

**Jihočeská univerzita v Českých Budějovicích  
Přírodovědecká fakulta**

# **Bakalářská práce**

**2018**

**Vojtěch Vacek**



**Jihočeská univerzita v Českých Budějovicích  
Přírodovědecká fakulta**

**Plugin do prohlížeče pro znemožnění  
profilování uživatele podle historie  
vyhledávání v Google**

Bakalářská práce

**Vojtěch Vacek**

Školitel: Ing Petr Břehovský

*České Budějovice 2018*

Jihočeská univerzita v Českých Budějovicích  
Přírodovědecká fakulta

**ZADÁVACÍ PROTOKOL BAKALÁŘSKÉ PRÁCE**

**Student:** Vojtěch Vacek

**Obor – zaměření studia:** Aplikovaná informatika - web a multimédia

**Katedra/ústav:** Ústav Aplikované Informatiky

**Školitel:** Ing. Petr Břehovský

**Garant z PŘF:** RNDr. Libor Dostálek

**Téma bakalářské práce:** Plugin do prohlížeče pro znemožnění profilování uživatele podle historie vyhledávání v Google.

Cíle práce:

- 1) Provést analýzu dostupných pluginů zabývajících se podobnou problematikou.
- 2) Vytvořit funkční plugin.
- 3) Otestovat funkčnost vzniklého pluginu.

Základní doporučená literatura:

- [1] CHAFFER, Jonathan a Karl SWEDBERG. *Mistrovství v jQuery: [kompletní průvodce vývojáře]*. 1. vyd. Brno: Computer Press, 2013. Mistrovství. ISBN 978-80-251-4103-8.
- [2] HOGAN, Brian P. *HTML5 a CSS3: výukový kurz webového vývojáře*. Vyd. 1. Brno: Computer Press, 2011. ISBN 978-80-251-3576-1.
- [3] SKOČEK, Jakub. *Analýza digitální identity na základě reálných uživatelských dat*. In: *Knihovna Akademie věd ČR [online]*. Praha: Knihovna Akademie věd ČR, 2016 [cit. 2016-11-03]. Dostupné z: [https://www.lib.cas.cz/casopis-informace/digitalni-identita/#\\_ftnref1](https://www.lib.cas.cz/casopis-informace/digitalni-identita/#_ftnref1)

Vedoucí práce: Ing. Petr Břehovský

podpis: 

Vedoucí ústavu: RNDr. Libor Dostálek

podpis: 

V Českých Budějovicích dne 20. 12. 2016 Podpis studenta: 

## **Bibliografické údaje**

VACEK V., 2018: Plugin do prohlížeče pro znemožnění profilování uživatele podle historie vyhledávání v Google. [Browser plugin to prevent user profiling by Google search history Bc. Thesis, in Czech.] – 27 p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic. Vedoucí práce Ing. Petr Břehovský.

## **Anotace**

Tato bakalářská práce přibližuje problematiku masivního sběru informací zanechaných uživateli ve formě digitálních stop a následném profilování. V teoretické části jsou vysvětleny pojmy digitální stopy a digitální identita. Dále se zabývá analýzou nasbíraných dat a možnou ochranou proti samotnému profilování. Praktická část se věnuje tvorbě samotného pluginu a testování s důkazem funkčnosti vzniklého rozšíření.

## **Klíčová slova**

digitální stopy, Google Chrome, javascript, plugin

## **Annotation**

This bachelor thesis describes issues of massive collection of information which has been left by users in the form of digital footprints and consecutive profiling. In the theoretical part terms – digital footprints and digital identity are explained. Next this part deals with analysis of collected data and potential protection against profiling itself. The practical part pays attention to creating of plugin itself and to testing on the basis of the proof connected to functionality of arisen extension.

## **Keywords**

digital footprints, Google Chrome, javascript, plugin

# Prohlášení

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne \_\_\_\_\_

\_\_\_\_\_ podpis

## **Poděkování**

Rád bych poděkoval svému vedoucímu práce panu Ing. Petru Břehovskému za rady a podmětné připomínky během psaní bakalářské práce.

# Obsah

1	Úvod .....	1
2	Digitální identita .....	2
2.1	OpenID .....	2
3	Digitální stopy .....	3
3.1	Aktivní digitální stopy .....	3
3.2	Pasivní digitální stopy .....	3
3.3	Možné hrozby .....	4
3.4	Správa digitálních stop .....	4
4	Analýza digitální identity .....	6
4.1	Ukládané informace .....	6
4.1.1	Co děláme .....	6
4.1.2	Co vytváříme .....	7
4.1.3	Kdo jsme .....	7
4.2	Analýza reálných uživatelských dat .....	7
5	Technologie znemožňující sledování .....	9
5.1	IXquick .....	9
5.1.1	StartPage .....	9
5.2	DuckDuckGo .....	9
5.3	Ghostery .....	10
5.4	AdBlock .....	10
5.5	Noiszy .....	11
5.6	TrackMeNot .....	11
6	Praktická část .....	13
6.1	Popis rozšíření .....	13
6.2	Architektura pluginu .....	15
6.3	Uživatelské rozhraní .....	15
6.4	Background komponenta .....	16
6.5	Content script .....	19
6.5.1	Získání adresy stránky .....	20
6.5.2	Získání náhodného slova .....	20



6.6	Manifest .....	21
6.7	Testování funkčnosti .....	23
6.7.1	Google Ads Settings .....	23
6.7.2	Výsledek .....	24
7	Závěr.....	26
8	Seznam literatury .....	28
9	Seznam obrázků .....	31
10	Seznam příloh.....	32

# 1 Úvod

V roce 2013 zveřejnil Edward Snowden v deníku The Guardian [1] informaci o masivním sledování lidí bezpečnostními agenturami. Hlavním zdrojem dat pro sledování a analyzování jednotlivých uživatelů jsou největší poskytovatelé služeb (Microsoft, Yahoo, Google, Apple, Facebook a další), kteří předávají informace bezpečnostním agenturám. V rozhovoru se Snowden zmiňuje o projektu PRISM. Tento projekt sbírá data od jednotlivých poskytovatelů služeb, zachytává telefonní hovory, analyzuje a následně ukládá data. Díky tomuto projektu získává národní bezpečnostní agentura mocný nástroj v podobě databáze uchovávající informace profilů jednotlivých uživatelů.

Mluvčí jednotlivých poskytovatelů služeb spolupráci s vládními organizacemi popírají. Google přiznává vládní žádost o instalaci tzv. zadních vrátek do svých produktů. Tu ale podle svých slov odmítá. Microsoft se ve své reklamní kampani reprezentoval sloganem: „Vaše soukromí, naše priorita“. Byl ale první, který se zapojil do plošného sběru informací o uživateli. Pro projekt PRISM začal pracovat v roce 2007. Mluvčí společnosti Apple tvrdí, že firma o projektu PRISM neslyšela. Podle Snowdena ale poskytují informace o uživateli od roku 2012 [1].

Po zveřejnění Snowdenova rozhovoru se otázka soukromí uživatelů stala jedním z nejdiskutovanějších témat. Lidé si začali uvědomovat nebezpečí sběru dat a následného profilování. Bylo potřeba vymyslet, jak zabezpečit vyhledávání na internetu a znemožnit (nebo alespoň omezit) únik citlivých informací. Z tohoto důvodu jsem se rozhodl vytvořit plugin do webového prohlížeče Google Chrome.

## 2 Digitální identita

Digitální identita nahrazuje naši fyzickou identitu v kybernetickém prostředí. Stejně jako se v reálném světě identifikujeme občanským průkazem, tak i v internetovém prostoru je potřeba prokázat naši totožnost.

Získáváme tedy dvě skupiny identit, reálná identita a digitální identita. Obě skupiny určují konkrétního jedince součtem jednotlivých charakteristik. Příklady reálné identity (např. datum a místo narození, pohlaví, velikost bot) zužují globální skupinu lidí na malou skupinku, v níž jakákoli další charakteristika povede k přesnému určení jedince. Některé informace se nikdy nezmění (krevní skupina, otisky prstů apod.) a jiné jsou v čase proměnné (výška, váha nebo třeba barva vlasů) [2].

Digitální identita neslouží pouze k důvěrné autentizaci. Navštívené webové stránky chtějí znát identitu uživatele za účelem cílené reklamy. IP adresa, stejně jako například místo narození, určuje pouze skupinu možných lidí z důvodu proxy serverů a sdílených připojení. Proto až chování uživatele na internetu odhaluje skutečnou identitu.

### 2.1 OpenID

Je standard provádějící autentizaci pomocí třetí strany. Při přihlašování na webový server jsme automaticky přesměrováni na přihlašovací stránku OpenID, kde provedeme autentizaci. OpenID pak zpraví webový server o tom, že za digitální identitou je skutečný člověk prokazující se danou identitou.

OpenID je tedy správcem digitální identity. Veškeré informace vedoucí k určení dané fyzické osobě jsou v jedné množině a uživatel může libovolně nastavit, které informace budou webovému serveru sděleny a které ne.

Použitím této technologie je odstraněno riziko nutné registrace na „pochybné“ servery. Heslo, jímž se uživatel přihlašuje ke službě, zná pouze správce digitální identity (OpenId). Ten pak potvrdí identitu uživatele webovému serveru, aniž by server potřeboval znát přihlašovací údaje.[3].

## 3 Digitální stopy

Digitální stopy jsou důsledkem běžné aktivity na internetu a jsou určeny souhrnem všech informací, které za sebou uživatel zanechá. Procházení zboží na e-shopech, komentování či postování příspěvků na sociálních sítích, využívání komunikačních aplikací (Skype, WhatsApp, e-mail) nebo jen brouzdání na webu. Pokaždé, když jsou tyto služby používány, zvětšuje se objem zanechaných dat (digitální stopy) v kyberprostoru a zpřesňuje se náš profil.

Pro společnosti, zabývající se sběrem dat o aktivitě uživatelů, je analyzování a vytváření profilů z online historie jednotlivých uživatelů důležité zejména pro zacílení reklamy. Ale informace se nemusí uchovávat pouze za účelem cílené reklamy. Osoby provozující veřejně dostupnou službu elektronických komunikací jsou podle § 97 zákona č. 127/2005 Sb. povinny uchovávat údaje, které jsou zpracovány prostřednictvím jejich veřejně dostupných sítí. Podle americké neziskové organizace Pew Research Center se zanechané stopy dají rozdělit do dvou základních skupin, aktivní digitální stopy a pasivní digitální stopy [4].

### 3.1 Aktivní digitální stopy

Aktivní digitální stopy jsou definovány jako osobní údaje, které uživatel vědomě zpřístupnil. Jedná se zde o sdílení na sociálních sítích. Tím jsou myšleny fotografie, videa nebo statusy uživatele. Dále se může jednat o příspěvky k otevřeným diskusím, on-line komentáře, weby vytvořené uživatelem, blogy apod.

### 3.2 Pasivní digitální stopy

Pasivní digitální stopy jsou podle organizace definovány jako osobní údaje, které byly zpřístupněny bez vědomí uživatele.

Při surfování na internetu za sebou uživatel zanechává právě tuto pasivní stopu, aniž by si uvědomil, že i tyto pasivní digitální stopy vedou k identifikaci uživatele. Mezi tyto stopy se dají zařadit navštívené webové stránky a doba strávená na nich, IP adresa uživatele (a tím zjištění přibližné polohy), poskytovatel připojení, otisk prohlížeče

a další. Pasivní, stejně jako aktivní, stopy prozrazují na uživatele určité informace. Tyto informace pak vedou k velmi přesnému určení profilu uživatele.

### 3.3 Možné hrozby

Neopatrný pohyb na internetu vede ke zvýšení rizika zneužití digitální stopy. Pokud potencionální útočník nasbírá dostatek informací o určité osobě, může například založit falešné účty na sociálních sítích nebo různé blogy a tím se dopustit krádeže identity. Prostřednictvím tohoto účtu se může útočník za „okradeného“ uživatele vydávat a osoby, s nimiž bude zloděj identit komunikovat, nepoznají, že na druhé straně není předpokládaná osoba. Do blogů pak může útočník vkládat obsah s extremistickou tematikou, poplašné nebo řetězcové zprávy. V krajní mezi vede krádež identity ke ztátě pracovního poměru nebo financí.

Protože náš pohyb ve webovém prostředí není anonymní a stopy, které po sobě každý zanechá, mnohé prozrazují, hrozí riziko omezení soukromí. Do tohoto omezení může spadat i vyzrazení určitých osobních údajů a postojů, například politické názory, sexuální orientace, postoj k určitému problému (adopce homosexuálními páry, eutanazie, migrační krize apod.) nebo intimní zdravotní problémy. Všechny tyto informace (kde se registujeme, co sdílíme, fotky které nahráváme) o nás prozradí více než životopis, a proto je běžně používají personalisté.

### 3.4 Správa digitálních stop

Pokud chceme využívat moderní komunikační technologie, musíme počítat se vznikem digitálních stop. Pokusem o snížení tvorby těchto stop je údaj v hlavičce http protokolu Do Not Track. Uživatelé tímto údajem informují webový server o tom, že nechtějí být sledováni. Provozovatelům webů ovšem tento požadavek nijak nebrání v pokračování sledování [5]. Další způsob, jak snížit nebo alespoň omezit šíření digitálních stop, je mazání cookies souborů. Nejefektivnější metoda zabráňující tvorbě aktivních digitálních stop je uvažování, které informace může uživatel sdělit, aniž by to ohrozilo jeho soukromí, a které ne.

Na webu je celá řada vyhledávacích enginů zaměřených na vyhledávání lidí (Spokeo, PeekYou, FindFace, ZoomInfo a další). Jejich úkolem je sbírat aktivní

digitální stopy, které uživatel zanechává napříč celým internetem, a zobrazit je na jednom místě. Kromě využití těchto vyhledávačů personalisty pro vyhledání podrobných informací o novém zaměstnanci mohou být použity pro kontrolu vlastní digitální stopy. Pro rozšířenou kontrolu si může uživatel vyžádat všechny informace, které o něm určitá společnost vlastní (Facebook information dowlands, Google dashboard nebo Twitter request your archive).

# 4 Analýza digitální identity

Analýza digitální identity zkoumá osobní data, jež uživatel vědomě či nevědomě poskytne při svém pohybu na internetu. Z dostupných dat lze sestavit přesný profil uživatele, který se za danou identitou vyskytuje. Nepromítají se zde údaje, které lze vyčíst z průkazu totožnosti, ale data, která tvoří zájmy, seznam přátel, dovednosti a další.

Společnost Google disponuje širokou škálou aplikací, které denně používají miliony lidí na celém světě. Veškeré informace, které uživatel společnosti poskytne, jsou ukládány a následně analyzovány [6]. V rámci získání právního krytí popisuje společnost Google jaké informace shromažďuje a jak s nimi nakládá. Uživatelé s tímto nakládáním souhlasí při potvrzení upozornění: Souhlasím s obchodními podmínkami.

Google svůj sběr dat kompenzuje zlepšením služeb. Například automatické doplňování či automatická oprava ve vyhledávacím řádku, předvyplněné formuláře a další. To vše jsou služby, které běžný uživatel uvítá, aniž by si uvědomil sběr informací o něm samotném, který probíhá na pozadí těchto služeb.

## 4.1 Ukládané informace

### 4.1.1 Co děláme

Při používání různých služeb a aplikací poskytujeme společnosti Google veškeré informace, které jsou s danou činností spojené. Například pokud vyhledáváme pomocí Google vyhledávače, sdělujeme, co vyhledáváme. Při kliknutí na odkaz je uložena navštívená webová stránka, přehraná videa na portálu YouTube nebo reklamy, na které jsme klikli. To je důležité pro následné cílení reklamy [7].

Zařízení, na kterém služby využíváme, odesílá další informace. Přesnou polohu v případě použití zařízení s GPS modulem nebo přibližnou skrze zaměření IP adresy. Zařízení kromě polohy také odešle informace o operačním systému a jeho verzi nebo jaký webový prohlížeč a jaké verze byl použit.

## 4.1.2 Co vytváříme

V případě přihlášení ke službám pomocí Google účtu je ukládán obsah všech prací vytvořených v Google aplikacích. Ukládány jsou přijaté a odeslané e-maily ve službě Gmail, upravované či prohlížené fotografie pomocí aplikace Google Fotky nebo programu Picasa. Ten mimo jiné disponuje i technologií rozpoznávání obličejů, čímž je umožněno spojování fotek z jiných zdrojů. Ve vlastnictví má společnost Google i dokumenty uložené na virtuálním disku Google Drive [7].

## 4.1.3 Kdo jsme

Již při vytváření nových uživatelských účtů sdělujeme jméno, mobilní telefon a zemi. Následují doplňující informace o dosaženém vzdělání, aktuální práci apod. Přidáváním přátel do kruhů (v případě Google+) nebo do skupin (Facebook) umožníme analýzu našich sociálních vazeb (přátelé, spolužáci, kolegové ...).

Ačkoli poskytnuté informace (jméno, příjmení či datum narození) nejsou data popisující chování nebo zájmy uživatele, poskytnou možnost spojení vytvořeného profilu s konkrétní osobou.

## 4.2 Analýza reálných uživatelských dat

Zařízení, na kterých používáme služby od Google, odesílají reálná uživatelská data. Uživatelskými daty jsou myšleny údaje mající určitou vypovídající hodnotu o konkrétních uživateli. Tato data jsou vstupními hodnotami pro následné analytické zpracování. Důkazem o této činnosti je prosté vyhledávání ve vyhledávači.

Při používání webového vyhledávače Google si nelze nevšimnout automatického doplňování hledaných výrazů nebo automatické opravy chyb v hledaných frázích. Tyto nástroje jsou výsledkem dlouhodobé analýzy historie našeho vyhledávání.

Díky sdílení informací napříč Google aplikacemi může uživatel jednoduše vyhledávat informace o své osobě ve vyhledávači. Promítají se zde emaily, kalendář, poznámky nebo například fotky v programu Picasa. Nyní se už jedná o zcela soukromá data, která uživatel poskytuje k analýzování.



Aplikace Google Maps nabízí možnost ukázky aktuálního provozu. To je docíleno zkoumáním GPS pozice, kterou zařízení odesílá, za jednotku času. Jestliže se více uživatelů pohybuje po silnici pomaleji než v jinou dobu, indikuje to na zvýšenou hustotu dopravy [7].

Všechny tyto analýzy prezentuje společnost Google jako vylepšení stávajících služeb [7]. Pokud si neuvědomíme možnosti zneužití dat poskytnutých k analýze, mohli bychom tyto věci chápat jako drobnosti usnadňující nám náš každodenní „online“ život. Spojíme-li ovšem všechny informace dohromady, dostaneme přesný popis konkrétního člověka. Jeho pohyb, korespondenci, zájmy, naplánované schůzky, fotografie, telefonní číslo, jméno, bydliště a další. Při takto podrobné analýze lze v průběhu delšího časového úseku a s jistou mírou přesnosti předpovědět budoucí kroky právě tohoto uživatele.

# 5 Technologie znemožňující sledování

Jak již bylo několikrát zmíněno, pohyb na internetu není zcela anonymní. Existují nástroje, kterými se dá tato anonymita a tím i soukromí uživatelů zvýšit.

## 5.1 IXquick

IXquick je webový vyhledávač, který nesdílí žádné soukromé informace a ani si je nikam neukládá. Vyhledávací engine tohoto vyhledávače pracuje následovně. IXquick nechá hledaný dotaz vyhledat pěti různými vyhledávači a na základě jejich shody zobrazuje výsledky. Jde tedy o metavyhledávač. Výsledky jsou následně označeny hvězdičkovým systémem. Čím více hvězd odkaz má, tím větší je shoda napříč různými vyhledávači a naopak. V seznamu nabídnutých výsledků má uživatel možnost zobrazit webovou stránku anonymně, tedy skrze proxy server nebo využít tzv. „highlight“ odkazu. Ten zviditelní vybraný příspěvek pro další vyhledávání [8].

Od roku 2008 je vyhledávač IXquick držitelem certifikátu European Privacy Seal. Tím společnost Surfboard Holding B.V., která vyhledávač vlastní, splňuje všechny podmínky týkající se ochrany soukromí spotřebitelů. Tyto podmínky určila Evropská unie [9].

### 5.1.1 StartPage

Společnost Surfboard Holding B.V. vlastní i další webový metavyhledávač StartPage. Tento metavyhledávač funguje obdobně jako IXquick, ovšem s výjimkou využití více vyhledávačů. StartPage používá pouze jeden vyhledávač, a to Google. Rozdíl mezi StartPage a Google je anonymita, kterou zaručuje certifikát European Privacy Seal stejně jako v případě IXquick [9].

## 5.2 DuckDuckGo

DuckDuckGo je dalším webovým vyhledávačem s vlastním vyhledávacím enginem. Tento vyhledávač klade velký důraz na bezpečnost soukromí uživatelů.

Neukládá žádná uživatelská data a soubory cookies omezuje na nutné minimum (nastavení webové stránky) [10].

Od roku 2013 se těší velké popularitě veřejnosti. To je pravděpodobně způsobeno únikem informací Edwarda Snowdena o masivním sledování a následném uchování informací uživatelů [11].

### 5.3 Ghostery

Společnost Ghostery, Inc. je další společností na trhu zabývající se ochranou soukromí uživatelů. Ghostery dnes spadá pod německou společnost Cliqz, která vyvíjí webový prohlížeč dbající na ochranu soukromí [12].

Mezi produkty firmy Ghostery patří mimo samotného webového prohlížeče i rozšíření právě pro prohlížeče. Toto rozšíření se snaží odhalit všechny trackery na navštěvovaných stránkách a následně je blokovat [13]. V přehledném uživatelském rozhraní vidíme, o jaké trackery se jedná (reklamní, analytické nebo tracker pro sociální média). Uživatel má pak možnost těmto trackerům důvěřovat (povolit), zakázat je na dané webové stránce nebo zakázat na všech webových stránkách.

Z důvodů odesílání informací o historii našeho procházení pomocí právě těchto trackerů může docházet k ohrožení našeho soukromí. Rozšíření Ghostery si tedy klade za cíl bezpečnější procházení webu [13].

### 5.4 Adblock

Adblock je další rozšíření pro webové prohlížeče od společnosti Eyeo. Snahou tohoto projektu je omezení reklamního sdělení skrze webové stránky. Projekt je inspirován rozšířením Adblock Plus a je k dispozici pro nejpoužívanější platformy webových prohlížečů [14].

Rozšíření pro odstranění reklamních sdělení funguje na principu porovnávání http dotazů s filtry. Pokud se adresa shoduje alespoň s jednou adresou ve filtrech, je požadavek zablokován. Adblock využívá EasyList [15], což je seznam nejzákladnějších filtrů. Znemožněním stahování reklamního sdělení je zvýšená rychlost načítání stránek. Tím se vyhledávání na webu stává efektivnějším [16].

Pro zefektivnění vznikl vedle seznamu filtrů EasyList jeho podmnožina EasyList Lite. Tam jsou uloženy filtry nejnavštěvovanějších stránek. Uživatelé, kteří chtějí pomoci s bojem proti reklamním spotům, mohou povolit anonymní zpracování historie jejich prohlížení. Touto metodou se optimalizuje seznam nejvyhledávanějších stránek a AdBlock nemusí procházet seznamem filtrů celého EasyListu [17].

V základním nastavení můžeme AdBlock pozastavit nebo ho na dané stránce či doméně zcela vypnout. Rozšířené nastavení nám umožní povolit určitou část reklam a editovat filtry.

## 5.5 Noiszy

Noiszy je plugin pro webové prohlížeče vytvořený z důvodu skrytí činnosti uživatele v šumu [18]. Toho je docíleno navštěvováním pseudonáhodných webových stránek. Základní seznam stránek je distribuován společně s pluginem. Po načtení webu získá plugin url adresy, které se na stránce vyskytují a které neodkazují na jinou doménu. Z nich vybere jednu, kterou Noiszy navštíví. Takto navštěvuje stránky v 75 % případů. Zbylé případy tvoří načtení nových webových stránek ze seznamu.

Seznam webových stránek Noiszy můžeme upravovat. Pokud je žádoucí, aby vyhledávání náhodných stránek probíhalo v českém jazyce, musíme ručně přidat všechny stránky, které budou později navštěvované. Stránky defaultního seznamu sice nelze odstranit, ale můžeme je skrýt. Tím nebudou použité při běhu pluginu.

## 5.6 TrackMeNot

Rozšíření fungující na stejné myšlence jako plugin Noiszy. Oba se snaží skrýt aktivitu uživatelů v šumu. Hlavní rozdíl mezi Noiszy a TrackMeNot je v navštěvovaných stránkách. Zatímco Noiszy navštěvuje přímo webové stránky a pohybuje se po nich v rámci jejich domény, TrackMeNot nechává vyhledávat jisté fráze webovým vyhledávačem [19].

Ke své práci využívá několik webových vyhledávačů jako Google, Bing, Yahoo nebo Baidu. Vyhledávací fráze jsou získávány z XML dokumentů určených pro RSS čtečky. Výhodou této slovní zásoby je získání smysluplných vět jako vyhledávacích frází a neustále se aktualizující obsah v dokumentech pro RSS čtečky. Pokud chceme

vyhledávat v mateřském jazyce, pak je zapotřebí nalézt weby, které mají k dispozici soubory pro RSS čtečky (RSS feeds). Odkazy na tyto weby musíme zadat v nastavení jako zdroj pro vyhledávací fráze.

## 6 Praktická část

V teoretické části byla pojmenována rizika související s využíváním webového vyhledávače Google. Cílem praktické části bylo vyvinout rozšíření do prohlížeče Google Chrome, znemožňující profilování uživatele na základě historie jeho vyhledávání. Plugin, který má toto profilování omezit, se nesnaží skrýt aktivitu uživatele, ale podsouvá společnosti náhodné dotazy. Tyto dotazy pak rozšíří základnu analyzovaných dat o řadu nesouvisejících hledaných výrazů, což by mělo vést ke zvětšení odchylky mezi skutečnou osobností a profilem vytvořeným analýzou.

### 6.1 Popis rozšíření

Funkce rozšíření spočívá v automatickém vyhledávání pomocí vyhledávače Google a probíhá v samostatné kartě prohlížeče. Vyhledávané výrazy jsou brány z textů webových stránek navštívených pluginem nebo ze slovní zásoby uložené v lokálním úložišti. Slovní zásoba je při běhu pluginu obměňována. Každému slovu je počítána četnost, která se při použití daného slova vždy navýší a po dosažení určité hodnoty je toto slovo odstraněno. Postupné mazání slov probíhá do dosažení minimální hranice počtu slov v seznamu. V režimu vyhledávání slov z textu jsou naopak nalezená slova ukládána do paměti, dokud není dosažena horní hranice počtu slov. Obě hranice (minimální a maximální počet slov v úložišti) si může uživatel nastavit sám.

Další konstanta, která lze do jisté míry nastavit v uživatelském rozhraní, je rychlost načítání nových stránek. Přesný čas mezi jednotlivými reloady je nežádoucí. Uživatel tedy vybírá pouze stupeň rychlosti a doba strávená na stránce je pak náhodná. To je pro napodobení vyhledávání reálnou osobou zcela zásadní.

Rozšíření ukládá náhodná slova z náhodných stránek, které později využije jako vyhledávací výrazy pro vyhledávač. Slova jsou ukládána na pevný disk zařízení i s jejich počtem použití. Ve složce na pevném disku, která přísluší pluginu, je přítomen soubor s příponou *.log*, který obsahuje veškeré informace uložené tímto rozšířením. Jedná se o seznam slov s jejich četností výskytu, počet navštívených stránek a nastavené uživatelské konstanty. Při prvním spuštění je do tohoto souboru nahrán základní výčet

slov, který je distribuován společně s pluginem, a následné přepsání četnosti po načtení první stránky.

```
H,ĚŮ.....wordsz. [{"frequency":0,"word":"traktor"}, {"frequency":0,"word":"emil"}, {"frequency":0,"word":"stroj"}, {"frequency":0,"word":"kombajn"}, {"frequency":0,"word":"buben"}, {"frequency":0,"word":"urabec"}]iĭleŮ.
.....wordsz. [{"frequency":0,"word":"traktor"}, {"frequency":1,"word":"emil"}, {"frequency":0,"word":"stroj"}, {"frequency":0,"word":"kombajn"}, {"frequency":0,"word":"buben"}, {"frequency":0,"word":"urabec"}].@&f.....
.....
```

Obrázek 1 Soubor se seznamem slov zobrazený v prohlížeči souborů Lister (Total Commander)

Horní polovina obrázku 1 obsahuje seznam slov nahraných při prvním spuštění. Zvýrazněný objekt, který obsahuje slovo a jeho četnost, ukazuje činnost pluginu při výběru náhodného slova. Zvýší četnost použití a následně uloží, což je vidět ve spodní polovině. Následující slova jsou výsledkem běhu pluginu. Tyto slova již byla vyhledána ve webovém prohlížeči a na disku slouží jako slovní zásoba.

<i>ukázka</i>	<i>významného</i>	<i>also</i>	<i>purposes</i>
<i>obsah</i>	<i>half</i>	<i>cambridge</i>	<i>stand</i>
<i>širších</i>	<i>15x</i>	<i>there's</i>	<i>were</i>
<i>nebo</i>	<i>samozřejmě</i>	<i>international</i>	<i>different</i>
<i>substantiva</i>	<i>vzniklou</i>	<i>bing</i>	<i>dictionary</i>
<i>bylo</i>	<i>laiky</i>	<i>mějte</i>	<i>minor</i>
<i>němčiny</i>	<i>prosince</i>	<i>leave</i>	<i>meaning</i>
<i>slovíček</i>	<i>stránek</i>	<i>překladu</i>	<i>[`mi:nɪŋ]</i>
<i>takového</i>	<i>glyphicon</i>	<i>překladateli</i>	<i>in'dɔr/</i>
<i>také</i>	<i>2018</i>	<i>překladatelskou</i>	<i>long</i>
<i>přečtete</i>	<i>11/17</i>	<i>reference</i>	<i>follow</i>
<i>policie</i>	<i>inventář</i>	<i>guides:</i>	<i>marsh</i>
<i>všechna</i>	<i>data</i>	<i>without</i>	

Podle výběru slov lze do budoucna připravovat přísnější specifikace pro výběr slov. Nyní se plugin zaměřuje pouze na délku slov, která musí být delší než tři znaky, a eliminaci některých speciálních znaků včetně interpunkčních znamének. Změna, která se pouhou optickou analýzou nabízí, je zaměření se na jazykový výběr. Neschopnost rozšíření rozpoznat jazyk vede, v případě výběru cizího slovíčka, k odklonění se od vyhledávání českých slov. V tomto případě se jedná o anglický jazyk, který není nejhorší volbou. Díky české verzi prohlížeče Google, který se snaží podstrkovat výsledky hledání v češtině, je veliká pravděpodobnost o znovunavrácení hledání do českého jazyka. Složitější případ nastane, když plugin navštíví stránky s obrázkovým písmem (čínština, korejština, japonština atd.). V tomto případě je jako slovo považován celý odstavec a jen v jedinečných případech je hledání navraceno zpět do mateřského jazyka. Řešením tohoto problému by mohlo být omezení horní hranice počtu znaků ve slovu nebo povolení jen určitých nebo výběr znaků z určitého rozsahu unikódu.

## 6.2 Architektura pluginu

Obecný plugin v prohlížeči Google Chrome se skládá z několika komponent. Background komponenta (tzn. běžící na pozadí) zpravidla obsahuje logiku celé aplikace, řídí další komponenty a v celém prohlížeči má jen jednu instanci. Do uživatelského rozhraní mohou patřit vyskakovací tzv. popup okna, která se objeví po kliknutí na ikonu pluginu v prohlížeči, i například nová karta či nové okno prohlížeče. Za pomoci standardního DOM modelu je možné číst a do jisté míry upravovat obsah stránek. K tomu se využívá tzv. content script. Ten je vložen a následně vykonán přímo v navštěvovaných stránkách. Nejdůležitější komponentou pluginu je Manifest. Jedná se o JSON soubor popisující jednotlivé komponenty. V Manifestu se mimo jiné uvádí i oprávnění, které plugin má.

## 6.3 Uživatelské rozhraní

Pro interakci s uživatelem se v tomto pluginu využívá popup okno, tedy vyskakovací okno, které se objeví po kliknutí na ikonu rozšíření. Ovšem důsledek běžícího pluginu je vidět v nové kartě, v níž se načítají webové stránky.



Aktivitu ze strany uživatele zachytává javascriptový soubor náležící popup oknu. Přijímá hodnoty od uživatele a předává je dále background komponentě.

```
var backGround = chrome.extension.getBackgroundPage();
```

Tento řádek kódu získává od prohlížeče odkaz na background komponentu, který uloží do proměnné. Odkaz zajistí přístup k background komponentě. Tím můžeme plugin ovládat pomocí vyskakovacího okna (popup.html).

## 6.4 Background komponenta

V background komponentě je uložena logika celé aplikace. Při spuštění aplikace se nejprve vytvoří nová karta v okně prohlížeče. K tomu je využito javascriptové API (tabs) sloužící pro práci s kartami a metoda `.create()` pro vytvoření nové karty.

```
chrome.tabs.create({});
```

Po vytvoření nové karty je zapotřebí získat odkaz na již vytvořenou kartu. K tomu nám postačí ID karty, které získáme dotazem na Chrome API (tabs) s volitelným parametrem ve zpětné funkci `array_of_Tabs`. Parametr obsahuje pole jednotlivých karet, kde na poslední pozici je námi vytvořená karta. Proměnná `.id` vrátí ID, které je následně uloženo do globální proměnné `tabId`.

```
chrome.tabs.query({}, function (array_of_Tabs) {  
    var tab = array_of_Tabs[array_of_Tabs.length - 1];  
    tabId = tab.id;  
});
```

Protože plugin načítá libovolné stránky, je riziko výskytu automatického přehrání videa nebo jiného zvukového projevu. Pomocí API (tabs) bude aktualizována karta, jejíž identifikační číslo je `tabId`. Kartě bude nastaven parametr `muted` na hodnotu `true`. Tím se kartě znemožní přehrát zvukový záznam.

```
chrome.tabs.update(tabId, {  
    muted: true  
});
```

Metoda `.update()` je využita i v případě načtení nové webové stránky. Ovšem v tomto případě bude použit parametr `url`, jehož hodnota odpovídá url nové stránky.

```
chrome.tabs.update(tabId, {  
    url: new_url  
});
```

V rámci inicializace pluginu je potřeba zkontrolovat přítomnost listu slov, která budou použita pro vyhledávání. Jestliže je plugin spuštěn poprvé anebo byl list odstraněn činností uživatele, bude zapotřebí tento seznam slov vytvořit. Následující kód získá z lokálního úložiště prohlížeče seznam slov v podobě pole.

```
chrome.storage.local.get({words : []}, function (result){  
    var words = result.words;  
    if(!words[0]){  
        //Seznam slov neexistuje  
    } else {  
        //Seznam slov existuje  
    }  
});
```

Pokud slovo `i` se svou četností na první pozici pole `words` neexistuje, znamená to absenci seznamu. Ten pak bude vytvořen a pomocí `XMLHttpRequestu`, který zde slouží pro výměnu dat mezi rozšířením a `JSON` souborem, se do něj uloží defaultní slovní zásoba. Tyto slova jsou uložena v samostatném souboru ve formátu `JSON`, který je distribuován společně s pluginem.

Při použití slova ze slovní zásoby je danému slovu zvýšena jeho četnost. To se děje z důvodu zamezení častému opakování stejných slov. Slova, která přesáhnou určitou mez, jsou ze seznamu odebrána. Pro odstranění konkrétního slova z pole je použita metoda `.splice`, která nahradí vyřazený prvek následujícím prvkem. Tímto postupem nebude žádná pozice v poli prázdná. Po změně je pole uloženo do lokálního úložiště díky Chrome API storage.

```
chrome.storage.local.set({klíč : nové_pole}, function(){  
    //Kód vykonaný po uložení  
});
```

Dalším úkolem této komponenty je volat content scripty s přístupem k obsahu stránky. Skript se spustí zavoláním metody *executeScript*. Tím se skript vloží do stránky načtené v kartě s identifikačním číslem *tabId*. Parametr *file* v této metodě určuje, jaký skript bude do stránky vložen a následně vykonán.

```
chrome.tabs.executeScript(tabId, {  
    file:'jméno content scriptu'  
});
```

K získání výsledku content scriptu je nutno v tomto pluginu přidat tzv. posluchače. Content script odesílá zprávu napříč celým pluginem. Parametr *request* obsahuje žádosti zaslané z content scriptu. Tyto žádosti jsou ve formátu JSON. Sender obsahuje informace o odesílateli. Protože je zpráva posílána přes celý prohlížeč, v informacích je tedy ID nebo URL karty, ze které byla zpráva odeslána. *SendMessage* je využit pro odeslání odpovědi zpět ke zdroji. Odpovědi musí být opět ve formátu JSON. V tomto pluginu se využívá pouze parametr *request*, kde hodnota *action* označuje skript, který zprávu zaslal, a hodnota *source* obsahuje návratovou hodnotu content scriptu.

```
chrome.runtime.onMessage.addListener(  
    function (request, sender, sendResponse) {  
        var result = request.source;  
        switch (request.action) {  
            case "url":  
                //Pokračování kódu  
                break;
```

```

        case "word":
            //Pokračování kódu

            break;
        }
    }
});

```

## 6.5 Content script

Content script je kód, který je vložen a následně spuštěn ve webové stránce. V tomto pluginu se používají dva content scripty. Jeden pro získání url adres pro následné navštívení webových stránek a druhý pro získání náhodného slova z textů v již navštívených stránkách. Tyto skripty nejsou přítomny v prohlížeči po celou dobu existence pluginu, jako je tomu u background komponenty, a ani nejsou definované v souboru Manifest. Pokud by byly uvedené přímo v Manifestu s přiřazenými právy, znamenalo by to vložení skriptů do všech stránek, které jsou (nebo budou) otevřené uživatelskou činností. To by bylo nežádoucí, a proto se tyto skripty volají z background komponenty pro určitou kartu.

Po spuštění skriptu se odešle skrze celý plugin zpráva, kterou zachytí posluchač přítomný v background komponentě. Pro posláni zprávy se využije Chrome API, tentokrát *runtime*. Toto API je využíváno pro vynucení nějaké události, získání informací z Manifestu nebo zaslání zpráv. Zaslání zprávy napříč pluginem je umožněno díky metodě *sendMessage()*. Zpráva je předána jako JSON soubor, ve kterém uvádíme klíčová slova a jejich hodnoty. V tomto případě jsou klíčovými slovy *action* pro jméno akce, která je vykonána a díky níž se v background komponentě může s výsledkem správně naložit, a *source* pro zdroj. V tomto případě je to funkce, jejíž návratová hodnota bude odeslána ve zprávě.

```
chrome.runtime.sendMessage({
    action: "název",
    source: funkce()
});
```

### 6.5.1 Získání adresy stránky

Skript pro získání adresy stránky využívá přístup k DOM stránky. Odkazy na stránky ve výsledku vyhledávání v Google jsou uloženy v elementech, jejichž společná třída je „r“. Z těchto elementů bude potřeba vyselektovat elementy odkazů `<a>` a z nich získat atribut `href`, ve kterém je uložena adresa stránky. Adresy se uloží do pole pomocí metody `push()`. Adresa nově načtené stránky bude náhodně vybrána z pole uložených adres. K náhodnému výběru je využita metoda `random()`, která vrací číslo od 0 (včetně) do 1 (kromě). Proto je třeba toto číslo vynásobit maximální hranicí. Tou je v tomto případě počet prvků v poli adres. Získání celého čísla bylo docíleno metodou `floor()`, která zaokrouhluje desetinná čísla směrem dolů.

```
var urls = new Array();
var rClass = document.getElementsByClassName("r");
for (var i = 0; i < rClass.length; i++) {
    var x = rClass[i].getElementsByTagName("a");
    if(x.length > 0){
        urls.push(x[0].getAttribute("href"));
    }
}
var url=urls[Math.floor((Math.random()*urls.length) + 0)];
```

### 6.5.2 Získání náhodného slova

Tento content skript, stejně jako předchozí script, má přístup k DOM stránky. Jeho úkolem je získání náhodného slova z webové stránky. Je předpokladem, že text

vhodný pro výběr slov se bude vyskytovat v elementech `<p></p>`. Metodou `.random()` je vybrán náhodný odstavec, jehož slova jsou metodou `.split()` vložena do pole.

```
var arrayPar = document.getElementsByTagName("p");
var par = arrayPar[Math.floor((Math.random() *
                                arrayPar.length) + 0)].innerHTML;
var words = par.toString().split(" ");
```

Z tohoto seznamu je vybráno náhodné slovo, jehož počet znaků je větší než tři. Důvodem této podmínky je eliminace spojek, zájmen nebo částic. Následně je slovo kontrolováno na přítomnost určitých znaků (`<`, `>`, `&`, `#`, `"`). Tyto znaky by mohly indikovat slovo z části vnořeného html elementu. K nalezení speciálních znaků je použita metoda `.search()`, jejíž parametr určuje hledaný výraz. V tomto případě bude parametr ve formě regulérního výrazu. Metoda v případě nálezů výrazu vrátí číslo určující první nalezenou pozici. Pokud ovšem nebude shody dosaženo, výsledná hodnota bude `-1`.

```
if(word.search(/[<|>|&|#|"|']/) > 0) {
    //Znak se ve slově vyskytuje
} else {
    //Znak se ve slově nevyskytuje
}
```

Z vybraného slova se nakonec odstraní interpunkční znaménka. Toho se docílí metodou `.replace()`, která nahradí tyto znaménka prázdným znakem. Metoda `.toLowerCase()` převede všechny znaky písmen na malá písmena.

```
word.replace(/[.,?!]/g, "").toLowerCase();
```

## 6.6 Manifest

Jedná se o soubor ve formátu JSON, který poskytuje důležité informace o rozšíření a jeho oprávněních. Podle potřeby pluginu se do Manifestu přidávají jednotlivé údaje. Některé údaje jsou povinné (`name`, `version`, `manifest_version`), některé

doporučené (description, icons, default\_locale), ale většina údajů je volitelných právě pro rozdílné potřeby jednotlivých pluginů.

Manifest\_version určuje verzi formátu Manifestu. V současné době je nejnovější verze označována číslem 2. Manifest verze 1 byl nahrazen už ve webovém prohlížeči Chrome verze 18 (současná verze prohlížeče je 65). Version udává aktuální verzi pluginu, kterou si vývojář určuje sám na základě vlastního principu verzování.

Sekce browser\_action je sice nepovinná, ale zásadní pro přiřazení ikony k rozšíření, ovládání rozšíření pomocí popup okna nebo informační popis zobrazený po najetí kurzoru myši. V tomto rozšíření je z tohoto výčtu použito nastavení ikony rozšíření a vyskakovací okno (popup). Výsledný zápis této části by vypadal následovně.

```
"browser_action": {  
    "default_icon": "icon.png",  
    "default_popup": "popup.html"  
}
```

V části background se uvádějí skripty nebo html stránky běžící na pozadí rozšíření. V celém rozšíření může být pouze jedna instance background komponenty. Vyjímkou je anonymní režim prohlížeče. Pokud je anonymní režim prohlížení využíván, pak jedna instance je pro běžný prohlížeč a druhá pro anonymní okna.

```
"background": {  
    "scripts": ["engine.js"]  
}
```

Jestliže chceme využívat jednotlivé API prohlížeče Chrome, pak tato rozhraní musíme uvést v poli permissions, které shromažďuje veškerá oprávnění, která pluginu udělíme. Ta se nemusí nutně týkat pouze využívání API. V tomto pluginu je definováno oprávnění pro využití karet prohlížeče a povolení všech url.

```
"permissions": ["activeTab", "tabs", "<all_urls>", "storage"]
```

Web\_accessible\_resources je pole obsahující cesty k souborům, které budou použity ve webových stránkách. V tomto rozšíření se jedná o dva content scripty.

```
"web_accessible_resources": ["getUrl.js", "getWord.js"]
```

V rámci bezpečnějšího nastavení pluginu je v Manifestu nastaven content security policy. V tomto nastavení uvádíme, z jakých adres se bude moci načítat obsah třetích stran na stránkách, které budou navštíveny pluginem.

## 6.7 Testování funkčnosti

Vzniklý plugin automaticky vyhledává náhodná slova a navštívuje různé stránky. Tím se zvětší objem dat, které se musí zpracovat pro vytvoření profilu. Důsledek celého tohoto procesu je ztracení našich skutečných dat v šumu nepodstatných věcí. Aby mohla být funkčnost potvrzena, je potřeba nelézt způsob, který dokazuje možné profilování. Společnost Google umožňuje stažení vlastních uživatelských dat, vzniklých vyhledáváním. Ovšem výsledkem této operace je soubor nepřehledného množství námi zadaných vyhledávacích dotazů včetně dotazů položených pluginem. Z tohoto souboru nelze vyčíst, zda se mění náš profil, který je vytvořen z našich vyhledávacích dat.

Jediná možnost, jak dokázat funkčnost pluginu přímo na probíhajícím profilování, je cílená reklama. Ta vzniká podrobným zkoumáním zájmů uživatele. Společnost Google tento proces nazývá jako personalizace reklam.

### 6.7.1 Google Ads Settings

Pro ověření správné funkčnosti pluginu, který má znemožnit profilování uživatele na základě jeho vyhledávání v Google, použijte službu Google Ads Settings. Jedná se o správu reklamního sdělení. Princip této služby spočívá v přiřazování štítků (frází podle nichž se zobrazují reklamy). Například pokud vyhledáváme tenisovou raketu a láhev na pití, dostaneme štítek sportovní náčiní. Systém pak ví, že jakákoliv reklama na sportovní potřeby by nás mohla zaujmout. Pro tvorbu těchto štítků se nevyužívá pouze Google vyhledávač, ale i sledování videí na YouTube nebo z webů třetích stran.

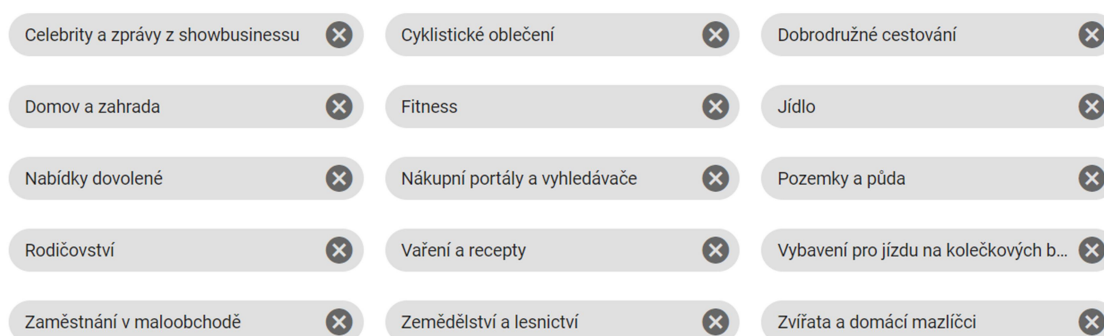
Google Ads Settings umožňuje i editaci těchto štítků. Pokud si uživatel nepřeje, aby se mu konkrétní reklama zobrazovala, může tento štítek odstranit. Neznamená to ovšem, že by společnost Google tento štítek smazala. Pouze se přestane zobrazovat



reklama, která by byla vybrána na základě tohoto štítku. Celý tento přehled a správu reklam můžeme chápat jako shrnutí informací, které o nás (na nás) Google ví.

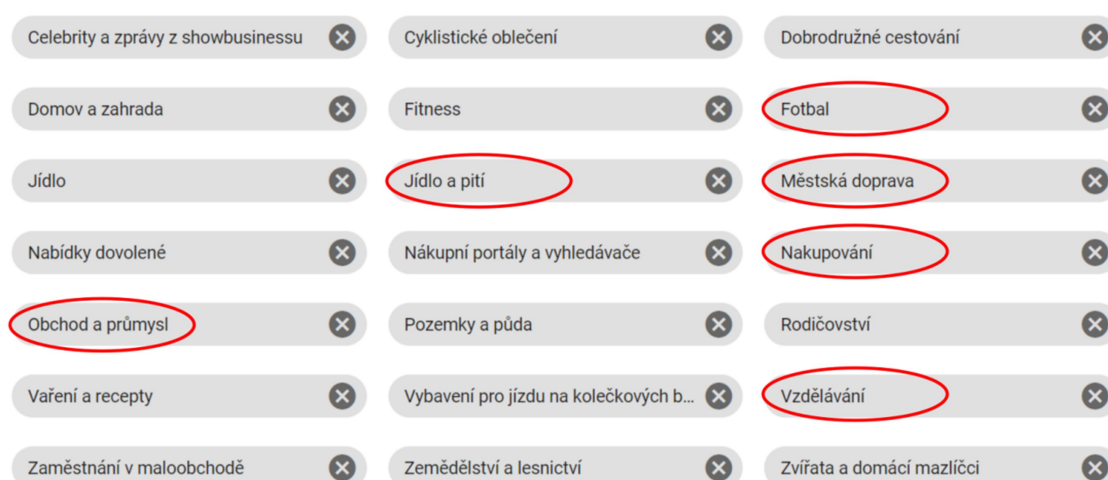
## 6.7.2 Výsledek

Správnou funkčnost rozšíření lze identifikovat pouhým pozorováním reklam a změnou štítků pro výběr reklam v Google Ads Settings. Při běhu pluginu by tedy měla nastat obměna štítků, aniž bychom sami vyhledávali.



Obrázek 2 Štítky v Google Ads Settings

Na obrázku 2 jsou k vidění štítky tak, jak je zobrazuje Google Ads Settings. Vznikly důsledkem skutečného vyhledávání před použitím pluginu. Po zachycení tohoto stavu byl na stejném počítači a pod stejným uživatelským účtem spuštěn plugin. Ten běžel po dobu jednoho dne. Za tuto dobu plugin navštívil 667 webových stránek.



Obrázek 3 Štítky v Google Ads Settings po použití pluginu

Štítky na obrázku 3 ukazují činnost samotného pluginu. Kromě původních štítků, které byly přiřazeny vyhledáváním uživatele, přibyly nové štítky vzniklé důsledkem běhu pluginu. Detailnějším zkoumáním slov, které plugin využil pro vyhledávání, a nově přidanych štítků, jež byly přidány společností Google po analýze našeho vyhledávání, jsem našel jistou podobnost mezi slovy a štítky.

Štítky se nerovnají vyhledávacím frázím, ale popisují okruh možného zájmu. V souboru, který uchovává stavové informace pluginu, jsem tedy hledal spojitost mezi vyhledáváním a štítky, které mi byly přiřazeny. Z tohoto souboru byla vybrána slova pro demonstraci funkčnosti pluginu.

```
{"frequency":2, "word": "výchova"}  
{"frequency":1, "word": "výuka"}  
{"frequency":1, "word": "přijímacím"}  
{"frequency":2, "word": "cambridge"}
```

Výše uvedený výčet slov naznačuje jistou spojitost s přidáním štítkem „výuka“. To je důkaz o změně profilu vytvářeným společností Google za účelem cílené reklamy.

Testování správné funkce pluginu byla prováděna na politice reklamního sdělení společnosti Google. Výsledkem a funkcí pluginu ovšem není zabránit zobrazování cílené reklamy, ale znemožnit vytváření profilu uživatele. Cílem tohoto rozšíření je tedy zvýšení ochrany soukromí uživatelů před masivním sledováním.

## 7 Závěr

V teoretické části práce byla představena rizika hrozící při běžném využívání moderních komunikačních technologiích. Zaměřil jsem se na seznámení s digitální stopou, jelikož spousta uživatelů bere takto zanechané informace na lehkou váhu nebo si existenci digitální stopy nepřipouští. V závěru této části jsem uvedl dostupné prostředky, kterými se lze sledování částečně bránit nebo alespoň omezit tvorbu digitálních stop.

V praktické části se věnuji tvorbě samotného rozšíření. Uvádím a rozebírám jednotlivé logické prvky pro celkovou představu funkce pluginu. Výsledný plugin představuje další možnost ochrany soukromí uživatelů před sledováním a proti sběru citlivých dat, jak je uvedeno v teoretické části.

Za zmínku stojí celkový přístup pluginu k ochraně soukromí uživatelů. Nesnaží se skrývat aktivitu uživatele, jako je tomu většiny technologií snažících se zvýšit soukromí uživatelů, ale veškerá aktivita je obklopena nepodměnými dotazy. Sledující strana je pak zahlcena nereálnými informacemi, což vede k odchýlení našeho vytvořeného profilu od skutečnosti. Stejný přístup mají i rozšíření Noiszy a TrackMeNot, jejichž princip fungování je představen v teoretické části. Oba tyto pluginy vyžadují počáteční přidání stránek jako výchozí zdroje. Noiszy pouze navštívuje webové stránky, naopak TrackMeNot provádí jen vyhledávání ve webových vyhledávacích. Mnou vytvořený plugin spojuje obě tyto funkčnosti. Provádí vyhledávání ve webovém prohlížeči a následně navštívuje nalezené webové stránky. Tento proces je navíc zautomatizován, není potřeba přidání odkazů na webové stránky jako výchozí zdroje.

Celý plugin nabízí několik možností, kam jeho funkcionalitu posunout. Nyní je projekt ve fázi posílání náhodných dotazů, jejichž výběr není nijak zásadně limitován a je jednoslovný. To může vést k nesmyslnému vyhledávání, které by časem mohlo být odhaleno. Účinnost pluginu by potom byla nulová. Rozšíření tohoto pluginu by mělo spočívat v tvorbě náhodných a smysluplných otázek jako vyhledávacích frází. Zajímavou myšlenkou je doplnit plugin o kategorizování slov podle významu

a analyzování reálných vyhledávacích dotazů zadaných uživatelem. Pokud bych například vyhledával informace o určitém zdravotním problému, plugin by automaticky vyhledával stránky s lékařskou tematikou.

Všechny tyto myšlenky možného vývoje jsou ovšem pouze teoretické. Není tedy jisté, zda plugin vyřeší záludnosti českého jazyka jako skloňování nebo slovosled. V případě nalezení schůdné cesty nesmíme zapomenout na omezený výpočetní výkon přidělený prohlížečem.

## 8 Seznam literatury

- [1] BALL, James a Dominic RUSHE. NSA Prism program taps in to user data of Apple,Google and others. In: The Guardian [online]. 2013 [cit. 2017-02-08]. Dostupné z: <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>
- [2] Manage Your Identity. Internet Society [online]. Geneva, c2018 [cit. 2018-02-25]. Dostupné z: <https://www.internetsociety.org/tutorials/manage-your-identity/>
- [3] OpenID: The Internet Identity Layer [online]. c2018 [cit. 2018-02-25]. Dostupné z: <http://www.openid.net/>
- [4] MADDEN, Mary, Susannah FOX, Aaron SMITH a Jessica VITAK. Digital Footprints: Online identity management and search in the age of transparency [online]. Washington: Pew Internet & American Life Project, 2007 [cit. 2016-11-25]. Dostupné z WWW: [http://www.pewinternet.org/files/old-media/Files/Reports/2007/PIP\\_Digital\\_Footprints.pdf.pdf](http://www.pewinternet.org/files/old-media/Files/Reports/2007/PIP_Digital_Footprints.pdf.pdf)
- [5] POLESNÝ, David a Stanislav JANŮ. Do Not Track je fiasko, služby vás na webu stopují dál Více na: <https://www.zive.cz/clanky/do-not-track-je-fiasko-sluzby-vas-na-webu-stopuji-dal/sc-3-a-172821/>. In: Zive [online]. 12. března 2014 [cit. 2018-03-25]. Dostupné z: <https://www.zive.cz/clanky/do-not-track-je-fiasko-sluzby-vas-na-webu-stopuji-dal/sc-3-a-172821/>
- [6] L. MITCHELL, Robert. What Google Knows About You: Google may know more about you than your mother does. Got a problem with that?. Computer World [online]. Framingham, c2018, May 11, 2009 [cit. 2018-02-21]. Dostupné z: <https://www.computerworld.com/article/2551008/networking/what-google-knows-about-you.html>
- [7] Je pro nás důležité, abyste porozuměli tomu, jak shromažďujeme a využíváme data. Naše služby na základě získaných dat neustále zlepšujeme. [online]. Mountain View, Kalifornie, USA [cit. 2018-03-02]. Dostupné z: <https://privacy.google.com/intl/cs/your-data.html>

- [8] Ixquick Protects Your Privacy!. Ixquick [online]. Zeist (Netherland), c2018 [cit. 2018-02-21]. Dostupné z: <https://www.ixquick.eu/eng/what-makes-ixquick-special.html?#hmb>
- [9] European Privacy Seal for Ixquick and Startpage. European Privacy Seal [online]. Bonn, c2008-2017 [cit. 2018-02-21]. Dostupné z: <https://www.european-privacy-seal.eu/EPS-en/Ixquick-startpage>
- [10] We don't collect or share personal information. DuckDuckGo [online]. Paoli, Pennsylvania [cit. 2018-02-21]. Dostupné z: <https://www.duckduckgo.com/privacy>
- [11] ARTHUR, Charles. NSA scandal delivers record numbers of internet users to DuckDuckGo. The Guardian [online]. London, c2018, 10 Jul 2013 [cit. 2018-02-21]. Dostupné z: <https://www.theguardian.com/world/2013/jul/10/nsa-duckduckgo-gabriel-weinberg-prism>
- [12] Ghostery is Acquired by Cliqz!. GHOSTERY [online]. New York, February 15, 2017 [cit. 2018-02-24]. Dostupné z: <https://www.ghostery.com/blog/ghostery-news/ghostery-acquired-cliqz/>
- [13] GHOSTERY [online]. New York [cit. 2018-02-24]. Dostupné z: <https://www.ghostery.com/>
- [14] Adblock [online]. Cologne Germany [cit. 2018-02-25]. Dostupné z: <https://getadblock.com/>
- [15] EasyList [online]. [cit. 2018-02-25]. Dostupné z: <https://easylist.to/>
- [16] CASSIDY, Rhana. Adblock: How does Adblock work?. Adblock [online]. 10 Aug, 2017 [cit. 2018-02-25]. Dostupné z: <https://help.getadblock.com/support/solutions/articles/6000087914-how-does-adblock-work->

- [17] CASSIDY, Rhana. AdBlock: How does AdBlock work?. AdBlock [online]. 8 Nov, 2017 [cit. 2018-02-25]. Dostupné z: <https://help.getadblock.com/support/solutions/articles/6000142821-what-is-easylist-lite->
- [18] Noiszy: About. Noiszy [online]. c2017 [cit. 2018-04-11]. Dostupné z: <https://noiszy.com/about-me/>
- [19] C. HOWE, Daniel. TrackMeNot. *NYU|courant COMPUTER SCIENCE* [online]. [cit. 2018-04-13]. Dostupné z: <https://cs.nyu.edu/trackmenot/>

## 9 Seznam obrázků

Obrázek 1 Soubor se seznamem slov zobrazený v prohlížeči souborů Lister (Total Commander)	14
Obrázek 2 Štítky v Google Ads Settings.....	24
Obrázek 3 Štítky v Google Ads Settings po použití pluginu .....	24



# 10 Seznam příloh

Příloha 1 – Zdrojové kódy

Příloha 2 – Návod