. Pro politické komunikace v online prostředí, které probíhají od konce 20. století, se běžně používají pojmy jako “postmoderní kampaně” nebo “digitální věk” (Macková, 2017).

V průběhu let se logicky politická komunikace měnila. S nástupem masových médií bylo nutné profesionalizovat kampaně a začali se najímat odborníci na vztahy s veřejností. Dochází k velké medializaci, se kterou souvisí i proměna žurnalistiky a dalších oborů. Online kampaně disponují řadou nesporných výhod. Hovoříme o přesnějším cílení na voliče, okamžité zpětné vazbě a lepší organizaci kampaní. Chování politiků samotných taktéž nezůstává stejné. Politici svoje chování, promluvy a časový harmonogram přizpůsobují médiím. Jelikož jsou média hlavním prostředníkem mezi nimi a voličem, snaží se je politici co možná nejlépe zaujmout. Zároveň je možné pozorovat snahu politiků ovládat mediální prostor. Největší důraz je přitom kladen na předvolební období. Prostor dostávají i dříve neslyšení politici, kteří zastupují menšinové názory nebo řeší okrajová témata a kterým dříve nebyl věnován mediální prostor.

Okamžitý přístup k obrovskému množství informací je pro dnešního voliče novinkou. Volič se na internetu sám rozhoduje, jakým informacím se chce věnovat, které vyhledávat, jaké znalosti prohlubovat a jakým tématům by se rád vyhnul a s jakým sdělením bude interagovat. Taková možnost značné samostatnosti a orientace v tématu se projevuje nedůvěrou lidí v mediálně podané informace nebo dokonce v politiky samotné. Jednoduché dualistické rozdělení občanů na politicky pasivní a aktivní již není možné. Zatímco v minulosti se angažovanost dala zřetelně zaznamenat dle činností občana, dnes to již měřitelné není. Valná většina občanů se

dnes nepovažuje za politicky aktivní, neangažují se v přímé politice, nicméně díky médiím zůstávají informovaní, z povzdálí pozorují politickou scénu, utvářejí si vlastní názor a do politiky vstupuji pouze ve chvílích, které sami uznají za vhodné. S politiky díky sociálním médiím mohou být v každodenním kontaktu nebo s nimi neinteragují vůbec. S proměnou mediálního světa tedy souvisí i proměna společnosti a z občanů se stávají kritičtí pozorovatelé.

3.6.2. Politická komunikace na internetu

Navzdory zásadnímu vlivu nových médií na politiku jako takovou, zůstával fenomén internetu poměrně dlouhou dobu mimo dosah hledáčků politických vědců. V posledních letech se situace mění a zájem o politickou komunikaci na internetu pozorujeme nejen u politologů, ale také u dalších humanitních oborů, kterými jsou psychologie nebo sociologie. Nejprve se odborníci zaměřili na dopad nových médií na politickou angažovanost voličů, později se začali věnovat také dopadu médií na podobu politických debat. (Macková, 2017).

S výzkumem nových médií vědci dochází k závěru, že ač mají média velký vliv, nejedná se o převratné změny. Dopady nejsou tak velké, jak se na počátku jevilo. Pozornost se proto přesouvá směrem k dílčím aspektům a zkoumají se zvlášť jednotlivé formy. Po výzkumu webových stránek se pozornost přesunula k blogům, následně ke streamovacím platofrmám (YouTube) a nakonec ke stále se rozvíjejícím sociálním médiím (Facebook, Twitter) (Macková, 2017).

Role sociálních médií začala sílit v roce 2008 s kampaní Baracka Obamy, který masivně využíval online platformy pro prakticky přímou komunikaci s voliči. Kampaň byla velice úspěšná svou inovací a proto se její unikátnost snažilo později napodobit více politiků, nicméně poměrně neúspěšně v porovnání s průlomem Baracka. Kampaň však zvedla zájem o výzkum dopadu sociálních médií na politickou komunikaci (Macková, 2017).

V České republice se již v roce 2009 začali mladí lidé sdružovat do skupin na Facebooku k vyjádření nesouhlasu s politikou Jiřího Paroubka. Sociální média postupně nabývala na důležitosti a začala se jim věnovat patřičná pozornost. V roce 2016 ČSSD dokonce důrazně doporučilo svým členům aktivitu na sociálních sítích, zejména na Twitteru. V Posledních letech jsou taková doporučení běžnou praxí i u ostatních politických stran. (Macková, 2017). Sociální média totiž jednak plní funkci plnohodnotné kampaně, která doufá naklonit si voliče na svojí stranu, jednak lze

pozorovat snahu vyplnit s pomocí sociálních médií vzniklou komunikační propast mezi politiky a demokratickými občany a to především přímou komunikací (Graham a kol., 2013).

Ač mohou sociální média být ku prospěchu komunikace mezi politickým aktérem a jeho voličem, je na místě zmínit další aspekty, které se mohou jevit jako negativní. V prvé řadě je politik nucen obeznámit se s fungováním a zacházením sociálních sítí, jsou-li pro něj novinkou. To je proces, který zabere čas. Pro mladší politiky je užívání sítí méně náročné a přirozenější, než pro ty starší. Proto jsou mladí politici zpravidla aktivnější nebo si užívání sítí osvojí dříve (Strandberg, 2013). Čas zabírá také samotné užívání a postování na sociální sítě. Je na místě upravit denní harmonogram podle aktivity ostatních uživatelů. Odpovídání na případné komentáře, zprávy a další ohlasy opět ukrajuje z pracovního času. Reakce na online statusy nejsou pouze pozitivní. Právě politická sféra je semeništěm trollů. Pokud politik nemá o existenci internetových škůdců valné informace, může mu jejich aktivita psychicky ublížit. Zároveň politik může při špatné komunikaci nejen s trolly ublížit sobě a své kampani. Proces zapojení se do online života je však nezbytný, pokud chce politik v současné společnosti předejít reputaci vzdálené persony, ke které chová nedůvěru, ba jí dokonce opovrhuje.

Aktivitou na sociálních sítích se politici přibližují dříve netykavé společnosti, znovu si získávají úctu a důvěru občana a navazují intimnější vztah, než jaký je možný skrze ostatní masová média. Na druhou stranu ukazují veřejnosti své soukromé životy a politická komunikace nabývá osobnější roviny. Politici nyní vystupují sami za sebe jako jednotlivci. Případné negativní ohlasy tím pádem také nabírají osobního rázu a kritika se může stát špatně stravitelnou. Obrovským zásahem do psychického zdraví se mohou stát příspěvky trollů, jsou-li v nevědomosti brány doslovně a vážně.

3.6.3. Obsah politické komunikace online

I když popularita sociálních médií stále roste, politická komunikace vedená skrze ně se zdá být stále konzervativní a jednosměrná. Slouží především ke sdílení informací a aktualizací a to především v předvolebním období. Ačkoliv jsou některé komunikace dlouhodobé, většinou intenzita kolísá. Političtí aktéři jsou nejaktivnější v komunikaci v období spuštění kampaní, které skrze sociální média představují (Larson, 2016). Kampaně vedené na sociálních médiích vnímají aktivnější politici

jako důležité, narozdíl od méně aktivních politiků, kteří nevěnují tomuto druhu kampaní žádný čas. Dále z výzkumů vyplývá, že ty kampaně, které jsou vedené aktivními politiky, jsou značně individualizované a vcelku vzdálené od centrální stranické kampaně. Politici se snaží protěžit hlavně sami sebe a zviditelnit svojí postavu (Karlsen, 2011).

Sociální média disponují nekonečným množstvím informací. Uživatelé si sami vybírají, jakým informacím jsou ochotni věnovat svůj čas. Twitter a ostatní sítě mají navíc potenciál dostat informace mezi velké množství lidí prostřednictvím organického sdílení, tj. sdílení, které se šíří “samo” bez další intervence politiků. Uživatelé jednoduše sdílí, respektive retweetnou, informace k sobě na profil, čímž se dostanou ke všem sledujícím, kteří mohou opět informace sdílet dál. Takové informace mají pro většinu uživatelů navíc větší váhu, protože se jedná o informace ze zdroje, ke kterému mají vztah a kterému věří. Sdílení zaručuje tok informací od uživatelů se silným informačním základem k těm méně informovaným (Boulianne, 2016).

3.6.4. Čeští politici na Twitteru

Do České republiky dorazil trend využívání sociálních médií pro politickou komunikaci později, než tomu bylo u ostatních západních států. Aktivních politických účtů existuje v poměru s ostatními zeměmi stále málo, i když mají vzrůstající tendenci. Twitterové účty navíc v České republice (zatím) nejsou natolik oblíbené, jako například facebookové účty. Politiky s větším počtem sledujících na Twitteru jsou například Karel Schwarzenberg nebo Andrej Babiš. Nejen kvůli podstatně většímu počtu potenciálních voličů preferují političtí aktéři komunikovat prostřednictvím Facebooku na místo Twitteru. Limitující je pro ně 280 povolených znaků a komunikaci vnímají jako jednostrannou. Taktéž tvrdí, že uživatelské prostředí Facebooku je přijatelnější a snazší na ovládání (Macková, 2017). Nicméně Twitter se i přes své limity stal nepostradatelným nástrojem, který politici užívají na denní bázi.

Výkonnou moc na Twitteru současný prezident Miloš Zeman nezastupuje. Oficiální účet má pouze mluvčí Jiří Ovčáček (@PREZIDENTmluvci) se 78,2 tisíci sledujících, nicméně stanovisko hradu se k němu staví neutrálně. Další složka výkonné moci, tj. vláda má svůj oficiální profil @strakovka, který sleduje 133,3 tisíce lidí. Kromě něj Twitter ke komunikaci využívají také jednotlivé osobnosti vlády, mezi

něž patří premiér Andrej Babiš s účtem @AndrejBabis (454,6 tisíc sledujících) a jednotliví ministři. Místopředsedkyně vlády a ministryně financí Alena Schillerová vystupuje pod účtem @alenaschillerov a má 38 tisíc sledujících. Ministr vnitra Jan Hamáček tweetuje pod účtem @jhamacek svým 87,9 tisícům sledujících. Ministr školství Robert Plaga zase využívá účtu @RobertPlaga a disponuje 15,2 tisíci sledujících. Znovu zvolený ministr zdravotnictví, Adam Vojtěch, vystupuje pod @adamvojtechano s 80,5 tisíci sledujícími. Ministr kultury Lubomír Zaorálek (@zaorálek) vystupuje na svém profilu a tweetuje pro 36,3 tisíc lidí.

Zákonodárnou moc tvoří Parlament České republiky, který je rozdělen na dvě komory. Poslanecká sněmovna má na Twitteru jeden zastřešující účet @snemovna s 20,3 tisíci sledujících, rovněž senát disponuje funkčním účtem (@SenatCZ) s 20,4 tisíci sledujících. Předseda poslanecké sněmovny Radek Vondráček vystupuje na účtu @Vondraczech pro pouhých 5 766 sledujících. Mezi aktivní předsedy různých politických stran patří například: Petr Fiala (předseda ODS) s účtem @P_Fiala a 90,8 sledujících, Markéta Pekarová Adamová (předsedkyně TOP 09) s účtem @market_a a se 73,4 tisíci sledujících, Ivan Bartoš (předseda Pirátů) s účtem @PiratIvanBartoš s 93,3 tisíci sledujících, Marian Jurečka (předseda KDÚ-ČSL) s účtem @MJurečka a téměř 30 tisíci sledujících a další. Za zmínku také stojí profil mladého bývalého poslance Dominika Feriho, který během covidové pandemie naprosto suverénně ovládl prostor sociálních médií, zejména Instagramu, který na Twitteru vystupuje pod @DominikFeri a sleduje ho 192,7 tisíc lidí. Pozornost je vhodné věnovat profilu bývalého ministra financí a bývalého předsedy TOP 09, Miroslava Kalouska. Kalousek vystupuje pod @kalousekm a jeho satirické tweety sleduje 193,3 tisíc lidí. Zajímavým uživatelem Twitteru je moderátor politických diskusí Václav Moravec, který není přímo politikem, ale v politické komunikaci České republiky hraje významnou roli. Na Twitteru vystupuje pod uživatelským jménem @vaclavmoravec s počtem sledujících 163,2 tisíc.

Nejen z počtu sledujících, ale také z počtu tweetů lze snadno vyčíst, které profily jsou aktivní a které méně. Nutno poznamenat, že počet sledujících nevyjadřuje pouze aktivitu a popularitu daného účtu, nýbrž může značit i jistou míru kontroverze. Zda je profil sledovaný a zda na jednotlivé tweety ostatní uživatelé reagují, je patrné z množství a obsahu komentářů, přičemž mezi komentující často nepatří pouze standardní uživatelé.

3.7. Typy uživatelů na Twitteru

Na Twitteru můžeme nalézt celou řadu různých uživatelů, kteří tuto platformu využívají dle svého. Kromě výše zmíněného užívání platformy k politické komunikaci je běžné všechny sociální média používat k marketingovým účelům, za žurnalistickým účelem, ke krizové komunikaci nebo k vědeckým účelům. Část z Twitterových účtů jsou softwarové programy, které se nazývají *boti*. Uživatelé, kteří platformy využívají pro osobní účely, se na Twitter chodí bavit, udržovat kontakt s přáteli nebo aby zůstali v obraze. Bohužel se najdou i tací, kteří záměrně v online prostředí škodí a znepříjemňují tak danou sociální síť jiným uživatelům. Jedná se o například o fenomén trollingu.

Pokud hovoříme o marketingové komunikaci, pak je na místě zmínit, že značky využívaly média ke svým kampaním vždy. S příchodem sociálních sítí a zejména pak Twitteru se však narušuje kontrola, kterou značky nad svou propagací dříve měly. Ruší se vztahy mezi majiteli značky, zákazníkem a konkurencí. S vidinou levného a dostupného marketingu většina značek udělala začátečnickou chybu, když se pokusily monetizovat rozrůstající se a stále oblíbenější prostor sociálních médií. Značky pouze přesunuly své již existující kampaně do online prostoru aniž by si uvědomily existenci dialogu, pro který jsou právě sociální sítě ojedinělé a tolik oblíbené. Ocitly se tak v prostoru, kde jsou snadno zranitelní, protože kdokoliv může přímo vyjádřit svůj negativní názor nebo zkušenost a to před zraky mnoha tisíců dalších uživatelů (Weller a kol., 2014). Taková konverzace mezi uživateli ale probíhá nezávisle na tom, jsou-li značky její součástí, či nikoliv (Thoring, 2011). Sociální média jako taková samozřejmě zaznamenala přítomnost značek na svých platformách a využila příležitosti. Brzy přišla s novou nabídkou firemních účtů, které se od těch osobních liší a jejichž rozhraní je přizpůsobené potřebám obchodníků.

Twitter je díky jednostranným vztahům mezi sledujícím a sledovaným vnímán spíše jako nástroj pro žurnalistické účely, kde uživatelé pracují se zprávami. Tradiční role novinářů, kterou zastávali v původních médiích, se s působením na Twitteru proměňuje. Novináři by nyní měli sloužit jako filtr, který je schopný oddělit fake news od seriózních zpráv a měli by užívat Twitter jako studnici inspirativních a nových informací. Žurnalisté využívají Twitter pěti způsoby. Promují na něm své webové stránky tím, že tweety obsahují odkaz a zviditelňují svou vlastní osobu. Dokáží zprostředkovat “živé vysílání” z místa konání určité události a v reálném čase o ní

tweetovat. Prostřednictvím Twitteru interagují s veřejností. Monitorují odezvu veřejnosti a podle toho upravují svou další komunikaci. Konečně s pomocí Twitteru pátrají a vykonávají výzkum. (Weller a kol, 2014)

Rychlý přístup k informacím a snadná dostupnost Twitteru a dalších sociálních médií z nich tvoří skvělý nástroj při krizové komunikaci. Zvláště v případě přírodních katastrof jako vlna tsunami, zemětřesení, záplavy, požáry a podobně. Uživatelé jsou již zvyklí vnímat Twitter jako platformu pro sdělování novinek, zpráv a aktualizací. Proto není divu, že takové zprávy jednak sami sdílejí a jednak se na Twitter uchylují potřebují-li nové informace zjistit. Hashtagy umožní označení tweetů, které se věnují svému tématu a takové tweety jsou dle hashtagů později snadno vyhledatelné. Twitter se stává výborným nástrojem nejen k dosažení informovanosti veřejnosti, ale také při snaze varovat občany.

Že je Twitter hojně využíván pro výzkum humanitních oborů je již zmíněno v předchozích kapitolách. Za zmínku stojí připomenout, že se data z Twitteru získávají pomocí API a dalších nástrojů, které jsou dostupné online.

Ačkoliv byl Twitter vyvinut výhradně pro mezilidskou komunikaci, mezi aktivními účty lze nalézt až 16% automatických a poloautomatických účtu, které jsou řízeny počítačovými softwary (Zhang a Paxon, 2011). Pro tento typ účtů je zaveden a běžně užíván termín *boti* (angl. *bots*). Boti automaticky přidávají obsah na svoje twitterové účty. Obsah může být předem definován vlastníkem účtu a automatické je pouze načasování sdílení, nebo může být obsah kompletně automaticky generován. Twitter podporuje užívání takových účtů, nejsou-li v rozporu s podmínkami užívání. Užitečných typů botů je několik. Jedná se o marketingové boty, které se užívají nejčastěji a jejichž obsahem je marketingová zpráva. Další užiteční boti jsou schopni sami tweetovat předpověď počasí, nebo varování o přírodních katastrofách. Dále existují boti pro zábavu, kteří si například s ostatními uživateli mohou zahrát interaktivní hry. Zakázané jsou takové automatické účty, které vykazují známky antisociálního chování. Mohou sdílet agresivní spamy, které jsou často využívány k nátlakovému marketingu. Mezi zakázané boty řadíme také falešné účty, díky kterým uživatel získá klamavý počet sledujících, což následně může ovlivnit jeho vystupování na internetu.

Většina uživatelů využívá Twitter k soukromým účelům. Již bylo předesláno, že lidé na sociálních médiích udržují kontakt s přáteli, se kterými by v opačném případě kontakt ztratili, využívají Twitter ke své informovanosti, sbírají inspiraci a

motivaci, vzdělávají se nebo si zkrátka krátí dlouhé chvíle. Pro lidi, kteří svůj volný čas věnují popichování ostatních uživatelů, se vžil termín internetových *trollů* (angl. *trolls*). Trollové jsou charakteričtí svým zájmem o podnícení ostré internetové debaty, která se často promění v nesmyslnou hádku plnou urážek a zesměšňování mezi jednotlivými uživateli.

4. Troll

V každodenním životě se běžně setkáváme se spoustou záškodníků, jejichž chování si neumíme vysvětlit a kroutíme nad ním hlavou. Internetové prostředí není výjimkou. I zde nalezneme lidi, kteří pro své potěšení kazí ostatním zážitek z užívání online prostředí. K pocitu nedosažitelnosti jim napomáhá i anonymita, kterou internet nabízí. Uživatelé sociálních sítí pak mají pocit, že za své činy nejsou odpovědní.

Kvůli krátkým větám a nedostatku neverbální komunikace snadno dochází k neúplnému porozumění mezi oběma stranami. Mnohdy je velice obtížné s naprostou jistotou detekovat internetového trolla nebo různé spamující boty.

Termín “troll” v rámci World Wide Webu byl poprvé užit na Usenetu, fóru, které bylo populární v osmdesátých letech (Fornaciari a kol., 2018). Ačkoliv sociologie popisuje každé antisociální chování termínem “troll”, jeho samotná definice je kluzká. Fenomén trollingu je do značné míry subjektivní záležitostí a je proto těžké najít jeho přesný popis. To ústí v nepochopení problematiky trollingu a vysvětluje malý zájem badatelské komunity o jeho zkoumání. Obecně termínem označujeme člověka, který vykazuje známky antisociálního chování v internetovém prostředí, nejčastěji na sociálních platformách. Ve své komunikaci využívá agrese a jeho jazyk je urážlivý. Svým jednáním provokuje a irituje ostatní uživatele. Jeho cílem bývá odvedení směru diskuze od probíraného tématu a často diskuzi zcela ukončí. (Tomaiuolo a kol., 2020) Trolling je tedy specifický typ zákeřného online jednání, které míní narušit interakce, rozčílit interagující partnery a odlákat je od přínosné argumentace. (Samory a Peserico, 2017).

Troll je specifický tím, že cíl jeho jednání není na první pohled zcela patrný. To je také jeden z důvodů, proč je jeho automatická detekce více problematická, než je tomu u automaticky řízených počítačových softwarů. Boti jsou poměrně snadno identifikovatelní vzhledem k faktu, že sdílejí rozpoznávací prvky na svých účtech. Oproti tomu trollové jsou účty skutečných lidí, kteří mají víceméně společné

charakteristické vlastnosti jednání s dalšími uživateli, ale člověk od člověka své chování demonstruje rozdílným způsobem. Dalším důvodem, proč se zatím zcela nedaří zautomatizovat detekci trollů je rozdíl v definici, kterou pro trolla užívají výzkumníci a kterou pro trolla užívají uživatelé sociálních médií. Taktéž admini, kteří mají za úkol usměrňovat tok diskuze, často selžou v detekování onoho manipulativního a klamavého jednání. Trollové jednají skrytě, často pouhým dobře mířeným popíchnutím spustí válku v komentářích.. Využívají naivity nezkušených uživatelů a neveřejné vedlejší komunikace. Odhalit trolling není jednorázová aktivita, proto je uživatel pod dohledem dlouhou dobu a k jeho usvědčení je třeba nasbírat mnoho nevyvratitelných argumentů a neprůstřelných důkazů. Jednotlivá komunikační vlákna, která skýtají trolla, není těžké detekovat, nicméně vystopovat jedince, který flaming započal, je náročné. (Samory a Peserico, 2017).

Trollové jsou lidé, kteří se vyžívají v obtěžování ostatních, vykazují známky sadismu, znovu prožívají svá traumata nebo trpí depresí (Buckels, 2014). Dalším typickým znakem jsou nefungující vztahy. Trollové často nezapadají do škatulek společnosti, nejsou si svou pozicí ve společnosti jistí, nebo žádnou nemají. Problémy reálného života promítají do online světa, kde si vybíjejí nahromaděnou zlost a s mottem “když nejsem šťastný já, nebudou ani ostatní” znepříjemňují sociální média psychicky stabilním a zdravým uživatelům.

4.1. Odlišení trolla od ostatních uživatelů

Navzdory špatné detekovatelnosti trollů, sdílí tito uživatelé jisté rysy, které je oddělují od ostatních běžných uživatelů. Trollové přispívají na konkrétní platformu pravidelně a nepřerušovaně v řádu několika měsíců či let. Jejich aktivita je frekventovanější a posty častější. Kvalita textu se taktéž liší. Trollové píšou méně čitelné texty s menším počtem znaků v porovnání s běžnými uživateli a upřednostňují tak kvantitu před kvalitou. Taktéž se snaží o větší koherentnost jednotlivých příspěvků za účelem získání si důvěry většinových uživatelů (Samory a Peserico, 2017). Samory a Peserico (2017) tvrdí, že trollové narozdíl od násilných uživatelů příliš nepoužívají vulgární termíny, potupné urážky nebo klení na denní bázi, že se vyznačují spíše chybějící empatií a snaží se o konfrontační komunikaci. Tomaiuolo (2020) naopak definuje trolla jako internetového uživatele, který jedná agresivně, nebojí se užít vulgárních výrazů, touží po konfliktu, s druhými jedná

provokativním tónem a nezdráhá se je urazit. Na druhou stranu při definování trollího jednání Tomaiulo uvádí, že troll se snaží o degradaci konverzace, ke které nutně nedochází ponížením jedince. Namísto toho jde spíše o chování, které zahrnuje posílání opakujících se, nesmyslných a irelevantních zpráv, které mají za cíl na sebe strhnout pozornost (Tomaiuolo a kol., 2020).

Mimo touhy škodit ostatním je troll motivován také úspěchem, který později sklízí. Může být ostatními uživateli oceněn za vtipnost a to nehledě na to, zda jsou jeho pravé záměry ostatními uživateli odhaleny, či nikoliv. Ať už se ostatní účastníci internetové komunikace chytí do pasti, kterou troll připravil nebo se snaží jeho záměry odhalit, stává se troll centrem pozornosti, po které touží (Tomaiuolo a kol., 2020).

Samory a Peserico (2017) poukazují na rozdílné užívání zájmen v porovnání s běžným i s násilným uživatelem. Tvrdí, že trollové se vyhýbají zájmenům první a třetí osoby čísla jednotného, které jsou typické pro násilné uživatele. Místo toho je možné pozorovat snahu o navázání kontaktu pomocí zájmen druhé osoby čísla jednotného a první osoby čísla množného a doufají v navázání vztahu s komunitou. Komunikují stejným způsobem, jako průměrný běžný uživatel, nicméně kvůli tomu, že se vyhýbají osobním tématům, se jim nedaří plně dosáhnout empatického propojení. Interakce trollů v komunikačním vlákně spočívá v taktickém vyčkávání a reagování na příhodný příspěvek. Snaží se zůstat nenápadní. Někdy dokonce natolik, že není možné s jistotou určit, zda se v daném příspěvku fenomén trollingu objevuje. Trollové sami nejsou iniciátory diskutovaných příspěvků, ale vybírají si napjaté konverzace a dychtivě se snaží stát se jejich součástí. Z toho vyplývá, že trollové nejsou zřetelně ofenzivní antisociální uživatelé, nýbrž se snaží zapadnout, třebaže vlivem nedostatku empatie působí chladně a drsně. (Samory a Peserico, 2017)

Nutno podotknout, že vlastnosti zmíněné v předchozím odstavci nejsou v českém prostředí typické pro jednání trolla. Není jasné, zda nejasnosti v definici trolla jsou způsobeny rozdílnou kulturou a způsobem komunikace jednotlivých zemí, nebo každý označuje trollem jiné chování na internetu.Tento problém byl ostatně diskutován již v předchozích kapitolách. Detekci trollů se nezdařilo doposud zcela úspěšně zautomatizovat i kvůli jeho rozdílné definici. V České republice je trollem chápán agresivní a vulgární uživatel, který záměrně ničí internetový prostor. Troll provokuje, uráží a jeho příspěvky do diskuze postrádají logickou strukturu a nejsou

obsahově úplné. Za trolla zcela jistě neoznačujeme slušně se vyjadřujícího člověka, kterému by chyběla dávka empatie a který by se chtěl pouze začlenit do kolektivu. Upřímně si takového člověka v komunikačním vlákně většina uživatelů ani nevšimne, jelikož ničím nevyčnívá a tak mu není věnována pozornost. Pokud je motivací trolla na sebe strhnou pozornost valné většiny, chová se velmi výstředně, jelikož pak zcela jistě nezůstane bez povšimnutí. Vykazovat známky antisociálního chování chce odvahu. Ti, kteří by rádi svým negativním komentářem do diskuze přispěli, ale pod svou pravou identitou si vystoupit netroufnou, si vytvářejí anonymní falešné účty. Anonymita nabízená internetem smyje poslední stopy váhavosti a troll začne neohroženě přispívat do všech vláken, která mu k jeho účelům připadnou vhodná.

4.2. Komunikační vlákna s trollem

V předchozí kapitole je definováno, jakým způsobem se trollové vměšují do konverzací a jaký jazyk ke komunikaci používají. Jejich jazyk není nutně urážlivý, nýbrž podněcuje emoční reakci ostatních uživatelů. Komunikují komplexně, což se vylučuje s klamavým stylem jednání. Trollové typicky nelžou a nesnaží se někoho podvést, ale v průběhu konverzačního vlákna jsou schopni změnit svůj postoj nebo hrát usmíření. Reagují na emoční rozpoložení ostatních uživatelů a dle toho řídí tok svých příspěvků, čímž maskují svou strategii (Samory a Peserico, 2017).

Ačkoliv trollové do konverzačních vláken přispívají s cílem znehodnotit probíhající dialog, typicky to nejsou právě oni, kdo založí oheň. Jejich taktikou zůstává trpělivě přihlížet a ve vhodnou chvíli oheň rozdmýchat. Při analyzování konverzačních vláken dochází ke zjištění, že příspěvky, které následují bezprostředně po příspěvku, ve kterém bylo možné detekovat fenomén trollingu, se v podstatě nemění. V těchto příspěvcích nepozorujeme náhlou změnu tónu, urážky nebo hanlivá oslovení. Na druhou stranu příspěvky zahrnují více inkluzivního jazyka, zkracují se a část z nich tvoří citování jiného příspěvku. To ústí v cyklickou a bezpředmětnou konverzaci, jež může být skvělým indikátorem přítomnosti trolla (Samory a Peserico, 2017). Pokud konverzace stále probíhá, je možné si všimnout, že se se vzadlováním se od trollího příspěvku uklidňuje a opět se užívají konverzační pravidla běžná pro slušnou komunikaci. Tento fakt opět směřuje k

nelehké detekovatelnosti trolla ve vlákně, jelikož se jeho příspěvek schová mezi svými neutrálními sousedy (Samory a Peserico, 2017).

Detekování vlákna, kde je přítomný fenomén trollingu je snadné. Lze je lehce oddělit od vláken, kde trollové nepůsobí. Co ovšem dělá problém, je detekovat právě ten příspěvek, který lze považovat za iniciátorký. Vytipovat si při observaci daného trolla a dále zkoumat jeho pozdější chování v online prostředí je jednou z variant, jak dosáhnout úspěchu.

4.3. Napáchané škody

Fenomén trollingu přesahuje problém provokativních útoků a páchá škody na zranitelných uživatelích internetu. Hardaker (2010) přichází s akademickou definicí trolla jako internetového uživatele, který se pouze dobře baví pod uměle vytvořenou záminkou zapadnout do společnosti a být součástí konverzace. Stejně jako trolling dokáže být neslušný a podlý, může vyvolávat i další pocity. Troll je frustrovaný, pokud uživatelé správně odhadnou jeho záměry, nejsou jeho chováním vyprovokováni a neodpovídají. Trolling může být zmařen, pokud ho uživatelé odhalí a reagují, nicméně reagují takovým způsobem, který trolla odhalí. Troll také neuspěje, pokud si ho žádný uživatel nevšimne a tak opět dostane nulovou reakci. Trolling je úspěšný pouze v případě, že jeho záměry ostatní uživatelé neodhalí a reagují přesně podle přání tohoto antisociálního uživatele (Hardaker, 2010).

Pro zranitelné a naivní uživatele může být setkání s trollem devastující pro jeho psychické zdraví. Citliví lidé si berou k srdci útoky, ponížení nebo urážky, které troll pro své pobavení sdílí v komunikačním vlákně. Vzhledem k asynchronní povaze internetu se nejedná pouze o časový diskomfort zraněného uživatele, negativní následky si může nést i v dlouhodobém časovém horizontu. Mohou hanobením hraničit s kyberšikanou, nebo skončit fatálně. Navzdory faktu, že se adolescenti opravdu těžko vyrovnávají s krutostí online světa, je těžké vymáhat spravedlnost. Trollové se často hájí svobodou názoru a právem svůj názor vyjadřovat. Proto navzdory tomu, že mohou skutečně vážně uškodit komunitě, jsou správci sociálních médií s jejich vyhošťováním opatrní. Jednak není lehké trolla bezpečně rozpoznat a jednak se takové vyhoštění může považovat za cenzuru. Odstraněním trolla se taktéž naruší přirozený běh konverzace (Fornaciari a kol., 2018).

Mluvíme-li o trollingu, je namístě také zmínit, že se nejedná pouze o jednotlivce, kteří stojí za těmito útoky. Trollové se seskupují a hromadně operují v online diskusích, čímž mají značný vliv na změnu názoru širokého okolí. Organizace velice účinně šíří pomluvy, fake news nebo dezinformace. Ovlivnit tak mohou široké spektrum lidí jak v pozitivním, tak negativním slova smyslu. Vmanipulovat uživatele do koupi báječných hrnců nebo ovlivnit jeho politické preference je na dosah ruky každého většího seskupení lidí. Rusko například bylo obviněno z ovlivnění voleb amerického prezidenta v roce 2016. S pomocí lidí i botů mělo řídit masovou manipulaci názorů. IRA (*Internet Research Agency*), přezdívaná jako farma trollů, je jednou z prvních odhalených organizací, která systematicky využívá trollingu pro politickou propagandu (Tomaiuolo a kol., 2020).

Ať už jsou původní intence trolla jakékoliv, měli bychom k jeho počinům přistupovat s notnou dávkou opatrnosti, jelikož následky mohou být nedozírné.

4.4. Jak se s trollem vypořádat?

Pokud se uživatelům skutečně podaří správně identifikovat trolla, což jsme dokázali, že je nelehkým úkolem, měli by se řídit jednoduchým pravidlem. Základní učebnicová poučka zní: “Nekrmte trolly!”. V podstatě znamená, že nejlepším způsobem, jak se s trollem vypořádat je nedopřát mu luxus naší pozornosti a vyhnout se jakémukoliv kontaktu. Obvykle diskuzní fóra jmenují tzv. moderátory, kteří řídí tok diskuse a mají právo vyhostit nebo zablokovat uživatele, které vyhodnotí jako trolly. Na sociálních sítích mají takovou pravomoc adminové, čili správci. Takový systém může fungovat pouze u malých nebo středních skupin uživatelů. Ve velkém měřítku je takřka nemožné detekovat fenomén trollingu ručně. Problém se částečně řeší možností nahlašování buď samotných trollů, nebo trollího chování. Každý uživatel může každý příspěvek nahlásit za nevhodnost. Dosáhne-li příspěvek daného počtu nahlášení, je automaticky smazán. Pokud se jedná o omyl a příspěvek byl pouze kontroverzní odpovědí, má autor příspěvku možnost odvolání, které následně posoudí správci dané sítě.

Takto decentralizovaný přístup navíc řeší problém subjektivity, jelikož není závislý na jedné jediné kontrolující autoritě. S takovým systémem obstarávání komunity a ochranou před trolly se pojí také zpomalení procesu. Všichni uživatelé nemohou reagovat ve stejný moment a bez časové prodlevy. Kvůli této prodlevě

však nejsou včas chráněni běžní uživatelé, kteří jsou tak nuceni se s trollingem setkat a tak jim může stále ublížit (Tomaiuolo a kol., 2020).

Nejlepším řešením se proto zdá být implementace algoritmu, který je schopný trolly automaticky detekovat. V minulosti se užívaly takové systémy, které detekovaly tzv. fake profily, tedy falešné. Detekce falešných profilů je poměrně snadná, jelikož se jedná o účty se shodnou IP adresou s dalším účtem, vyznačují se také velkým počtem zamítnutých žádostí a přátelství/sledování a často nedisponují sítí kontaktů ani profilovým obrázkem. Pod rouškou anonymity falešných profilů lze zaznamenat antisociální chování poměrně často. Uživatelé falešných profilů se s novou identitou již nezdráhají trollit, jako by to snad ostatním uživatelům škodilo méně. V poslední době se detekci trollingu věnuje obor umělé inteligence a proces těžby dat. Odborníci pracují s různými druhy analýzy dat (Tomaiuolo a kol., 2020).

4.5. Detekovací metody

Proč je snaha o úspěšnou detekci trollů na místě bylo podrobněji vysvětleno v předchozí kapitole. Nejen, že působení trollů může mít vliv na veřejné mínění a změnu názoru uživatelů, ale může také fatálně ovlivnit psychické zdraví. Badatelé proto přirozeně hledají cesty nejúčinnější detekce.

Nejprve se začalo pracovat s automatickým prozkoumáním online obsahu skrze strojové zpracování přirozeného jazyka, tzv. NLP (*Natural Language Processing*). NLP analyzuje obsah a z nalezeného počtu negativních slov vyvodí celkové naladění příspěvku. Později se přidaly i více sofistikované metody analýzy, kterou je analýza postoje (angl. *sentiment analysis*). Podstata postojové analýzy spočívá v přiřazení emoce slovům ve větě, z čehož následně vyhodnocuje postoj napsaného příspěvku. Mezi nežádoucí emoce, které výsledně vedou k označení uživatele za trolla je hněv, zlost, zuřivost a další. Při detekci se také využívá síťová analýza SNA (angl. *Social Network Analysis*), která spočívá v prohledání a analýze dat uživatelů v rámci jedné komunity. (Tomaiuolo a kol., 2020)

Nejvhodnějším řešením složité analýzy obsahu a následné detekce trollů je kombinace různých metod analýzy. Je možné kombinovat analýzu stylu psaní, postoje, chování, sociálních interakcí, přiložených médií a času sdílení (Tomaiuolo a kol., 2020). Přístup na základě strojového učení je velice účinným při detekci.

Metody detekce trollů lze roztřídit dle typu informace, kterou pro detekci užívají. Analyzují se příspěvky uživatelů, celá komunikační vlákna, chování uživatelů a vztahy v komunitě.

4.5.1. Analýza příspěvků

Uživatelé se na sociálních sítích vyjadřují formou příspěvků ve psané formě, proto je vhodné předpokládat, že taktéž troll sdílí své myšlenky pomocí textu. Z toho vyplývá, že analyzovat informace obsažené v textu příspěvku bude správný přístup k odhalení antisociálního chování. K analýze dochází například s využitím a následným vyhodnocením ARI (*Automated Readability Index*), jelikož trollové komunikují jinak v porovnání s běžným uživatelem (viz. 4.1.) Badatelé textový obsah zkoumají s využitím nástrojů postojové analýzy (Tomaiuolo a kol., 2020).

Younus a kol. v roce 2014 provedli studii politické debaty na Twitteru pomocí analýzy postoje tweetů s využitím dvou algoritmů. Ke své studii vybrali uživatele, kteří se pokaždé umístili na spodních příčkách a snažili se zjistit, zda-li jsou daní uživatelé skutečnými trolly. Došli k závěru, že mnoho z nich jsou běžní uživatelé, kteří mají na daný politický diskurz krajně nepopulární názor, který navíc ofenzivně hájí. (Younus a kol., 2014)

Angiani a kol. (2016) ve své anylýze emocí na Twitteru rozdělují tweety do třech levelů. Určují, zda jsou tweety objektivní nebo subjektivní, pozitivní nebo negativní v rámci subjektivních a zda v rámci pozitivních vyjadřují lásku, radost nebo překvapení a zda se v rámci negativních jedná o strach, zlost nebo smutek (Angiani a kol., 2016).

Nelze proto soudit, zda je postojová analýza spolehlivým ukazatelem trollingu na Twitteru a to navzdory bohaté Twitter API, nýbrž je vhodné do analýzy příspěvků zahrnout další poznatky chování trollů, které jsou známé. Hovoříme o faktu, že trollové píšou krátké příspěvky a často, čímž upřednostňují kvantitu před kvalitou. Pozornost je na místě věnovat autorovi, klíčovým slovům, datu příspěvku, délce příspěvku a podobně. Ačkoliv samotná analýza metadat získaných z tweetů je méně přesná, než samotná postojová analýza, kombinace obou se zdá být vhodným řešením.

4.5.2. Analýza komunikačních vláken

Vzhledem k nepřesnostem a limitům při separátní analýze jednotlivých příspěvků je přesun k analýze celých komunikačních vláken logický. Takový přístup využívá stejné metody jako analýza příspěvků a doplňuje je o další informace, které získává z kontextu komunikace. Roli v detekci fenoménu trollingu hrají citace ostatních uživatelů v příspěvcích.

De-La-Peña-Sordo a kol. (2013) ve své studii poukazuje na možnost charakterizovat příspěvky podle následujících kritérií: nesmyslnost obsahu, stupeň kontroverze a zda-li se drží původního tématu. Každá zpráva je tak přiřazena k hodnotám, které přesně identifikují tyto tři charakteristiky. Tato měření jsou aplikována na všechny přispívající v komunikační vlákně a určují míru nepřátelského chování. Extrahovaná data jsou následně seskupena a porovnávají chování mezi uživateli. Na základě výsledků je pak možné identifikovat fenomén trollingu.

4.5.3. Chování uživatelů

Analýza celkového chování uživatele a vztahů v komunitě je dalším metodou při detekci trollů. Tento přístup se neomezuje pouze na sdílené příspěvky či komunikační vlákna a na informace z nich získané, ale zaměřuje se na celkové chování uživatele v dané komunitě, na jehož základě je schopný dedukovat, zda si uživatel počíná antisociálně. Uživatelé, kteří se vyžívají v trollingu, se do sítě zaregistrují a jejich chování v průběhu užívání se zásadně neliší. Zákeřné chování je z jeho profilu patrné, jelikož se jedná o soustavný vzorec, nikoliv o ojedinělou reakci na určité téma. Při analýze dat se badatelé nesoustředí na jednotlivé obsahy příspěvků, ale snaží se trolla odhalit na základě jeho pozorování jako celku dostatečně brzy na to, aby minimalizovali škody nebo zabránili jeho dalším aktivitám. Metoda zkoumání chování uživatelů kombinuje všechny přístupy uvedené v předchozích kapitolách (4.5.1.,4.5.2.), tzn. pracuje s analýzou postoje i s analýzou kontextu a přidává k nim obecnou analýzu chování (Tomaiuolo a kol., 2020). K závěru, že dosud užívané metody pravděpodobně nebudou dostačující, došli výzkumníci pomocí závěrů, které z prací vycházeli. K efektivní detekci trollů je zkrátka nutné využít analýzu postoje společně s dalšími metodami.

Aby byla badatelská komunita schopna zahrnout do analýzy specifikace pro odhaleni fenoménu trollingu, bylo nutné nejdříve zjistit, co přesně trolla od ostatních uživatel odlišuje. Cheng a kol. v roce 2015 vypracovali studii, která se soustředí na

charakteristiky antisociálního chování v online diskusích, na charakteristiky podlých uživatelů, vývoj fenoménu trollingu a na popis vyvolané reakce u ostatních. Spoléhají na samotnou komunitu a na moderátory diskuzí, díky nimž rozdělí členy komunity na ty, co vykazují známky podlého chování a na ty, co známky nevykazují. Toto rozdělení vychází z počínání moderátorů nebo adminů, kteří nahlašují nevhodné příspěvky. Následně porovnávají obě kategorie uživatelů (nahlášení a nenahlášení) a zjišťují konkrétní prvky, které jsou typické pro jednání trollů. Jazyk užívaný oběma skupinami se výrazně liší. Dochází k závěru, že trollové píší komentáře, které jsou nepochopitelné, nedrží se tématu a užívají více irelevantních slov. Dalším charakteristickým znakem je nezájem trollů o začlenění se do více komunikačních vláken, tzn, často reagují pouze v jednom komunikačním vlákně, ve kterém se ovšem angažují o poznání více. Píší více komentářů s čímž souvisí fakt, že se jim dostává také více odpovědí od ostatních. Při analýze profilu uživatele, který vykazuje známky antisociálního chování, dochází ke zjištění, že příspěvky trolla jsou často blokovány moderátory a že číslo blokovaných příspěvků v čase roste, zatímco u běžných uživatelů zůstává relativně stejné. Nálada a diskuzní kontext jedince vysvětlují fenomén trollingu často lépe, než analýza historie jedince (Cheng a kol., 2015)

S využitím takto podrobné analýzy charakteristických prvků je možné sestavit metodu vhodnou pro detekci trollů. Metoda stojí na analýze prvních deseti příspěvků uživatele a ty podrobuje zkoumání. Všímá si čtyřech prvků. Prvním je obsah příspěvku. Hodnotí se počet slov, čitelnost, obsah, apod.. Jako druhá se zkoumá aktivita uživatele, tzn. denní počet sdílených příspěvků, počet odpovědí v komunikačním vlákně, celkový počet příspěvků, apod. Třetím prvkem je reakce komunity a čtvrtým zkoumaným faktorem jsou činy moderátorů vzhledem k danému uživateli. Ve čtvrtém kroku zkoumáme, zda-li došlo k blokování komentářů, kolik komentářů bylo odstraněno, apod (Kumar a kol., 2014). Metoda správně odhalí trolla v 80% případů, přičemž nejefektivnějším prvkem, dle kterého je možné detekci řídit, je chování moderátorů. Taktéž aktivita uživatele je příhodným indikátorem antisociálního chování a to především počet příspěvků za den. Nejslabším indikátorem je obsah příspěvku, což postojovou analýzu řadí do ne příliš užitečných metod detekce trollů.

Výzkumů a testů se samozřejmě provedlo více a vykonávali je různí badatelé. Někteří se výše zmíněnou metodou inspirovali, jiní vytvořili metodu zcela novou,

nicméně všechny závěry se víceméně shodovali ve svých objevech. Z čehož vyplývá, že fenomén trollingu je všudypřítomným problémem se stejnými prvky, které vychází ze správného datasetu. V našem výzkumu v praktické části práce bohužel nedošlo k hodnocení chování moderátorů, jelikož tento druh dat k dispozici nebyl. Věnovali jsme se zbývajícím třem metodám detekce. Hodnotili jsme obsah příspěvků, aktivitu uživatele a dotkli jsme se zkoumání chování komunity (lajky, retweety).

4.5.4. Vztahy v komunitě

Poslední a nejsložitější metodou při detekování antisociálního chování je zkoumání vztahů v komunitě. Metoda stojí na nástrojích SNA (*Social Network Analysis*).

Mihaylov a kol. (2015) byli prvními, kdo se vztahy v komunitě zabýval. Svou studii provedli na portále Slashdot Zoo, kde uživatelé označují ostatní členy za přátele nebo nepřátele. Získala se data, která jsou užitečná pro identifikaci nepopulárních uživatelů na základě důležitosti jednotlivých uzlů. Z nepopulárních uzlů jsou badatelé schopni vytřídit takové uživatele, kteří vykazují známky trollingu. Na síť je pohlíženo jako na množství vztahů, kde lze nalézt nejen uzly s pozitivní vahou, ale také uzly s negativní vahou. Váhy uzlů vyjadřují, zda si uživatelé navzájem věří, či nikoliv, z čehož je možné usoudit míru důvěryhodnosti každého uživatele (Mihaylov a kol., 2015).

Twitter API umožňuje vrátit se po časové ose a vytěžit data z 400 posledních příspěvků uživatele. Těží tweety a retweety. To je skvělý nástroj pro získávání informací o uživateli. Získat můžeme také celkový počet příspěvků nebo frekvenci přispívání. To vše společně s obsahem posledních čtyř set příspěvků může badatelům prozradit mnoho o chování uživatele (Tomaiuolo a kol., 2020).

PRAKTICKÁ ČÁST

Praktická část této diplomové práce se zaměřuje na detekci internetových trollů na českém Twitteru, respektive na jejich přispívání do diskuzních vláken na politických profilech.

Výběr Twitteru jako zkoumaného sociálního média byl ovlivněn několika faktory. Jednak jsou twitterové příspěvky krátké vzhledem k omezenému počtu znaků, jednak je twitterový obsah tvořen převážně texty, dále je Twitter vnímán v posledních letech jako součást profesní komunikace a konečně Twitter poměrně snadno umožňuje těžbu dat pomocí Twitter API. (viz. 3.3.1.)

Na české politiky a političky jsme se zaměřili primárně kvůli předpokladu vysokého počtu zmíněných záškodníků. Taktéž by politické příspěvky ze své podstaty měly být formálně i obsahově ucelené, čímž se v internetovém prostředí stávají texty vyšší úrovně, které je vhodné zkoumat na akademické půdě. Předpokládaný vysoký počet trollů pramení také z přesvědčení, že politická sféra prostupuje do různých úrovní života většinového obyvatelstva republiky, čímž způsobuje střet nejrůznějších sociálních skupin. Tento střet následně často ústí v ostrou diskuzi mezi členy skupiny vyhraněných názorů. Politika se týká všech obyvatel bez rozdílu rasy, náboženství, pohlaví. nebo věku. Proto jsme pokládali její výběr pro detekci trollů za vhodnou a objektivní možnost.

Detekce trollů probíhala prostřednictvím sběru dat z Twitteru, která se následně vyhodnocovala. U příspěvků vybraných politiků se sledovala vlákna odpovědí a detekce probíhala ručně, ovšem za určitých pravidel. Postup byl vždy stejný a byl zvolen právě tak, aby jej bylo možné zautomatizovat. Diplomová práce, ač nazvaná Automatická detekce trollů na českém Twitteru, nezahrnuje vyhotovený software, který by byl schopen automatické detekce uživatelů s vychýleným jednáním. Avšak slouží návodně k potenciálnímu stvoření softwaru v budoucnosti. Jedná se o sborník informací o chování internetových trollů, jejichž chování je prakticky ověřeno. Diplomovou práci lze taktéž považovat za soubor dílčích kroků jednoho z možných postupů při tvorbě automatické detekce, jelikož se při ruční detekci kroky v průběhu neměnily, neopakovaly a měly jasně daný řád.

V praktické části se podařilo ověřit teoretický popis trollů. Na základě jasně definovaných znaků v chování byl zvolen postup pro detekci, která zpětným ověřením podezřelých účtů dokázala hypotézu o antisociálním chováním některých

uživatelů, tzn společné znaky a důvody takového jednání. Cílem výzkumu tudíž není detekovat naprosto všechny trolly, nýbrž s naprostou jistotou detekovat alespoň některé.

5. Metoda výzkumu

5.1. Google Colaboratory

Výzkum probíhal pomocí internetové aplikace Google Colaboratory. Jedná se o online nástroj přístupný skrze prohlížeč, který umožňuje všem uživatelům s Google účtem snadno pracovat v programovacím jazyce Python a pomocí něj sbírat potřebná data. V Colabu si uživatel vytváří tzv. sešity, které kromě kódu mohou obsahovat taktéž texty, obrázky, HTML a další. Sešit je přehledně rozdělen do buněk s kódy, které je nutné při každém otevření spustit pro správné fungování vloženého kódu. Vytvořené sešity jsou uloženy na Google Disku, což z nich činí snadno dostupné soubory, které lze bez velkých obtíží sdílet s dalšími uživateli.

Colaboratory nevyžaduje žádné předchozí nastavení prohlížeče. Vezmeme-li v potaz velkou databázi předepsaných částí kódů v Pythonu, jež je dostupná všem zdarma na internetu, pak není divu, že uživatelská obec této aplikace sílí a Colaboratory se stává oblíbeným nástrojem nejen studentů, nýbrž i dalších datových vědců a UI výzkumníků.

5.2. Tři fáze výzkumu

Záměrem výzkumu bylo ověřit teoretický podklad čtvrté kapitoly, zejména pak zaměřit se na jiné detekovací metody, než na pouhou analýzu obsahu. Snažili jsme se pracovat se všemi atributy chování trolla v online prostředí, což se odráží na zvoleném postupu. (viz. 4.5.) Provedenou detekci je možné rozdělit na tři fáze, ve kterých probíhala.

První fáze byla přípravná. Bylo nutné zajistit přístup k Twitter API, připravit kódy v Google Colaboratory, zajistit ukládání CSV souborů na Disk, připravit podklady k výzkumu a určit jednoznačný postup.

Ve druhé fázi došlo ke zkoumání politických účtů, vytipování příspěvků, které by v konverzačním vlákně mohly obsahovat trolling a k podrobnému přezkoumání

komunikačních vláken vytipovaných příspěvků. V komunikačních vláknech byl kladen důraz na nalezení fenoménu trollingu, popřípadě flamingu, který k trollingu odkazuje. Následně byli označeni takoví uživatelé, kteří splnili předem dané parametry a tím se stali podezřelými z vykazování antisociálního chování.

Ve třetí a poslední fázi došlo k vytvoření kompletního seznamu potenciálních trollů, který prošel dalším tříděním. Filtrování vedlo k vytvoření seznamu uživatelů, kteří se v komunikačním vlákně s naprostou jistotou chovali antisociálně. Zda-li se jednalo o trolly či pouze o běžné uživatele, kteří se nekorektně vyjádřili jednorázově, se odhalilo při zpětném vhledu do jejich celkové aktivity na Twitteru. Toto zpětné ověření uživatelského chování vyústilo v úspěšnou detekci trollů, kterou by bylo možné zautomatizovat.

5.2.1. První fáze

Před samotným zahájením výzkumu bylo nutné připravit potřebné podklady. Nejprve jsme vytvořili v Google Colaboratory kódy, které zajišťovaly extrakci dat z Twitteru. O přístup k Twitter API (viz. 3.3.1.) musí výzkumník zažádat na webové stránce https://developer.twitter.com, kde mu po vyplnění formuláře správci Twitteru vygenerují vlastní přístupové kódy, díky kterým je následné prohlížení dat možné, stejně jako jejich těžba a vyhodnocení. Po získání zmíněného přístupu je možné data v Colabu těžit, nicméně pro jejich lepší vizualizaci jsme je převedli do CSV (Comma-Separated Values) souboru, tedy takového formátu, který umožňuje vytvoření tabulky a její otevření v MO Excelu nebo Tabulkách Google. Vytvořené soubor se ukládají na Disk Google.

Taktéž bylo vybráno deset twitterových účtů českých politiků, které jsme zkoumali a ve kterých jsme se snažili tolling detekovat. Politici byli vybráni tak, aby v seznamu bylo zastoupeno co možná nejvíce funkcí a politických stran. Ohled se bral taktéž na aktivitu příslušného účtu a na počet sledujících. S rostoucím počtem sledujících úměrně roste aktivita na profilu a s vysokým počtem komentářů u příspěvků roste šance k výskytu trolla. U kontroverzních nebo vyhraněných politiků se šance ještě zvyšuje. Původním záměrem bylo zkoumat 10 účtů s nejvyšším počtem sledujících, to se však v průběhu výzkumu ukázalo být neefektivním přístupem.

Z politiků, kteří jsou na Twitteru aktivní (viz. 3.6.4.), jsme do výzkumu zahrnuli ty, kteří jsou v následující tabulce ve sloupci Výzkum označeni slovem “ANO”. Jedná

se o 10 profilů s vysokým počtem sledujících. Heslem “není aktivní” jsou označeny takové profily, které příspěvky nepřidávají na denní bázi, jsou proto s ohledem na týdenní časový horizont (viz. 5.2.2.) pro výzkum nevyhovující. Úřad vlády ČR sice aktivní je, nicméně z výzkumu byl vyřazen kvůli neaktivním sledujícím a chybějícím komentářům u retweetovaných příspěvků. Uživatel nevytváří vlastní obsah a nemá engagement publika. Připravený kód v Google Colaboratory z neznámého důvodu nešel spustit v případě Aleny Schillerové, proto taktéž nebyla do výzkumu zahrnuta.

Jméno	**Odkaz (uživatelské jméno)**	**Sledující (tis.)**	**Výzkum**
Andrej Babiš	https://twitter.com/AndrejBabis	454,6	ANO
Miroslav Kalousek	https://twitter.com/kalousekm	193,3	ANO
Dominik Feri	https://twitter.com/DominikFeri	192,7	není aktivní
Václav Moravec	https://twitter.com/vaclavmoravec	163,2	není aktivní
Úřad vlády	https://twitter.com/strakovka	133,3	pouze RT, nemá engagement
Ivan Bartoš	https://twitter.com/PiratIvanBartos	93,3	ANO
Petr Fiala	https://twitter.com/P_Fiala	90,8	ANO
Jan Hamáček	https://twitter.com/jhamacek	87,9	ANO
Adam Vojtěch	https://twitter.com/adamvojtechano	80,5	ANO
Jiří Ovčáček	https://twitter.com/PREZIDENTmluvci	78,2	ANO
Markéta Pekarová Adamová	https://twitter.com/market_a	73,4	ANO
Karel Schwarzenberg	https://twitter.com/schwarzenberg_k	51,8	není aktivní
Alena Schillerová	https://twitter.com/alenaschillerov	38	nelze spustit
Lubomír Zaorálek	https://twitter.com/ZaoralekL	36,3	ANO
Marian Jurečka	https://twitter.com/MJureka	30	ANO
Tomio Okamura	https://twitter.com/tomio_cz	29,4	NE
Senát	https://twitter.com/SenatCZ	20,4	NE
Poslanecká Sněmovna	https://twitter.com/snemovna	20,3	NE
Robert Plaga	https://twitter.com/RobertPlaga	15	NE
Radek Vondráček	https://twitter.com/vondraczech	5,7	NE

tab. 1 (viz. Příloha)

5.2.2. Druhá fáze

Po zkušebním testu funkčnosti nejen kódu, nýbrž také postupu výzkumu došlo na adekvátní úpravy. Následně započala druhá fáze.

Z Twitteru se extrahovala aktivita příslušného uživatele (viz. 5.2.1.), respektive jeho posledních 400 tweetů. Kromě uživatelského jména a textu příspěvků byl

zkoumán také datum a čas jednotlivých příspěvků, počet sledovaných a sledujících, počet lajků a retweetů, ID konverzace a ID tweetu. Stažená data byla zkoumána v Tabulkách Google a měla následující podobu:

	Uživatel	Sledujících	Sledovaných	Text tweetu	ID_tweetu	Čas	Lajky	Retweety	conversation_id
6	Andrej Babis	452756	241	Mám zároveň ohromnou radost, že nám skvěle fungují očkovací centra bez registrace. Na pražském Chodově i na Hlavním nádraží je stále velký zájem. V pondělí proto otevřeme další očkovací místo v obchodním centru Nový Smíchov, ve středu v obchodních centrech v Brně a v Ostravě.	1416040296907948032	2021-07-16 14:21:30	78	5	1416040295108644867
7	Andrej Babis	452756	241	Děkuji Maďarsku, Viktoru Orbánovi i maďarskému velvyslanci za to, že nám na mou žádost dnes půjčili další vakcíny. V @uvn_praha jsme vyložili 200 tisíc vakcín Pfizer/BioNTech. Celkově nám tak už Maďarsko půjčilo 340 tisíc vakcín, které teď potřebujeme a do konce roku vrátíme. https://t.co/qSpCyhaFlp	1416040295108644867	2021-07-16 14:21:29	141	9	1416040295108644867

tab. 2 (viz. Příloha)

Tabulka byla dále upravena do požadované podoby. Pracovali jsme pouze s příspěvky, které nebyly starší než jeden týden od extrakce dat. Při těžbě a vyhodnocování konverzačních vláken u příspěvků starších než jeden týden docházelo k nesprávnému fungování kódu, výsledky by tudíž byly značně zkreslené. Taktéž došlo k odfiltrování příspěvků s nízkým dosahem. Spodní hranice pro počet lajků byla stanovena na čísle 50, pro retweet na čísle 5. Původní plán výzkumu počítal s hranicí 100 pro lajky a 10 pro retweety, nicméně taková čísla se v našem prostředí ukázala jako příliš vysoká. Vhodné by jistě bylo zahrnout taktéž počet komentářů, respektive odpovědí, nicméně v programovacím jazyce Python nelze taková data extrahovat a pokud ano, jejich pravdivost není spolehlivá.

V dalším kroku druhé fáze jsme podrobně pracovali s každým jednotlivým příspěvkem, který prošel filtračním sítem. K těžbě dat bylo užito ID konverzace (*conversation ID*), tedy specifické číslo celého konverzačního vlákna. ID konverzace je stejné pro původní příspěvek i všechny jeho přidružené příspěvky, tzn. odpovědi a vnořené odpovědi. Kromě samotného obsahu textů jednotlivých příspěvků a ID konverzace bylo pro výzkum důležité taktéž znát uživatelské jméno a ID autora

příspěvku. Sloupec “level” v tabulce označuje míru vnoření odpovědi. Level 1 je označení pro přímou odpověď původnímu tweetu, level 2 značí odpověď na odpověď, atd. Poslední sloupec obsahuje odhad celkového vyznění příspěvku (pozitivní, negativní nebo neutrální), vygenerovaný algoritmem automatické postojové analýzy. Uvedený algoritmus využívá principy strojového učení a v tomto konkrétním případě je bohužel natrénován na uživatelských recenzích, nikoliv na politických komentářích, neboť vhodná data nebylo možné získat, a tak přesnost uvedeného odhadu není příliš vysoká. Pro účely výzkumu však algoritmus zcela postačil.

Troll vyvolává konflikt, uráží ostatní uživatele, provokuje a odvádí tok diskuze, čímž obvykle spustí flaming. Přispívá pravidelně a často, většinou v krátkých příspěvcích, čímž upřednostňuje kvantitu před kvalitou. Jeho cílem je na sebe strhnout pozornost množstvím irelevantních zpráv. (viz. 4.1., 4.2.) Pro snazší detekci výzkum pracoval pouze s těmi příspěvky, které postojová analýza vyhodnotila jako negativní. V dalším kroku se po abecedním seřazení přispěvatelů označili ti uživatelé, kteří přispěli třikrát a vícekrát do jedné diskuze, nebo dvakrát a vícekrát do různých diskuzí. Častí přispěvatelé byli pro přehlednost označeni růžovou barvou, která značí možný výskyt trolla. V posledním kroku jsme se zaměřovali na obsahovou analýzu jednotlivých příspěvků a za možného trolla označili každého, kdo ve své komunikaci užil vulgarismu. Pomocí postojové analýzy, vyhodnocení četnosti příspěvků a jejich obsahu by se dle teoretického podkladu mělo podařit odhalit antisociální chování.

Nutno podotknout, že se v této fázi výzkumu stále pracovalo pouze s jedním vybraným politickým účtem. Ačkoliv se v jedné tabulce nachází souběžně několik konverzačních vláken, jedná se vždy o konverzační vlákna na profilu jednoho politika.

	conversation_id	level	tweet_id	author_id	author_name	text	Sentiment
295	1414265981690683394	1	1414281472970743811	1305294469705728001	_3RUR	@AndrejBabis Dneska 3/10, chce to víc lží a nenávisti. Opřete se víc a agresivněji do všech, nejen Pirátů.	neg
61	1414265981690683394	2	1415018966532173826	3559582877	115Lada	@PavelMaron @AndrejBabis Jak jsi přišel na to,že tzv.duchodci neumí s PC,hlavou chytrá? Vzdělaný člověk umí,nevzdelany háze lopatou bez rozdílu věku	neg
366	1414265981690683394	2	1415016795610787846	3559582877	115Lada	@Filmgigant @AndrejBabis Jdi s tím do prdele	neg
373	1414265981690683394	2	1415017690486423564	3559582877	115Lada	@Lukas__Valenta @AndrejBabis Ale hovno	neg
341	1414265981690683394	1	1414276532793708545	3158799633	11Kalmartin	@AndrejBabis @matej_pluhar Máro, jdi do PIČI už.	neg

tab.3 (viz. Příloha)

5.2.3. Třetí fáze

V poslední fázi se výzkum soustředil na ověření potenciálních trollů. Všechny barevně označené příspěvky z naprosto všech komunikačních vláken byly umístěny do jednoho souboru, kde proběhla druhá obsahová analýza.

Postupně byli vyloučeni uživatelé, kteří sice byli označeni za potenciálního trolla, ale ukázali se v této fázi být spíše naštvanými uživateli. Vyloučili se tedy tací, kteří byli označeni pouze jednou napříč všemi zkoumanými politickými účty, jelikož užití vulgárních nebo neslušných výrazů v malém množství nesouhlasí s teoretickou definicí trolla.

Taktéž byli vyloučeni uživatelé, kteří přispěli do komunikačního vlákna mnohokrát, nicméně jejich příspěvky byly vždy slušně napsané. Takový uživatel přispěje mnohokrát, jelikož konstruktivně diskutuje o dané problematice, nejde mu tedy o odklonění toku diskuze.

Dále byly vyřazeny účty, které ve svých příspěvcích užívaly mnoho vulgarit a nadávek, ale obsah jejich příspěvků byl koherentní, texty poměrně dlouhé a logicky

strukturované. Obsah těchto textů nepostrádal smysl a nebyl konstruován tak, aby vyvolal při, nýbrž byl jen napsán velmi hrubě.

Po tomto přezkoumání zbylo z původních tří set devadesáti pěti potenciálních škůdců pouhých sedmnáct účtů, které se z doposud dostupných informací jevily jako trollové. Těmto uživatelům byla věnována další pozornost. Stejně jako byla ve druhé fázi výzkumu vytěžena data, díky kterým bylo možné vizualizovat veškerou aktivitu politiků, i v tomto kroku byla data vytěžena k monitorování veškeré aktivity trollů. Z následné obsahové analýzy nejnovějších příspěvků vyplynulo, kteří uživatelé skutečnými trolly jsou, kteří jsou trolly jen možná a u kterých došlo k omylu. Následující tabulka zobrazuje kompletní seznam sedmnácti podezřelých profilů, u kterých došlo ke zpětnému přezkoumání. V tabulce je rovněž ukázka jednoho příspěvku s náležitým počtem lajků a retweetů, který se kromě řady dalších parametrů stal jedním z důvodů označení uživatele za potenciálního trolla.V tabulce se dále nachází ID konverzace, ID samotného tweetu, počet sledujících, sledovaných a také sloupec, který označuje finální zhodnocení uživatele, tzn. vykazuje-li nebo nevykazuje-li známky antisociálního chování.

Uživatel	Sledujících	Sledovaných	Text tweetu	ID_tweetu	Čas	Lajky	RT	conversation_id	Troll
CapekCapekjiri	10	79	@AndrejBabis Bureši jsi strašná svině	1421890417931468800	2021-08-01 17:47:47	3	0	1421888351024599044	ANO
CtiradMasin	1	0	@SociologPetrHam @PREZIDENTmluvci Už aby ta nahnědlá krysa táhla z Hradu!	1422429295302037512	2021-08-03 5:29:06	1	0	1422260846646939657	ANO
dp26498978	16	152	@EdmontMoudry @tomio_cz Přesně tohle na tý mikině je, ty kreténe.	1420138018933981184	2021-07-27 21:44:23	0	0	1418960075331411974	NE
Garmix1	0	24	@nojumper66 @tessina19 Lol tak jestli má baba pěknou kundu jsem nikdy nehodnotil.	1409382604734926850	2021-06-28 5:26:12	0	0	1408964493544148996	MOŽNÁ
geor_andrews	36	128	@stepan_hanzlik @tomio_cz Normální, původní, tedy nikoli tupá migrující africká praopice nebo musulmán.	1419398937102925829	2021-07-25 20:47:32	0	0	1418960075331411974	MOŽNÁ

Isabell21525070	72	266	RT @ivokubes: No co, volby se blíží. Debilova vláda potřebuje prodat úspěchy.	1421491139295776770	2021-07-31 15:21:12	0	1	142149113929576770	MOŽNÁ
Jen_jdu_okolo	1967	644	@wambudambu79 Kdy vydáš sbírku wambásní?	1422946748260048897	2021-08-04 15:45:16	2	0	142288945641210675 6	NE
JiriMart	8	21	@PREZIDENTmluvci A zase jsi to Ty blbe spletl! Žijeme v diktatuře Zamanostátu hradního idiota!!!	1421017423411654657	2021-07-30 7:58:49	0	0	142078722004733 5424	ANO
JosefzizkaR	308	81	@adamvojtechano @SUKLcz Táhni zmrde https://t.co/TIPAE06Ory	1422951987960819717	2021-08-04 16:06:05	1	0	142293449994025 3697	ANO
Ken98811667	0	0	Diskriminační píčoviny to vymyslí je kretén jako je Vojtěch!	1421507823410765824	2021-07-31 16:27:30	0	0	142150782081029 7352	MOŽNÁ
Mengele85170837	3	9	@AndrejBabis @jhamacek @KarelHavlicek_ Tornádo už vás omrzelo?	1423003051917467649	2021-08-04 19:29:00	0	0	142283570010639 9744	ANO
mch_mch_msl_mch	154	719	@petrrossan "Vyližte mi prdel."	1423193915700695040	2021-08-05 8:07:25	0	0	142317948133491 9168	NE
Mrdamtwitr	13	114	@KarelHavlicek_ Ptám se: Kdy půjdeš do prdele, mrdko?	1423229680501633025	2021-08-05 10:29:32	0	0	142321977510744 4737	ANO
PalzerRadim	0	5	@alenaschillerov @KarelHavlicek_ Ty seš fakt ubohá hloupá kráva	1422233221010448387	2021-08-02 16:29:58	1	0	142218653864186 2658	ANO
pavel_krupicka	0	54	@PREZIDENTmluvci Ty informuješ pravdivě zmrde?seš prolhanej jak když ty tvé noviny tisknou.	1414842338481000449	2021-07-13 7:01:14	0	0	141321495454778 5735	ANO
RazyHea1	0	25	@novinkycz A co takhle naočkovat tvojí hubu a celou tvojí famílií ZMRDE!!	1414180553193103360	2021-07-11 11:11:32	0	0	141376015487766 3953	ANO
SpendlikRoman	10	142	@jhamacek Jsi debil ale to ty nepocopíš	1421196093652717577	2021-07-30 19:48:47	1	0	142119376114569 2164	ANO

tab. 4 (viz. Příloha)

Slovem ANO ve sloupci “Troll” byly označeny takové twitterové profily, jejichž příspěvky jsou pravidelné, z valné většiny vulgární, někoho urážejí, jsou nesmyslné a nelogické nebo provokují k další reakci ostatních uživatelů. Jinými slovy příspěvky přesně kopírují teoreticky položený základ pro koncept fenoménu trollingu.

Slovem MOŽNÁ ve sloupci “Troll” jsme označili účty, které přispívají pravidelně na denní bázi, obsah se však liší. Zhruba polovina příspěvků se jeví jako antisociální komunikace. Provokuje a uráží, je vulgární a vysmívá se, do diskuze nepřispívá validními argumenty, nýbrž pouze útočí na druhou osobu. Často zde můžeme pozorovat kromě argumentů ad hominem i další argumentační fauly. Druhá polovina příspěvků, většinou směrem k neveřejným osobám, je však podaná slušně a přátelsky. Lze se domnívat, že se většinou jedná o lidi, které daný uživatel zná z offline života, tedy kolegy, přátele nebo rodinu (viz. Příloha). Z toho plyne, že účty zařazené do MOŽNÁ skupiny se antisociálně chovají jen v některých případech a nelze je proto s naprostou jistotou zařadit mezi trolly.

Označení NE ve sloupci “Troll” získaly twitterové účty, u kterých při detekci došlo ke zjevnému omylu. Zřídka se sice v souhrnu jejich aktivity objevují vulgarismy nebo nic neříkající příspěvky, je jich však tak málo, že je lze považovat za výjimky. Ve většině případů uživatelé komunikují slušně, podávají relevantní informace, logicky strukturují argumenty v diskuzi, zbytečně neútočí na další uživatele, ba dokonce nepřispívají tak pravidelně jako tomu bylo u předchozích skupin uživatelů. Mnohdy nejsou aktivní každý den a jejich záměrem nebývá strhnout na sebe pozornost nebo započít válku v diskuzním vlákně.

6. Výsledky

6.1. Hodnocení příspěvků

Začátek výzkumu se odrážel od deseti českých politických účtů, přičemž u každého účtu bylo zkoumáno posledních čtyři sta příspěvků. Ve druhé fázi (viz. 5.2.2.) se z původních čtyř tisíc příspěvků vyfiltrovalo sto dvacet osm, které se podrobně zkoumaly. Filtrování proběhlo dle datumu, kdy příspěvky nemohly být starší než jeden týden a dle počtu lajků a retweetů, kdy lajků musel mít příspěvek minimálně padesát a retweetů minimálně pět. Každý politický účet měl přirozeně v konečném čísle jiný podíl příspěvků, které se do výzkumu hodily. Nejzastoupenějším politikem co do počtu zkoumaných komunikačních vláken se stal Jiří Ovčáček s dvaceti čtyřmi příspěvky. Naopak nejméně zastoupeným politikem se ve výzkumu stal Jan Hamáček s pouhými třemi příspěvky (viz. Příloha). Zajímavý je zde fakt, že Jana Hamáčka sleduje o více než deset tisíc víc uživatelů oproti Jiřímu Ovčáčkovi

(viz. tab. 1). S vyšším počtem sledujících by se dala očekávat větší aktivita u příspěvků, což by vedlo k frekvetovanějšímu výskytu internetových trollů. Po ověření jsme došli k závěru, že na vině není nízká aktivita sledujících Jana Hamáčka, nýbrž Hamáčkova aktivita. Hamáček na svůj twitterový účet nepřispívá denně a počet sdílených příspěvků je výrazně nižší než u Jiřího Ovčáčka.

Při hodnocení komunikačních vláken se pracovalo s naprosto všemi příspěvky se stejným komunikačním ID, tzn. se všemi dostupnými reakcemi na příspěvky, které splnili podmínky předešlého kroku. Reakce se dále filtrovaly dle postojové analýzy. Zkoumaly se pouze takové odpovědi, které postojová analýza vyhodnotila za negativní. Z negativních odpovědí se za potenciálního trolla označili uživatelé, kteří přispěli vícenásobně a uživatelé, kteří ve svých projevech použili vulgarismy.

6.1.1. Politické účty

V případě Andreje Babiše bylo zkoumáno celkem tři sta sedmdesát tři příspěvků do šestnácti různých komunikačních vláken, z nichž sto čtyřicet tři příspěvků bylo označeno za potenciální trolly. V konečné fázi byli za trolly označeni tři uživatelé: @Garmix1 @JosefzizkaR a @jen_jdu_okolo. Při zpětném ověření byl uživatel @jen_jdu_okolo vyloučen ze seznamu trollů a uživatel @Garmix1 byl označen za možného trolla. V příspěvcích Andreje Babiše jsme s takto nastaveným postupem byli schopni detekovat s naprostou jistotou jednoho internetového trolla.

U uživatelského profilu Ivana Bartoše jsme zkoumali sedmnáct komunikačních vláken a pět set sedmdesát sedm příspěvků od různých uživatelů. Sto devadesát pět bylo následně označeno za potenciální trolly, z nichž pouze jeden uživatel byl zařazen do finálních sedmnácti. Uživatele @Isabell21525070 jsme nakonec označili za možného trolla, jelikož příspěvky na jeho profilu jsou různého rázu. S naprostou jistotou se na profilu Ivana Bartoše nepodařilo detekovat žádného trolla.

U Petra Fialy jsme zkoumali osm komunikačních vláken s celkovým počtem osmdesáti příspěvků, z nichž pouze osm bylo označeno za potenciální trolly. Žádný z nich však nebyl ve finálním kroku označen za trolla. U Petra Fialy se tedy s naprostou jistotou nepodařilo detekovat trolla.

Pouze tři komunikační vlákna z profilu Jana Hamáčka a dvacet devět příspěvků bylo zkoumáno. Šest z nich bylo označeno za potenciální trolly a pět z nich napsali uživatelé, kteří ve finální fázi byli označeni za trolly. Příspěvky napsali čtyři uživatelé: @JosefzizkaR, @Mrdamtwitr, @PalzerRadim a @SpendlikRoman. Všechny čtyři jsme s naprostou jistotou zařadili mezi trolly. V příspěvcích Jana Hamáčka jsme během jednoho sledovaného týdne detekovali čtyři trolly.

Marian Jurečka napsal během jednoho týdne třináct příspěvků, které vyhovovaly požadavkům a které jsme podrobili zkoumání, přičemž u třech z nich se jednalo o odpovědi na tweety J. Hamáčka a J. Maláčové. Tyto odpovědi dále tvořily svoje komunikační vlákno. Třináct komunikačních vláken obsahovalo sto dvacet devět příspěvků, které postojová analýza vyhodnotila za negativní. Třicet šest příspěvků bylo označeno za vytvořené potenciálními trolly a dva z nich zůstaly označeni i v konečné fázi. Jednalo se o příspěvky uživatelů @SpendlikRoman a @PalzerRadim. Nutno podotknout, že se zmínené příspěvky nacházely právě ve vláknech, kam M. Jurečka taktéž přispíval odpovědí. V komunikačních vláknech, která přímo navazovala na autorské příspěvky Jurečky se za týdenní aktivitu nepodařilo detekovat žádného trolla.

U Miroslava Kalouska jsme zkoumání podrobili čtrnáct komunikačních vláken a dvě stě devadesát osm příspěvků, z nichž šedesát devět bylo označeno za potenciální trolly a dále jim byla věnována pozornost. Žádný z označených příspěvků však v konečné fázi označení nepotvrdil a u Miroslava Kalouska jsme žádného trolla detekovali.

V případě Tomia Okamury jsme pracovali s dvanácti komunikačními vlákny a dvě stě dvaceti jedna příspěvky, z nichž osmdesát jedna bylo označeno k dalšímu přezkoumání k nalezení fenoménu trollingu. Pět příspěvků bylo v konečné fázi označeno za napsané trollem. Autoři takových příspěvků byli @geor_andrews a @dp26498978. Ani jeden ze zmíněných nebyli nebyli s naprostou jistotou označeni za trolla. Uživatel @dp26498978 byl v posledním kroku ze seznamu vyloučen. Uživatel @geor_andrews byl označen za možného trolla, jelikož se styl jeho komunikace s každým příspěvkem liší. S naprostou jistotou jsme ani u Tomia Okamury nedetekovali žádného trolla.

Při zkoumání účtu Jiřího Ovčáčka jsme pracovali s dvaceti čtyřmi komunikačními vlákny a čtyři sta čtyřiceti třemi příspěvky. Jako potenciální trolly jsme označili dvě stě dva příspěvků dvacet sedm z nich prošlo i do konečné fáze.

Příspěvky napsalo šest uživatelů: @CapekCapekjiri, @CtiradMasin, @JiriMart, @JosefzizkaR, @mch_mch_msl_mch a @pavel_krupicka. Jediný @mch_mch_msl_mch byl po zpětném ověření ze seznamu trollů vyloučen. U Jiřího Ovčáčka se s naprostou jistotou podařilo detekovat pět trollů.

U Markéty Pekarové jsme vybrali šest komunikačních vláken. Z dvě stě třiceti sedmi příspěvků jsme za potenciální trolly označili osmnáct. Žádný z nich však v konečné fázi za trolla označen nebyl. U Pekařové se s naprostou jistotou nepodařilo detekovat žádného trolla.

Adam Vojtěch napsal za týden čtrnáct příspěvků, které vyhověly požadavkům zkoumání. Jejich komunikační vlákna obsahovala osm set osmdesát sedm příspěvků a dvě stě dva z nich bylo určeno k dalšímu zkoumání. Finální počet příspěvků, které obsahovaly známky antisociálního chování byl roven dvaceti dvoum. Diskuzí se účastnili uživatelé @Isabell21525070, @JosefzizkaR, @Ken98811667, @Mengele85170837, @mch_mch_msl_mch, @Mrdamtwitr, @PalzerRadim a @RazyHea1. Uživatel @mch_mch_msl_mch byl po zpětném ověření twitterového účtu ze seznamu trollů vyřazen. Uživatelé @Isabell21525070 a @Ken98811667 byli označeni za možné trolly. U Adama Vojtěcha jsme s naprostou jistotou detekovali pět trollů.

Jméno	Uživatel. jméno	Vlákna	Příspěvky	Potenc. trollové	Troll příspěvky	Troll - uživatelé
Andrej Babiš	@AndrejBabis	16	373	143	7	1 (@JosefzizkaR)
Ivan Bartoš	@PiratIvanBartos	17	577	195	1	0
Petr Fiala	@P_Fiala	8	80	8	0	0
Jan Hamáček	@jhamacek	3	29	6	4	4 (@JosefzizkaR. @Mrdamtwitr, @PalzerRadim, @SpendlikRoman)
Marian Jurečka	@MJureka	13	129	36	2	0
Miroslav Kalousek	@kalousekm	14	298	69	0	0
Tomio Okamura	@tomio_cz	12	221	81	5	0
Jiří Ovčáček	@PREZIDENTmluvci	24	443	202	27	5 (@CapekCapekjiri, @CtiradMasin, @JiriMart, @JosefzizkaR, @pavel_krupicka)

Markéta P. Adamová	@market_a	6	237	18	0	0
Adam Vojtěch	@adamvojtechano	14	887	202	22	5 (@JosefzizkaR, @Mrdamtwitr, @Mengele85170837, @PalzerRadim, @RazyHea1)

tab. 5 (viz Příloha)

Ač bylo zkoumáno nejvíce vláken na twitterovém účtu Jiřího Ovčáčka, nejvíce příspěvků obsahovala komunikační vlákna Adama Vojtěcha. Z toho vyplývá, že mluvčí prezidenta ČR je sice velice aktivní, ale sledující Adama Vojtěcha píší dvojnásobný počet komentářů. Odhadujeme, že k tzv. flamingu zde dochází poměrně často, tudíž by měl být vysoký i počet výskytu antisociálních uživatelů. S nalezením celkem pěti trollů ve vláknech Vojtěcha se hypotéza potvrzuje.

Na profilu Jiřího Ovčáčka jsme detekovali výskyt pěti internetových trollů, což je vůbec nejvyšší počet výzkumu. Trolly jsme dále detekovali u Andreje Babiše a Jiřího Hamáčka. U ostatních politiků se žádné internetové trolly s naprostou jistotou detekovat nezdařilo. Buď se jejich identita coby internetových škůdců nepotvrdila při zpětném ověřování vůbec nebo jen částečně.

Velice překvapující je vysoký výskyt trollů v konverzačních vláken Jiřího Hamáčka. V pouhých dvaceti devíti příspěvcích se podařilo odhalit hned čtyři internetové trolly. V poměru zkoumaných příspěvků s detekovanými trolly byla detekce na tomto účtě jednoznačně nejúspěšnější. Většina detekovaných uživatelů je přitom shodná pro Hamáčka a Vojtěcha. Tři stejní uživatelé ruší u zmíněných politiků internetovou diskuzi.

Dalším nepředvídaným překvapením se stal fakt, že se nezdařilo odhalit žádného trolla na profilu Ivana Bartoše. V pěti stech sedmdesáti sedmi příspěvcích bylo za potenciální příspěvky s fenoménem trollingu označeno sto devadesát pět, což je třetí nejvyšší číslo označených příspěvků. Vyšší číslo označených příspěvků lze nalézt na profilech Vojtěcha a Ovčáčka, přičemž na obou zmíněných se detekce trollů zdařila.

Na profilu Petra Fialy se nejen nezdařilo žádného trolla detekovat, ale také počet zkoumaných vláken a příspěvků je velmi nízký oproti ostatním zkoumaným politikům. Z toho lze usuzovat, že Fiala na svém účtu není denně aktivní. Ačkoliv ho sleduje 90,8 tisíc lidí, což se v rámci našeho výzkumu řadí do horní poloviny, engagement sledujících vysoký není, čímž se snižuje i možnost detekce trolla. Nejen, že se s vyšším počtem přispěvatelů úměrně zvyšuje i pravděpodobnost jeho

výskytu, ale v zájmu trolla není zapojovat se do “neviditelných” diskuzí. Troll na sebe chce strhnout pozornost, chce být viděn a chce změnit diskuzi. Takové aspirace by zůstaly u příspěvků bez reakcí nenaplněny.

6.2. Hodnocení trollů

Při zpětném ověřování ve třetí fázi výzkumu jsme byli schopni detekovat čtrnáct uživatelů, kteří vykazují známky antisociálního chování. Mezi trolly řadíme uživatele, kterým ve čtvrté tabulce (viz. 5.2.3.) byla přiřazena hodnota ANO nebo MOŽNÁ. Při jejich označení byl brán ohled nejen na obsahovou analýzu příspěvků, ale také na četnost a pravidelnost postování. Troll by měl přispívat každý den mnohými texty, které na sebe nutně nemusí navazovat a nemají tak logické uspořádání. Pozornost byla věnována rovněž počtu sledujících a sledovaných. Ačkoliv je obvykle trollem definován uživatel, který se snaží začlenit do online kolektivu a buduje si tak stálou fanouškovskou základnu na sociálních médiích, v českém prostředí nalezneme i uživatele využívající možnou anonymitu na internetu. Podezřelé se proto stávají i účty s malým počtem sledujících nebo s žádnými sledujícími. Kolik twitterových účtů daný uživatel sleduje bylo pro výzkum méně důležité než jaké účty to jsou. Obvykle můžeme pozorovat, že je troll aktivní u více účtu v rámci určité komunity. Příkladem, který lze vztáhnout k našemu výzkumu, by mohla být četná aktivita na mnoha politických účtech.

Pro další ověření faktu, že se jedná o skutečného trolla a nedošlo k omylu, by mohla sloužit taktéž obsahová analýza reakcí ostatních uživatelů na daný příspěvek. Tento výzkum ale s analyzováním odpovědí komunity pracoval pouze namátkově, nejedná se proto o systematické ověření. Přesto by byl tento krok při budování softwaru automatické detekce možné zařadit. Tok diskuze by po příspěvku trolla měl být odkloněn. U jindy slušně se vyjadřujících uživatelů bychom mohli pozorovat agresivní reakce, které u nich příspěvek trolla vyvolal. Jinými slovy by po takovém příspěvku mělo dojít k flamingu, který by se měl po nějakém čase opět uklidnit a tok diskuze by se měl znovu stočit k původnímu tématu.

6.2.1. Uživatelé, co trolly nejsou

Ve čtvrté tabulce (viz. 5.2.3.) se nachází tři uživatelé, kteří se ve druhé fázi zkoumání zdáli být trolly, nicméně ve třetí fáze se tato podezření vyvrátila. Uživatelé @dp26498978, @Jen_jdu_okolo a @mch_mch_msl_mch se ve svých příspěvcích

nevyjadřují vulgárně, příspěvky mají logickou strukturu a jsou významově obsáhlé. Počet sledujících i sledovaných se s každým uživatele různí, nejedná se však o anonymní profily, které by sledovalo nula účtů.

Uživatel	Sledujících	Sledovaných	Počet příspěvků	Politici	Troll
dp26498978	16	152	2	Okamura	NE
Jen_jdu_okolo	1967	644	2	Babiš	NE
mch_mch_msl_mch	154	719	5	Ovčáček, Vojtěch	NE

tab. 6 (viz. Příloha)

Uživatel @dp26498978 je aktivní každý den, účastní se politických diskuzí na mnoha politických profilech a lze u něj detekovat vulgarity. Nicméně neučástí se pouze politických diskuzí, nýbrž je aktivní v mnoha oblastech. Nesleduje pouze vyhraněnou skupinu, kde by jeho účelem bylo otravovat diskuzi. Přispívá konstruktivně a vulgaritu nebo provokaci u něj pozorujeme spíše výjimečně. Ve výzkumu byl detekován ve vlákně Tomia Okamury a původně byl označen za trolla kvůli příspěvkům jako “Přesně tohle na tý mikině je, ty kreténe.” (viz. Příloha). Ze seznamu byl po zpětném ověření pro tyto atributy vyloučen.

Pro uživatele @Jen_jdu_okolo taktéž platí aktivita ve všemožných kruzích. U příspěvků s politickou tématikou můžeme jako u jediných pozorovat známky agrese a užívání vulgárních výrazů. Spíše než o záměrnou provokaci se zde ale jedná o rozhořčené vyjádření politických názorů. Uživatel má vzhledem k poměrně vysokému počtu sledujících málo lajků a prakticky nikdo další jeho příspěvky neretweetuje. Tato skutečnost poukazuje na to, že uživatel pravděpodobně nebaží po pozornosti, jak by tomu bylo v případě trolla. Ve výzkumu byl za trolla označen ve vlákně Andreje Babiše, např. kvůli následujícímu příspěvku: “Jdi do prdele, hovado estébácký! V tomhle Debilistáně se člověk nemůže ani normálně otestovat, když už to vy kreténi nesmyslně nařizujete i lidem s protilátkama!!!”. I díky faktu, že jsou jednotlivé příspěvky dlouhé, jsme usoudili, že zmíněný uživatel je patrně autentický svému já a na Twitteru záměrně nechce kazit požitek z online prostředí dalším uživatelům.

Třetím vyřazeným uživatelem ze seznamu trollů je @mch_mch_msl_mch. Jedná se o velmi aktivního uživatele, který denně napíše desítky twitterových

příspěvků. Příspěvky nesplňují kritéria, která jsme určili pro bezpečné označení trolla. Zmíněný uživatel se dokonce zřídka vyjadřuje k politické scéně nebo k veřejným osobám obecně. Z jeho slušné komunikace lze vyvodit, že Twitter užívá ke svým osobním účelům v souladu s pravidly. Za trolla byl označen kvůli jeho aktivitě ve vlákně Jiřího Ovčáčka, kde se projevil tímto nevhodným způsobem: “Tohle čti tý starý kurvě před spaním, blbečku. https://t.co/uKEj20LLpp”. Detekovaný byl také u Adama Vojtěcha.

6.2.2. Uživatelé, co možná trolly jsou

Ze sedmnácti označených trollů byla čtyřem přiřazena hodnota MOŽNÁ. Tito uživatelé některé atributy antisociálního chování splňovali, jiné ne. Nelze o nich s naprostou jistotou rozhodnout, zda skutečnými tolly jsou nebo zda s trolly některé vlastnosti sdíli zcela náhodně. Jedná se o uživatele @Garmix1, @geor_andrews, @Isabell21525070 a @Ken98811667.

Uživatel	Sledujících	Sledovaných	Počet příspěvků	Politici	Troll
Garmix1	0	24	4	Babiš	MOŽNÁ
geor_andrews	36	128	4	Okamura	MOŽNÁ
Isabell21525070	72	266	3	Bartoš, Vojtěch	MOŽNÁ
Ken98811667	0	0	2	Vojtěch	MOŽNÁ

tab. 7 (viz. Příloha)

Uživatel @Garmix1 nepřispívá denně, je aktivní v průměru zhruba jednou za tři dny a přispívá do diskuzí veřejně známých osobností, ale i do diskuzí ostatních uživatelů. Obsahově se jeho příspěvky různí. Některé provokují, některé jsou vulgární, některé urážejí, ale jiné jsou zcela v normě. Tento uživatel nemá žádného sledujícího a sám sleduje pouze dvacet čtyři lidí. Některými svými znaky připomíná chování internetového trolla a jinými toto přesvědčení vyvrací.@Garmix1 byl detekován v komunikačním vlákně Andreje Babiše.

Podobně si stojí @geor_andrews. Tento uživatel přispívá denně velmi hojně po určitou dobu. Následně se na pár dní odmlčí a nepřispívá ničím. Po pár dnech opět přichází vlna aktivity a tak stále dokola. Do politické scény na Twitteru příliš nevstupuje a reaguje na různé podněty do konverzačních vláken ostatních uživatelů. Část příspěvků je vesměs urážlivá, xenofobní a rasistická, občas je užit vulgární

termín. Další část ale tvoří dlouhé a logicky strukturované příspěvky, s čímž se u trolla nesetkáváme. Zároveň má třicet šest sledujících, což není mnoho, ale není to ani vyloženě málo. Nelze proto s jistotou tvrdit, zda se o trolla jedná nebo ne. Detekován byl v příspěvcích Tomia Okamury.

V příspěvcích @Isabell21525070 rovněž nalézáme ambivalentní chování. Některé příspěvky jsou slušné, obsahově úplné, jiné jsou pouhým vyslovením nadávky nebo urážky. Uživatel má poměrně velkou fanouškovskou základnu, sleduje ho sedmdesát dva účtů. Zároveň je @Isabell21525070 aktivní každý den, přispívá hojně a pravidelně. Těžko soudit, zda se jedná o mnohdy naštvaného uživatele, který vyjadřuje svůj názor poměrně bezskrupulózním způsobem nebo se jedná o dobře maskovaného trolla. Tento uživatel se vyskytuje v komunikačních vláknech Adama Vojtěcha a Ivana Bartoše.

Posledním z možných trollů je @Ken98811667. Uživatel přispívá téměř každý den a v příspěvcích mnohdy uráží, mnohdy provokuje nebo podrývá jiné argumenty. Vyloženě vulgárních termínů občas užije, ale nelze je nelézt často. Věnuje se kontroverzním tématům a svůj názor dává razantně najevo. Vyjadřuje se téměř výhradně pouze v politických diskusních vláknech a nemá žádné sledující ani sledované. Je možné, že má uživatel zmíněný účet vytvořen právě za účelem ostré kritiky politické situace a v roušce anonymity se cítí bezpečně pro vyjadřování názorů, které by nebyl schopen pod svou pravou identitou vyslovit. Je ale také možné, že se jedná o prostého trolla, který chce účelně rozvířit vody v politických diskuzích. Možného trolla jsme detekovali u Adama Vojtěcha.

6.2.3. Uživatelé, co trolly jsou

Podařilo se s naprostou jistotou odhalit deset škodících uživatelů. Vezmeme-li v potaz fakt, že byly u každého z deseti politiků zkoumány pouze příspěvky v týdenním časovém rozpětí, odhalení deseti trollů lze považovat za pozitivní výsledek výzkumu a metodu detekce tedy za úspěšnou. Jedná se o uživatele @CapekCapekjiri, @CtiradMasin, @JiriMart, @JosefzizkaR, @Mengele85170837, @Mrdamtwitr, @PalzerRadim, @pavel_krupicka, @RazyHea1 a @SpendlikRoman.

Uživatel	Sledujících	Sledovaných	Počet příspěvků	Politici	Troll
CapekCapekjiri	10	79	4	Ovčáček	ANO
CtiradMasin	1	0	6	Ovčáček	ANO
JiriMart	8	21	5	Ovčáček	ANO
JosefzizkaR	308	81	7	Babiš, Vojtěch, Ovčáček, Hamáček	ANO
Mengele85170837	3	9	7	Vojtěch	ANO
Mrdamtwitr	13	114	2	Vojtěch, Hamáček	ANO
PalzerRadim	0	5	6	Vojtěch, Hamáček	ANO
pavel_krupicka	0	54	4	Ovčáček	ANO
RazyHea1	0	25	3	Vojtěch	ANO
SpendlikRoman	10	142	2	Hamáček	ANO

tab. 8 (viz. Příloha)

Uživatel @CapekCapekjiri si pomalu buduje síť lidí, které obsah jeho příspěvků zřejmě baví. V tuto chvíli ho sleduje deset účtů. Oproti doposud zmíněným účtům se @CapekCapekjiri těší velkému zájmu publika. Počet lajků u příspěvků je vysoký nejen vzhledem k počtu sledujících. Denně se vyjadřuje ve výhradně politických diskusních vláknech. Způsob vyjadřování je vulgární, nelogický, obsah příspěvků se často opakuje v různých vláknech, uráží a je prázdný. Za cíl si patrně neklade předat informaci nebo myšlenku, nýbrž pouze provokovat. V našem výzkumu jsme tohoto uživatele detekovali ve vlákně Jiřího Ovčáčka. Všemi atributy proto splňuje podmínky, které byly k detekci trolla vyhrazeny.

Jako další byl za trolla označen účet @CtiradMasin. Uživatel je s pouze jedním sledujícím takřka anonymní. Sám žádné účty nesleduje, přitom je poměrně aktivní v politických a dalších veřejných vláknech. To může naznačovat, že existuje na Twitteru pod svou pravou identitou a kdykoliv cítí potřebu se k něčemu vyjádřit nebo prostou příležitost k rýpání, přepne z osobního účtu na tento. Nepřispívá sice denně, ale přispívá hojně, často formou obrázků a memes. V příspěvcích se vysmívá, uráží, vyvolává konflikt. Některé příspěvky jsou konstruktivním vyjádřením názoru, ale mnoho jich není. Taktéž zde nenalezneme příliš mnoho vulgarismů. I

tento uživatel byl přítomen ve vlákně Jiřího Ovčáčka. Většina příspěvků odpovídá teoretickému popisu trolla.

Dalším detekovaným trollem je učet uživatele @JiriMart, který přispívá pravidelně každý den. Oproti předchozím trollům je méně aktivní, stále však pravidelně přispívá do výhradně politických debat nebo směrem k veřejně známým osobám. Uživatel v twitterové komunitě moc oblíbený není. Sleduje ho osm účtů a lajky u příspěvků skoro nenalézáme. Příspěvky jsou poměrně vulgární a provokativní. Hrubě se vyjadřuje skoro ke každé diskuzi, příspěvky na sebe nenavazují a nemají řád. V našem výzkumu se objevil ve vlákně Jiřího Ovčáčka.

Uživatel @JosefzizkaR přispívá velmi hojně každý den. Má vybudovanou širokou fanouškovskou základnu, sleduje ho tři sta osm účtů. Obsah příspěvků se ale zřejmě sledujícím nelíbí, v průměru má tento troll přibližně jeden lajk u příspěvku. Z dostupných informací lze usuzovat, že příliš reakcí jeho dráždivé příspěvky nevyvolávají, což je v rozporu s tím, čeho si většina trollů přeje docílit. Uživatel je aktivní na profilu mnoha politiků. Reaguje na příspěvky Adama Vojtěcha, Andreje Babiše nebo Aleny Schillerové. V našem výzkumu se objevil ve vlákně všech čtyřech politiků, u kterých jsme ve finální fázi přítomnost trolla s naprostou jistotou definovali, čímž se stává nejzastoupenějším ze všech nalezených trollů. Aktivní byl u Andreje Babiše, Jana Hamáčka, Jiřího Ovčáčka i Adama Vojtěcha.

Pátým odhaleným trollem se stal @Mengele85170837. Tento uživatel přispívá velmi často každý den, má tři sledující a u příspěvků většinou lajky nemá. Obsah jeho příspěvků není vulgární, ale je natolik provokativní, že se o trolla jedná. Aktivní je téměř výhradně v politických diskuzích, například u Adama Vojtěcha nebo Andreje Babiše. V našem výzkumu byl detekován v konverzačním vlákně Adama Vojtěcha. Dále se vyjadřuje v komunikačních vláknech českých médií jako CNN Prima nebo iDNES.cz. Svými kontroverzními příspěvky se snaží rozpoutat další debatu a dává jasně najevo svůj extremistický postoj vzhledem k citlivým tématům.

@Mrdamtwitr přispívá krátkými nestrukturovanými příspěvky. Aktivní je na denní bázi, sleduje ho třináct lidí, kteří ovšem příspěvky nelajkují. Tento troll, narozdíl od předchozího, je detekovatelný na první pohled. Zatímco předchozí uživatel provokuje své publikum skrytými ale dobře mířenými popichy, @Mrdamtwitr hraje zcela otevřenou hru. Veřejně nadává ostatním uživatelům, které ponižuje. Svůj postoj vyjadřuje také skrze emoji. Zcela zřetelně upřednostňuje kvantitu nad kvalitou příspěvků. Dané reakce v komunikačních vláknech nejen politiků, ale i dalších

známých i neznámých uživatelů neobsahují žádnou informační hodnotu, jedná se pouze o ponižovací posty s cílem ublížit nebo vyvolat reakci. V našem výzkumu byl odhalen u Jana Hamáčka a Adama Vojtěcha.

@PalzerRadim je aktivní skoro každý den a nemá žádné sledující. Občas je jeho příspěvek olajkován. Vyjadřuje se k politickým účtům Adama Vojtěcha, Jana Hamáčka nebo Aleny Schillerové. Ve svých příspěvcích uráží, ponižuje a vulgárně napadá další uživatele. Některé příspěvky jsou krátké a lrom nadávek nic dalšího neobsahují, jiné jsou delší, nicméně o informační hodnotu bohatší nejsou.

Uživatel @pavel_krupicka patří mezi anonymní trolly, kteří jsou detekovatelní až po podrobném zkoumání příspěvků. Nemá žádné sledující ani lajky a příspěvky jsou na první pohled běžné. Nejsou příliš krátké a plně vulgarit, nicméně i tak jsou obsahově prázdné. Jen několik málo z nich v sobě ukrývá informaci nebo názorový postoj, většinou se jedná o změť prázdných slov. Uživatel chce očividně provokovat další uživatele a tak reaguje na citlivá témata skutečnosti, zaujímá postoj proti čemukoliv, kolem čeho se zrovna debata odvíjí. Vulgární slova sice příliš neužívá, ale občas se v jeho příspěvcích objeví. Reaguje především na Jiřího Ovčáčka, v jehož vláknech byl detekován při výzkumu.

Účet @RazyHea1 byl zjevně založen za účelem trollingu. Uživatel není příliš aktivní, přihlásí se vždy zhruba jednou za týden, napíše řadu nevhodných příspěvků a vrací se opět po týdnu. Žádný účet ho nesleduje a nikdo jeho příspěvky nelajkuje. Tomuto trollovi zjevně nejde o pozornost nebo reakci komunity. Dle dostupných indicií se jedná o člověka, který nutně potřebuje svou zlost vyjádřit v online prostředí. Valná většina příspěvků obsahuje minimálně jedno vulgární oslovení, příspěvky nemají strukturu a působí jako zlostné výkřiky do tmy. Troll byl detekován ve vlákně Adama Vojtěcha

Posledním detekovaným trollem je @SpendlikRoman, kterého jsme detekovali ve vlákně Jana Hamáčka. Jedná se o uživatele s deseti sledujícími, jehož příspěvky jsou v poměru k počtu sledujících oblíbené a lajkované. @SpendlikRoman nepřispívá pravidelně, aktivní je zhruba jednou za dva dny. Mezi trolly byl zařazen vzhledem k obsahové analýze příspěvků. Některé příspěvky jsou zcela v normě, nicméně převažují příspěvky, které jsou typickým projevem antisociálního chování. Buď podrývají argumenty ostatních, provokují a ničí tím diskuzi, kterou odvádí jinam, nebo jsou pouhým shlukem vulgárních oslovení. Troll byl detekován na profilu Jana Hamáčka, na něhož uživatel často reaguje.

6.3. Shrnutí

Pro detekci samotných trollů bylo nejprve potřeba vytipovat taková komunikační vlákna, pro která je pravděpodobnost jejich výskytu vysoká. Vhodná komunikační vlákna jsme získávali z předem určeným politických účtů a vybírali je dle přesných kritérií. Příspěvky, na které komunikační vlákna byla napojena, byly přidány v týdenním časovém rozpětí, měly minimální počet padesáti lajků a pěti retweetů. Každé komunikační vlákno disponuje unikátním ID komunikace.

Za použití ID komunikace byla jednotlivá vlákna, včetně všech vnořených odpovědí, dále zkoumána. Výzkum pracoval pouze s příspěvky, které postojová analýza vyhodnotila za negativní. Růžovou barvou došlo k označení všech příspěvků, které jsme považovali za potenciální trolly. Označeny byly příspěvky, které obsahovaly vulgární výrazy a všechny příspěvky od stejného autora, pokud byl jejich počet v jednom komunikačním vlákně vyšší než dva nebo v různých vláknech vyšší než jeden. Nejvíce komunikačních vláken bylo zkoumáno u Jiřího Ovčáčka, nejméně u Jana Hamáčka. Nejvíce příspěvků obsahovala komunikační vlákna Adama Vojtěcha, nejméně jsme jich nalezli u Jana Hamáčka. Sledující Adama Vojtěcha jsou nejvíce aktivní, každý jeho příspěvek má v komunikačním vlákně šedesát tři odpovědí. Nejméně aktivní jsou sledující Jana Hamáčka a Mariana Jurečky, kde se průměrný počet odpovědí pohybuje kolem devíti. Nejvíce potenciálních trollů bylo označeno na profilech Jiřího Ovčáčka a Adama Vojtěcha s počtem dvě stě dva, nejméně potenciálních trollů bylo nalezeno u Jana Hamáčka s počtem šest. V poměru počtu zkoumaných příspěvků a příspěvků označených za napsané potenciálními trolly, byl největší výskyt trollů u Jiřího Ovčáčka, kde bylo označeno přibližně 45% příspěvků. Nejmenší koncentraci trollů pozorujeme u Markéty Pekarové Adamové, kde byl výskyt trollů pouze v 7% zkoumaných příspěvků.

Označené příspěvky následně prošly dalším třízením. Odfiltrovaly se příspěvky, jejichž autor přispěl pouze jednou do všech zkoumaných komunikačních vláken na profilech deseti různých politických účtů a byl zřejmě označen pro použití vulgárního výrazu. Vulgaritu ale ve zkoumaném použil pouze jednou, proto se nejspíše jednalo o výjimku. Taktéž se odflitrovaly příspěvky, jejichž autor byl označen kvůli svém vysoké aktivitě v mnoha komunikačních vláknech, nicméně po obsahové analýze příspěvků vyšlo najevo, že komunikuje slušně, strukturovaně a do diskuze

přispívá validními a srozumitelnými argumenty. Tito autoři byly obvykle velmi zastoupení napříč všemi vlákny. Počet označených příspěvků slušného přispěvatele se obvykle pohyboval kolem deseti, ale spíše číslo deset přesahoval. Naopak přispěvatelé, kteří po tomto kroku zůstali označeni za trolly, přispěli v průměru přibližně pětkrát. Při tomto vyřazování pro výzkum nevhodných příspěvků došlo k potvrzení teoretického popisu trolla. Troll totiž není přímo účasten diskuzí a hádek, jeho cílem není prosadit a vysvětlit svůj postoj, jeho cílem bývá rozpoutání války v komunikačních vláknech, nad kterou se následně v tichosti baví.

V konverzačních vláknech u Petra Fialy, Miroslava Kalouska ani Markéty Pekarové Adamové nezůstaly po výše zmíněném filtrování označeny žádné příspěvky. Nejvíce stále označených příspěvků bylo nalezeno u Jiřího Ovčáčka a Adama Vojtěcha. V této fázi výzkumu došlo na zpětné ověření trollů pomocí síťové a obsahové analýzy. Někteří autoři byli ze seznamu vyloučeni, nebo označeni za možné trolly. To vysvětluje, proč například Tomio Okamura měl ve svých vláknech označeno pět příspěvků po filtrování trollů, ale ve finálním fázi celého výzkumu nebyl uveden ani jeden uživatel-troll, který by do komunikačních vláken Okamury přispíval.

Trollové byli detekování pouze u Andreje Babiše, Jana Hamáčka, Jiřího Ovčáčka a Adama Vojtěcha. U Andreje Babiše byl detekován jeden troll. U Jana Hamáčka se potvrdili při zpětném ověření všichni čtyři označení trollové. U Jiřího Ovčáčka i Adama Vojtěcha bylo u každého detekováno pět trollů. Vzhledem k faktu, že Jiří Ovčáček a Adam Vojtěch se průběžně usazovali na prvních příčkách co se počtu zkoumaných vláken nebo příspěvků týče, lze usuzovat, že vysoký počet detekovaných trollů s tím souvisí. Ve vyšším počtu příspěvků je vyšší pravděpodobnost výskytu a detekce trolla. Účty kontroverzních politiků Jiřího Ovčáčka a Adama Vojtěcha se zdají být správným prostředím pro působení internetových trollů.

Nejvíce detekovaných nevhodných příspěvků napsali uživatelé @JosefzizkaR a @Mengele85170837. První zmíněný má zároveň nejvíce sledujících a detekovali jsme ho u čtyřech různých politiků, což je nejvyšší počet výzkumu. Zbývající trolly jsme obvykle detekovali na jednom či dvou politických profilech.

Obsahová analýza příspěvků je pro úspěšnou detekci trollů zásadně nejdůležitější. Pouze z obsahu textu lze odvodit, zda uživatel jedná mimo hranice slušného chování a zda tak činí systematicky, nikoliv náhodně. Obsahová analýza

však není jediným vodítkem, které trolla pomůže rozeznat od ostatních. Kromě kvality příspěvků jsme v předchozích kapitolách zmínili taktéž kvantitu. Naprosto všichni detekování trollové se vyznačují častou aktivitou na Twitteru. Většina z nich je aktivní každý den, ostatní téměř každý. Zároveň nemluvíme o jednom příspěvku denně. Detekování trollové přispívají ve velkém množství krátkými příspěvky, které na sebe mnohdy nenavazují a nemají mezi sebou žádnou spojitost. Dalším faktorem, který je možné při detekci zohlednit, je trollova síť sledovaných účtů. Všichni trollové nepřispívají pouze jednomu náhodně vybranému politikovi, ale cíleně si vybírají skupinu politických či jinak veřejně známých osob, které následně urážejí a ponižují. Výzkum naopak nepotvrdil hypotézu, která tvrdí, že troll se snaží najít svém místo v komunitě. Dle této teorie mezi trolly nepatří anonymní profily, nýbrž profily se silnou fanouškovskou základnou, která adekvátně na příspěvky reaguje. V našem seznamu se nachází trollové, které mají aktivní publikum, ale také trollové, jejich profil je zcela anonymní, nemají žádné sledující ani lajky. Pro detekci jsou tato čísla nepodstatná.

Při pokusu o automatickou detekci trollů na českém Twitteru, který by se opíral o výsledky tohoto výzkumu, by bylo vhodné zaměřit se zejména na nadměrné užívání vulgárních a urážlivých slov a na aktivitu, která je ve větším časovém úseku nadprůměrně vysoká co do počtu příspěvků za den a zároveň pravidelná a kontinuální. Dále by bylo vhodné zaměřit se na síťovou analýzu uživatelského účtu, pomocí které lze zjistit provázanost s dalšími twitterovými účty, tedy zjistit, kam troll přispívá a zda-li má vyhraněný zájem o určitou komunitu. Přidaným faktorem, který tento výzkum nezohlednil, by mohla být taktéž detekce emoji, především domácích zvířat, exkrementu a dalších. Tyto emoji se ve zkoumaných příspěvcích trollů vyskytovaly nejvíce. Dalším krokem pro úspěšnou automatickou detekci trollů by mohlo být již zmíněné přezkoumání reakce komunity na daný příspěvek, respektive co se odehrává v komunikačním vlákně po přidání příspěvku s fenoménem trollingu. Obvykle se v reakcích o pár příspěvků níže spustí flaming.

.

Závěr

Tato diplomová práce s názvem “Automatická detekce trollů na českém Twitteru” je rozdělena na dvě části - teoretickou a praktickou. V teoretické části se zabývá sociálními médii, respektive Twitterem. Zmiňuje jeho vznik a historii, popisuje uspořádání a vysvětluje možnosti sběru a těžby dat, která jsou využívána pro zkoumání a deskripci společnosti. Dále teoretická část zmiňuje popis soudobé společnosti. Zmíněný popis obsahuje rovněž využití Twitteru jako platformy pro profesní komunikaci zahraničních i českých politiků. V další části se práce zabývá teoretickým popisem internetového trolla. Zmiňuje jasně definované charakteristiky jeho chování v online prostředí, které dále využíval výzkum v praktické části. Kromě čistě popisných pasáží fenoménu trollingu obsahuje teoretická část úvahu nad napáchanými škodami na ostatních uživatelích společně s prakticky ověřenými radami jak při setkání s trollem reagovat. Na konci teoretické části jsou vysvětleny různé metody detekce internetového trolla z nichž většina byla použita v praktickém výzkumu.

Praktická část diplomové práce se zabývá detekcí trollů na twitterových účtech českých politiků. Detekce začíná výběrem vhodných příspěvků ke zkoumání. Komunikační vlákna vybraných příspěvků jsou dále posuzována pomocí postojové analýzy. Příspěvky do komunikačních vláken, které jsou negativně ohodnoceny postojovou analýzou jsou posuzovány na základě obsahové analýzy. Obsahuje-li text vulgární slovo, je označen za příspěvek potenciálního trolla. Všechny takto označené příspěvky prošly druhou fází filtrace, přičemž bylo vybráno sedmnáct uživatelů, které výzkum zpětně ověřil. Při zpětném ověření došlo k celkové analýze uživatelské profilu. Byla zkoumaná celková aktivita trolla a to jak z hlediska obsahu psaných textů, tak z hlediska frekvence přispívání. Výzkum se taktéž zaměřil na to, do jakých diskuzí se troll zapojil a které tématické okruhy ho zajímaly. Došlo k analyzování příspěvků po obsahové i vizuální stránce, k analyzování komunikačních vláken i k analyzování vztahů v komunitě. Z toho vyplývá, že výzkum pracoval s téměř všemi detekovacími metodami uvedenými v teoretické části. Nezohlednil pouze chování ostatních uživatelů v reakci na trolla.

Praktická část této diplomové práce detekuje trolly mechanicky. Způsob detekce je nicméně zvolen tak, aby jej bylo nejen možné, ale rovněž poměrně snadné zautomatizovat. Úspěšná detekce trollů, která se opírá o teoretické ukotvení,

vypovídá o úspěšnosti výzkumu, který tak potvrzuje předem dané hypotézy o fenoménu trollingu v online prostředí. Nalezení deseti trollů nejen ověřuje teoretický popis trolla v první části práce, ale také prokazuje, že postup zvolený ve druhé části práci je vhodným nástroj k detekci. Automatizace zmíněného postupu je tudíž smysluplná a proveditelná.

Cílem této diplomové práce bylo prozkoumat možnosti automatické detekce trollů v prostředí českého Twitteru pomocí síťové analýzy a strojového zpracování jazyka a následně otestovat účinnost zmíněných metod. Metody síťové analýzy využíval výzkum k ověření systematického vystupování trolla ve veřejných kruzích komunikace. Metody strojového zpracování jazyka jsou vhodné pro obsahovou analýzu příspěvků. Praktická část testováním metod dokázala jejich účinnost při detekci internetových trollů. S ohledem na výše zmíněné lze tvrdit, že cíle diplomové práce byly úspěšně naplněny.

.

Příloha

V přiloženém odkaze do online aplikace Tabulky Google lze nalézt tabulky uspořádané do čtyřech složek.

Složka *Tweety* obsahuje tabulky, které byly používány ve druhé fázi výzkumu. Každý z vybraných politických účtů disponuje vlastní tabulkou. V tabulkách jsou seznamy příspěvků, které zbyly po filtrování dle datumu a počtu lajků i retweetů. Příspěvků, jejichž komunikační vlákna se podrobně dále zkoumala.

Složka *Vlákna* obsahuje, jak již název vypovídá, příspěvky do vybraných komunikačních vláken. Opět má každý politický účet svojí tabulku s kompletním výčtem příspěvků do všech předem vybraných komunikačních vláken z jeho účtu. V tomto kroku proběhla filtrace potenciálních trollů dle četnosti přispívání autora a dle obsahu textů. Růžově jsou označeni všechny příspěvky, jejichž obsah je vulgární nebo jejichž autor přispěl do komunikačního vlákna vícenásobně. Složka dále obsahuje tabulku *Všechna vlákna*, kde se nachází seznam naprosto všech růžově označených příspěvků. Růžové příspěvky prošly druhým kolem třízení. V tomto kole se ze seznamu odstranili autoři, kteří přispěli jednorázově a autoři, kteří se nevyjadřovali vulgárně. Dále byli odstraněni autoři, kteří svým jednáním zjevně neprovokovali ani nepobuřovali. Zbylé příspěvky od sedmnácti různých autorů se nachází ve složce *Výběr trollů final.*

Složka *Trollové* obsahuje vždy jednu tabulku pro každého ze sedmnácti detekovaných trollů. V tabulkách nalezneme všechny tweety a odpovědi, kterými daný troll na svém twitterovém účtu. Tyto příspěvky výzkum zpětně ověřoval ve třetí fázi.

Poslední složka *Tabulky v DP* obsahuje všechny uvedené tabulky ve vlastním textu této práce.

DP Míková 2021

(https://drive.google.com/drive/folders/1-UADQjGzqW8zp1dInhSTj0nN3jeK0olM)

Bibliografie

ANGIANI, Giulio, Stefano CAGNONI, Natalia CHUZHIKOVA, Paolo FORNACCIARI, Monica MORDONINI and Michele TOMAIUOLO, 2016. Flat and Hierarchical Classifiers for Detecting Emotion in Tweets. *AI*IA 2016 Advances in Artificial Intelligence Lecture Notes in Computer Science* [online]. pp. 51–64. Retrieved z: doi:10.1007/978-3-319-49130-1_5

BERNSTEIN, Michael, Justin CHENG, Cristian DANESCU-NICULESCU-MIZIL and Jure LESKOVEC, 2017. Anyone Can Become a Troll. *American Scientist* [online]. vol. 105, no. 3, p. 152. Retrieved z: doi:10.1511/2017.105.3.152

BISHOP, Jonathan, 2014. Representations of 'trolls' in mass media communication: a review of media-texts and moral panics relating to 'internet trolling'. *International Journal of Web Based Communities* [online]. vol. 10, no. 1, p. 7. Retrieved z: doi:10.1504/ijwbc.2014.058384

BÖHMOVÁ, Lucie, Antonín Pavlíček, 2013. Personalistika a budoucnost sociálních sítí v ČR. *Scientific Papers of the University of Pardubice* [online]. Series D, Faculty of Economics and Administration. 27 (2/2013).

BOULIANNE, Shelley, 2016. Online news, civic awareness, and engagement in civic and political life. *New Media & Society* [online]. vol. 18, no. 9, pp. 1840–1856. Retrieved z: doi:10.1177/1461444815616222

BUCKELS, E.e., P.d. TRAPNELL and D.l. PAULHUS, 2014. Trolls Just Want to Have Fun. *PsycEXTRA Dataset* [online]. Retrieved z: doi:10.1037/e520722015-006

ANON., 2013. Closing the gap? Twitter as an instrument for connected representation TODD G RAHAM , MARC E L BROER SMA A ND KARIN HAZE LHO F F. *The Media, Political Participation and Empowerment* [online]. pp. 87–104. Retrieved z: doi:10.4324/9780203381113-15

COLES, Bryn Alexander and Melanie WEST, 2016. Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior* [online]. vol. 60, pp. 233–244. Retrieved z: doi:10.1016/j.chb.2016.02.070

DE-LA-PEÑA-SORDO, Jorge, Igor SANTOS, Iker PASTOR-LÓPEZ and Pablo G. BRINGAS, 2013. Filtering Trolling Comments through Collective Classification. *Network and System Security Lecture Notes in Computer Science* [online]. pp. 707–713. Retrieved z: doi:10.1007/978-3-642-38631-2_60

FORNACCIARI, Paolo, Monica MORDONINI, Agostino POGGI, Laura SANI and Michele TOMAIUOLO, 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* [online]. vol. 89, pp. 258–268. Retrieved z: doi:10.1016/j.chb.2018.08.008

HARDAKER, Claire, 2010. Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture* [online]. vol. 6, no. 2. Retrieved z: doi:10.1515/jplr.2010.011

Cheng, J.; Danescu-Niculescu-Mizil, C.; Leskovec, J. Antisocial behavior in online discussion communities. In Proceedings of the Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015; pp. 61–70.

JUNGHERR, Andreas, 2014. Twitter in Politics: A Comprehensive Literature Review. *SSRN Electronic Journal* [online]. Retrieved z: doi:10.2139/ssrn.2402443

KARLSEN, Rune, 2011. A Platform for Individualized Campaigning? Social Media

and Parliamentary Candidates in the 2009 Norwegian Election Campaign. *Policy & Internet* [online]. vol. 3, no. 4, pp. 1–25. Retrieved z: doi:10.2202/1944-2866.1137

KUMAR, Srijan, Francesca SPEZZANO and V.s. SUBRAHMANIAN, 2014. Accurately detecting trolls in Slashdot Zoo via decluttering. *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* [online]. Retrieved z: doi:10.1109/asonam.2014.6921581

LARSSON, Anders Olof, 2014. Online, all the time? A quantitative assessment of the permanent campaign on Facebook. *New Media & Society* [online]. vol. 18, no. 2, pp. 274–292. Retrieved z: doi:10.1177/1461444814538798

MACHOVÁ, Kristína and Dominik KOLESÁR, 2019. Recognition of Antisocial Behavior in Online Discussions. *Advances in Intelligent Systems and Computing Information Systems Architecture and Technology: Proceedings of 40th Anniversary International Conference on Information Systems Architecture and Technology – ISAT 2019* [online]. pp. 253–262. Retrieved z: doi:10.1007/978-3-030-30604-5_23

MACKOVÁ, Alena, 2017. *Nová média v politické komunikaci: politici, občané a online sociální sítě*. B.m.: Masarykova univerzita, Fakulta sociálních studií, Mezinárodní politologický ústav.

MIHAYLOV, Todor, Georgi GEORGIEV and Preslav NAKOV, 2015. Finding Opinion Manipulation Trolls in News Community Forums. *Proceedings of the Nineteenth Conference on Computational Natural Language Learning* [online]. Retrieved z: doi:10.18653/v1/k15-1032

SAMORY, Mattia and Enoch PESERICO, 2017. Sizing Up the Troll. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* [online]. Retrieved z: doi:10.1145/3025453.3026007

STRANDBERG, Kim, 2013. A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections. *New Media & Society* [online]. vol. 15, no. 8, pp. 1329–1347. Retrieved z: doi:10.1177/1461444812470612

TAO, Ke, Fabian ABEL, Claudia HAUFF and Geert-Jan HOUBEN, 2012. Twinder: A Search Engine for Twitter Streams. *Lecture Notes in Computer Science Web Engineering* [online]. pp. 153–168. Retrieved z: doi:10.1007/978-3-642-31753-8_11

TEEVAN, Jaime, Daniel RAMAGE and Merredith Ringel MORRIS, 2011. #TwitterSearch. *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11* [online]. Retrieved z: doi:10.1145/1935826.1935842

THORING, Anne, 2011. Corporate Tweeting: Analysing the Use of Twitter as a Marketing Tool by UK Trade Publishers. *Publishing Research Quarterly* [online]. vol. 27, no. 2, pp. 141–158. Retrieved z: doi:10.1007/s12109-011-9214-7

TOMAIUOLO, Michele, Gianfranco LOMBARDO, Monica MORDONINI, Stefano CAGNONI and Agostino POGGI, 2020. A Survey on Troll Detection. *Future Internet* [online]. vol. 12, no. 2, p. 31. Retrieved z: doi:10.3390/fi12020031

WELLER, Katrin, Axel BRUNS, Jean BURGESS, Merja MAHRT and Cornelius PUSCHMANN, 2014. *Twitter and society*. B.m.: Peter Lang.

YOUNUS, Arjumand, M. Atif QURESHI, Muhammad SAEED, Nasir TOUHEED, Colm O'RIORDAN and Gabriella PASI, 2014. Election trolling. *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*

[online]. Retrieved z: doi:10.1145/2567948.2577352
ZHANG, Chao Michael and Vern PAXSON, 2011. Detecting and Analyzing Automated Activity on Twitter. *Passive and Active Measurement Lecture Notes in Computer Science* [online]. pp. 102–111. Retrieved z: doi:10.1007/978-3-642-19260-9_11