PROBLÉMY VYHĽADÁVANIA INFORMÁCIÍ V KONTEXTE SÚKROMIA

Mgr. Mirka Pastierová, PhD.; mirka.pastierova@uniba.sk; (Katedra knižničnej a informačnej vedy, Filozofická fakulta, Univerzita Komenského v Bratislave)

Techniky zberu, monitorovania a analýzy dát o dotazoch a používateľoch prieskumových strojov sa stali zdrojom množstva opodstatnených obáv spojených s bezpečnosťou, súkromím a ochranou osobných údajov. Praktiky bežných prieskumových strojov generujú problémy ako filtračné bubliny, cielenie reklamy, ale aj potenciálne riziko úniku dát o vyhľadávaní. Článok objasňuje konkrétne techniky anonymizácie dát spolu s dostupnými riešeniami na úrovniach od operačného systému po špecializované aplikačné riešenia. Samostane sa venujeme charakteristike vybraných anonymných a súkromných prieskumových strojov predstavujúcim alternatívu voči bežným vyhľadávačom, ktoré často stavajú svoje monetizačné modely na zdieľaní a predaji používateľských dát tretím stranám.

https://doi.org/10.52036/1335793X.2021.1-2.24-30

PROBLÉMY VYHĽADÁVANIA INFORMÁCIÍ V KONTEXTE SÚKROMIA

Techniky zberu, monitorovania a analýzy dát o dotazoch a používateľoch prieskumových strojov sa stali zdrojom množstva opodstatnených obáv spojených s bezpečnosťou, súkromím a ochranou osobných údajov. Praktiky bežných prieskumových strojov generujú problémy ako filtračné bubliny, cielenie reklamy, ale aj potenciálne riziko úniku dát o vyhľadávaní. Článok objasňuje konkrétne techniky anonymizácie dát spolu s dostupnými riešeniami na úrovniach od operačného systému po špecializované aplikačné riešenia. Samostane sa venujeme charakteristike vybraných anonymných a súkromných prieskumových strojov predstavujúcim alternatívu voči bežným vyhľadávačom, ktoré často stavajú svoje monetizačné modely na zdieľaní a predaji používateľských dát tretím stranám.

PREDSTAVENIE PROBLÉMU

Zneužívanie osobných údajov a obavy súvisiace so zabezpečením dát citlivých na ochranu sa stali v post-Snowdenovskej ére problémami, ktoré podnecujú v spoločnosti opodstatnené znepokojenie. Nie sú ojedinelé prípady sledovania online a telefónnej komunikácie vládami, ale aj súkromnými firmami. V nadväznosti na otázky súvisiace s ochranou súkromia a zabezpečenia osobných údajov bol realizovaný výskum (Auxier et al. 2019), ktorý poukázal na to, že väčšina respondentov zastáva názor, že má veľmi malú alebo žiadnu kontrolu nad údajmi, ktoré o nich zhromažďuje vláda (84 % respondentov) a firmy (81 % respondentov). Približne dve tretiny respondentov uviedlo, že ich znepokojujú metódy používania zhromažďovaných osobných údajov. Až 72 % respondentov vyjadrilo pocit, že všetko alebo takmer všetko, čo robia online či pri využívaní mobilu je monitorované inzerentmi, technologickými firmami a inými spoločnosťami.

Problematiku súkromia v kontexte digitálneho prostredia podrobne analyzoval z rôznych pohľadov vo svojej štúdii Šušol (2019). Poukázal aj na iniciatívy dôležité v oblasti legislatívnej regulácie s priamym dopadom na jednotlivcov. Cieľom nasledujúceho textu je poukázať na témy prepojené na súkromie z perspektívy vyhľadávania informácií a predstaviť konkrétne techniky a nástroje, ktoré prispievajú k anonymizácii používateľov webových vyhľadávacích služieb.

DVE TVÁRE PERSONALIZÁCIE

Je známym faktom, že moderné webové prieskumové stroje sledujú, analyzujú a využívajú informácie o lokácii, demografické údaje, históriu vyhľadávania a ďalšie behaviorálne signály na účely vytvárania osobných profilov, následnej personalizácie výsledkov a cielenia reklamy. Samotná história zadávaných dotazov môže odhaliť meno používateľa, bydlisko, informácie o práci, rodine, záujmoch či plánoch. Okrem toho sa do histórie spolu s dotazom zvyčajne ukladajú aj údaje o dátume a čase vyhľadávania spolu s ďalšími dátami ako IP adresa, používateľský agent (špecifikácia typu a verzie prehliadača, typ operačného systému), údaje o poskytovateľovi internetového pripojenia a unikátnom identifikátore, ktorý je uložený v pre-

hliadači pomocou cookie (Feigenbaum 2012, We [bez dátumu]). Pri prihlásení na používateľský účet sú to aj údaje o používateľskom mene a emailovej adrese.
Cieľom využívania týchto dát je lepšie pochopenie zámeru a kontextu vyhľadávania a vo výsledku zefektívnenie prístupu k relevantnému obsahu, ktorý je šitý na mieru danému používateľovi. Techniky personalizácie sú však dvojsečnou zbraňou, na jednej strane môžu priniesť zvýšenú užitočnosť danej vyhľadávacej služby, ale na strane druhej generujú riziko narušania súkromia (Ahmad et al. 2018). Ďalším negatívnym efektom personalizácie prieskumovými strojmi sú aj filtračné bubliny, ktoré limitujú pri prístupe k širokému spektru informácií. Namiesto toho vytvárajú stav intelektuálnej izolácie spôsobenej algoritmami, ktoré filtrujú výsledky vyhľadávania na základe profilu (modelu) používateľa.
Na tieto problémy poukázali aj výsledky výskumu (Purcell et al. 2012) potvrdzujúce, že ľudia sú síce stále spokojnejší s kvalitou vyhľadávaných výsledkov, ale majú aj obavy zo zberu osobných informácií prieskumovými strojmi. Celkovo 65 % respondentov sa vyjadrilo, že nesúhlasia so zberom informácií o vyhľadávaní, ktoré sú následne využité na určenie poradia výsledkov vyhľadávania v budúcnosti, pretože to limituje prístup k informáciám online a k výsledkom vyhľadávania viditeľných na výstupe. Až 73 % respondentov personalizáciu vyhľadávania vnímalo ako metódu, ktorá je invazívna vzhľadom na súkromie a taktiež 68 % si neželalo, aby na nich bolo cielená reklama, pretože nesúhlasia s tým, aby bolo ich online správanie sledované a analyzované.

ÚNIK DÁT O VYHĽADÁVANÍ

Využívanie bežných prieskumových strojov okrem filtračných bublín a cielenia reklamy generujú ďalší podstatný problém. Je ním potenciálny únik dát o vyhľadávaní a ich zdieľanie či predaj tretím stranám. V minulosti bolo medializovaných množstvo prípadov, kedy došlo k zámernému alebo aj nezámernému úniku dát z používateľských profilov či logov vyhľadávania. Jeden z najznámejších je prípad, keď spoločnosť AOL (Arrington 2006) zverejnila bez súhlasu používateľov logy vyhľadávania, ktoré obsahovali dáta o dotazoch zadaných do tohto vyhľadávača počas obdobia troch mesiacov. Práve logy webového vyhľadávania obsahujú dáta extrémne citlivé na ochranu (Hong et. al 2012). Napriek tomu, že logy uvádzali používateľov pod anonymnými číselnými identifikátormi, niekedy ich bolo možné na základe zadávaných dotazov (uvádzali napr. svoje celé meno, adresu ap.) identifikovať. Celkovo súbor obsahoval 20 miliónov dotazov vyhľadávaných 650 tisíc používateľmi. Po niekoľkých dňoch AOL tieto dáta prestal bez ďalšieho vysvetlenia zverejňovať, medzitým však boli distribuované ďalej.
K úniku dát o vyhľadávaní dochádza bežne aj po kliknutí na výsledok, kedy je špecifický dotaz následne zasielaný navštíveným stránkam pomocou poľa sprostredkovateľ v rámci HTTP hlavičky. Pri využívaní stránok prieskumových strojov s podporou enkrypcie pomocou HTTPS protokolu sa zvyčajne nezasielajú dotazy navštíveným stránkam. Po kliknutí na stránku, ktorá však takisto využíva HTTPS, je dotaz aj tak zaslaný danej stránke s tým rozdielom, že neunikne a nie je dostupný iným počítačom.
Okrem zasielania údajov o dotaze pri návšteve akejkoľvek stránky posiela počítač automaticky aj informácie o IP adrese a používateľskom agentovi (We [bez dátumu]).

TECHNIKY A NÁSTROJE ANONYMIZÁCIE VYHĽADÁVANIA

Aktuálne sú vyvíjané rôzne prístupy na riešenie ochrany súkromia používateľov pri vyhľadávaní na webe s využitím rozmanitých techník anonymizácie dát. Spoliehajú sa primárne na kryptografické techniky a/alebo na techniky zahmlievania (z angl. obfuscation techniques) (Ahmad et al. 2018, Boutet et al. 2016). Techniky zahmlievania sú založené na maskovaní identity používateľa alebo generovaní falošných dotazov a dát o kliknutiach. Obe tieto techniky majú veľa nevýhod spojených s vysokými nárokmi na spracovanie (keďže do úvahy sa berú všetky dotazy nezávisle od ich obsahu), ale aj na réžiu a latenciu siete. Tieto techniky taktiež ignorujú to, že pôvodný dotaz používateľa môže obsahovať obsah citlivý na ochranu napr. informácie o vierovyznaní, politických preferenciách, zdravotnom stave ap. Na základe toho bol navrhnutý modul (Boutet et al. 2016), ktorý sémanticky analyzuje citlivosť dotazu a následne anonymizuje a chráni len tie dotazy, ktoré si to vyžadujú.
Techniky anonymizácie dát môžu byť súčasťou algoritmu samotného prieskumového stroja, ale fungujú aj na ďalších úrovniach počnúc operačným systémom, prehliadačmi na prezeranie webu popr. ďalšími špecializovanými softvérovými aplikáciami.
Apple (App 2021) vo svojom najnovšom mobilnom operačnom systéme OS 14 prišiel so zásadnými zmenami, ktoré ovplyvňujú monitorovanie a zber používateľských údajov aplikáciami. V rámci ATT (App Tracking Transparency) je poskytovateľ aplikácie povinný zobraziť upozornenie s požiadavkou na povolenie monitorovania dát o koncovom používateľovi alebo jednoducho reflektovať stanovené pravidlá ochrany

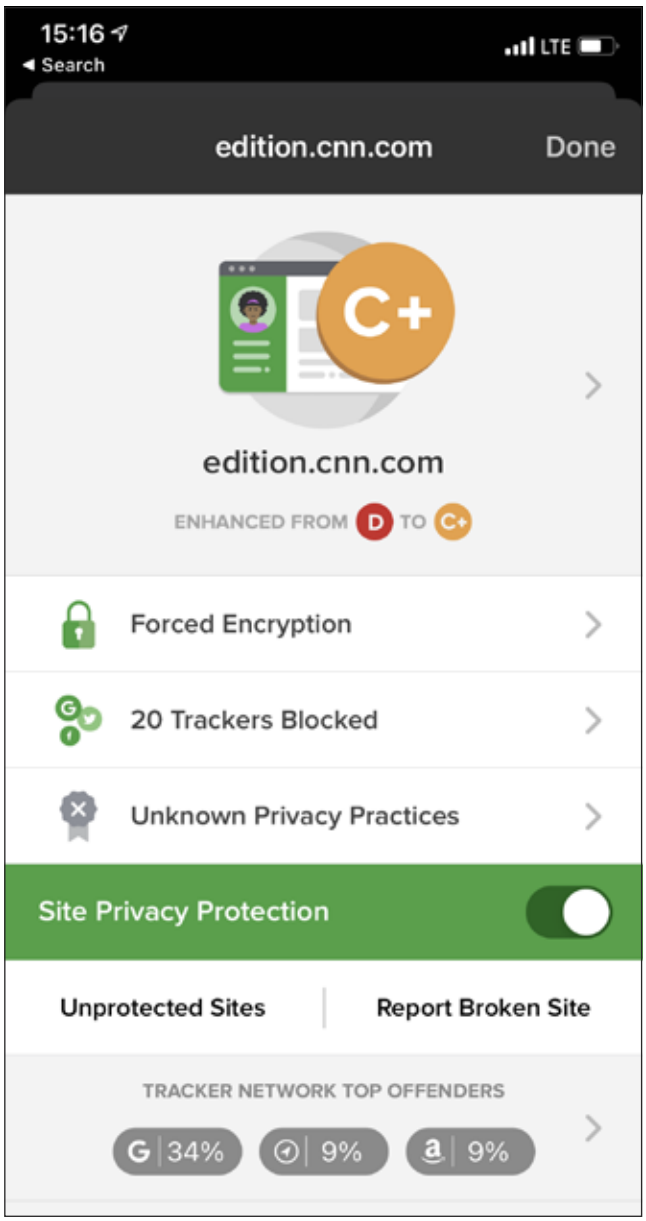
súkromia. To platí aj pre prípady, že používateľ už súhlasil s predchádzajúcimi podmienkami využívania aplikácie. Google mobilné aplikácie, medzi ktoré patrí aj vyhľadávanie, prestávajú napr. využívať IDFA (Identifier for Advertisers). Ten mal za úlohu identifikovať zariadenie používateľa a monitorovať dáta za účelom personalizácie reklamy (Combette 2021).
Na úrovni prehliadača v závislosti od konkrétneho typu (Firefox, Safari, Chrome ap.) je možné využívať režim anonymného prehliadania pomocou tzv. privátneho (inkognito) módu. Jednou z ďalších alternatív je využitie špecializovaného anonymného prehliadača webu, akým je napríklad Tor (2021). Tento prehliadač poskytuje proxy server, a tým aj zabezpečenie pred sledovaním, stálym dohľadom a cenzúrou. Ďalej predstavíme špecializovanú kategóriu vyhľadávacích nástrojov, ktoré aplikujú metódy anonymizácie dát a podporujú rôzne funkcie, ktoré riešia problémy spojené so súkromím a osobnými údajmi.

ANONYMNÉ A SÚKROMNÉ PRIESKUMOVÉ STROJE

Aktuálne sú dostupné rozmanité webové prieskumové stroje, ktoré neuchovávajú dáta o používateľovi a vznikli ako odpoveď na problémy súvisiace s ich ochranou. Sú známe aj pod označením anonymné alebo súkromné prieskumové stroje. Ďalej špecifikujeme vybrané funkcie nástrojov spadajúcich do tejto kategórie vyhľadávačov v kontexte ochrany súkromia. Nezaoberáme sa všeobecnými možnosťami a nastaveniami vyhľadávania ani spôsobmi zobrazovania, radenia či filtrovania výsledkov. Charakterizujeme reprezentatívne prieskumové stroje, ktoré sa najfrekventovanejšie umiestňujú v rebríčkoch aktuálne najlepších nástrojov podporujúcich anonymizáciu dát a súkromné vyhľadávanie – DuckDuckGo (2021a), Startpage (2021) a Swisscows (2020).

DUCKDUCKGO

Jedným z hlavných hráčov v oblasti súkromného a anonymného vyhľadávania je známy alternatívny metaprieskumový stroj DuckDuckGo, ktorý funguje od roku 2008. Počet dotazov zadaných prostredníctvom tohto vyhľadávača rastie exponenciálne a aktuálne (k marcu 2021) predstavuje denný priemer takmer 98 miliónov (DuckDuckGo 2021b).
Z hľadiska anonymného vyhľadávania DuckDuckGo nezbiera a nezdieľa osobné informácie. V nastaveniach nástroja je možné nastaviť POST požiadavky, ktoré nezobrazujú dotazy v prehliadači a nezasielajú ich tým pádom ani iným stránkam navštíveným z výsledkov vyhľadávania. V niektorých starších prehliadačoch je potrebné presmerovať kliky cez server DuckDuckGo, aby



Obr. 3 DuckDuckGo – zabezpečenie súkromia na navštívenom webovom sídle

dáta o dotazoch neunikli tretím stranám. Táto funkcia sa dá takisto špecifikovať v nastaveniach.
DuckDuckGo funguje nielen ako webový vyhľadávač, ale aj ako rozšírenie pre vybrané prehliadače a mobilná aplikácia, ktoré umožňujú blokovanie sietí trackerov na navštívených stránkach. Taktiež poskytuje funkciu automatického zobrazovania verzie stránky s enkrypciou a ukazuje aj skóre zo služby TOSDR (Terms of Service Didn't Read) (Protecting 2021). TODSR (2021) sleduje a hodnotí podmienky používania a bezpečnosti z pohľadu ochrany osobných údajov používateľa. Obrázok 3 ukazuje náhľad z mobilnej aplikácie DuckDuckGo, ktorá pri návšteve stránky cnn.com

Obr. 4 DuckDuckGo – zoznam trackerov

zablokovala 20 trackerov a na ďalšom obrázku (4) potom vidno, o aké spoločnosti monitorujúce údaje ide.

Napriek všetkým týmto užitočným možnostiam je potrebné upozorniť na to, že DuckDuckGo ukladá históriu vyhľadávania, ktorú však agreguje a očisťuje o dáta identifikujúce konkrétneho používateľa. Takisto monetizuje z partnerstiev s najväčšími komerčnými hráčmi na trhu ako je Amazon či Ebay, ktoré následne zbierajú anonymizované informácie o produktoch.

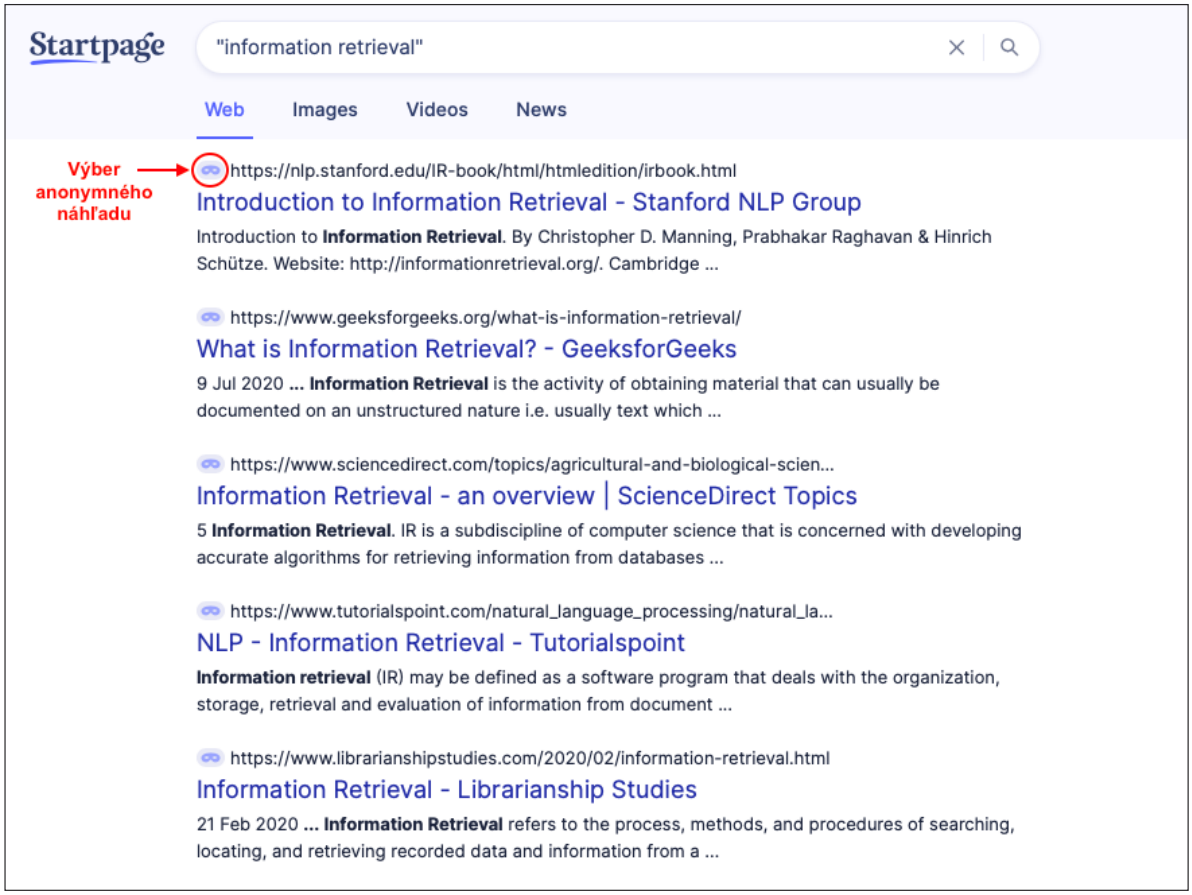
V prípade, že preferujete anonymný prieskumový stroj, ktorý je aj nezávislý, odporúčame využiť iné alternatívy spomenuté ďalej.

STARTPAGE

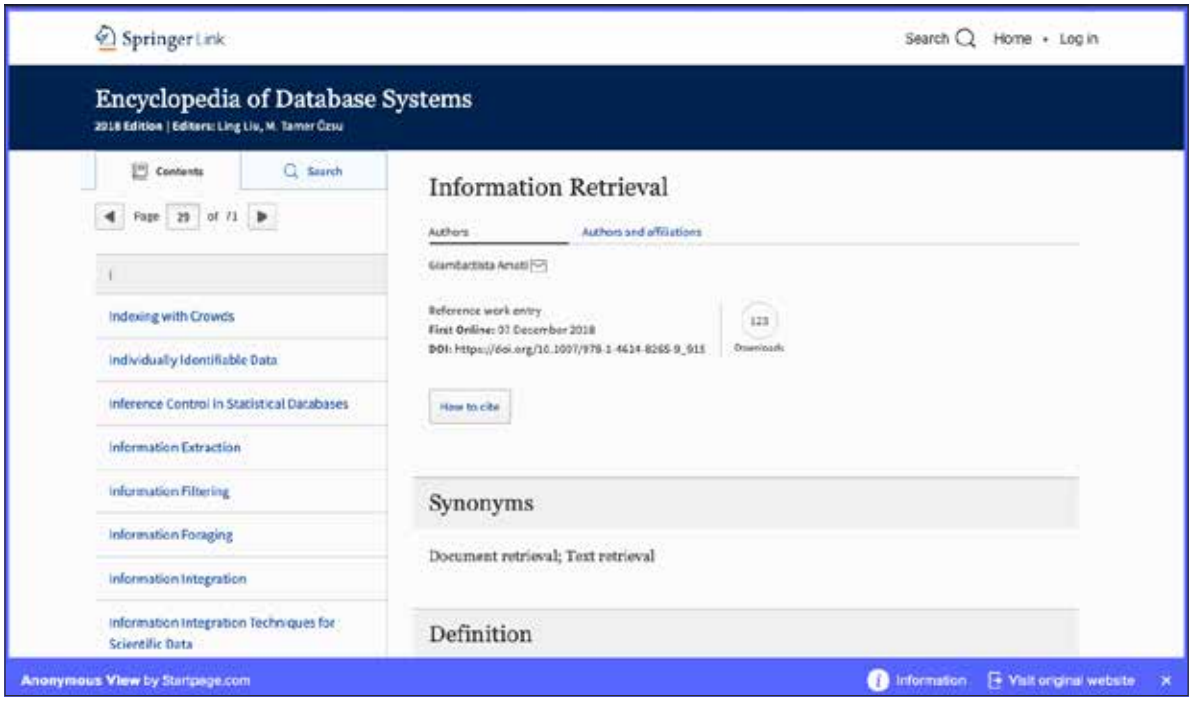
Tento prieskumový stroj na súkromné vyhľadávanie vznikol v roku 2006, ale jeho začiatky sa datujú ešte do roku 1998, kedy fungoval pod názvom Ixquick.com. Startpage poníma súkromie ako základné ľudské právo, a tým pádom neloguje, nemonitoruje, nezdieľa a hlavne nepredáva osobné dáta používateľov. Kedže Startpage nevytvára personalizované profily používateľov, umožňuje vyhľadanie širšieho záberu výsledkov. Je tak súčasťou riešenia problému filtračných bublín, ktoré zužujú výsledky vyhľadávania na základe dlhodobého sledovania správania používateľov.
Pri samotnom vyhľadávaní sú dotazy automaticky očistené o metadáta, ktoré by obsahovali napríklad IP adresu a ďalšie identifikujúce informácie. Anonymizovaný dotaz je následne zaslaný do Google a používateľovi vráti výsledky samozrejme bez logovania.
Na stránke s výsledkami Startpage zobrazuje vybrané reklamy, tie ale nie sú personalizované. V nastaveniach je potom možné zobrazovanie reklamy úplne obmedziť. Ďalšie nastavenia súvisiace s ochranou a bezpečnosťou predstavujú rodinný filter, výber metódy HTTP požiadaviek a regiónu servera. Takisto je dostupná aj funkcia bezpečnostných odporúčaní, po ktorých aktivácii Startpage používateľa upozorňuje na stránky s malvérom. Tieto nastavenia spolu s ďalšími možnosťami sa dajú uložiť ako cookie prehliadača alebo pod vlastnou URL adresou.
Tento nástroj poskytuje nielen ochranu pri vyhľadávaní výsledkov, ale aj po výbere a kliknutí na výsledky vyhľadávania. Na tento účel je dostupná funkcia anonymného náhľadu (obr. 1), ktorý poskytuje zabezpečené prezeranie webových stránok (obr. 2).

SWISSCOWS

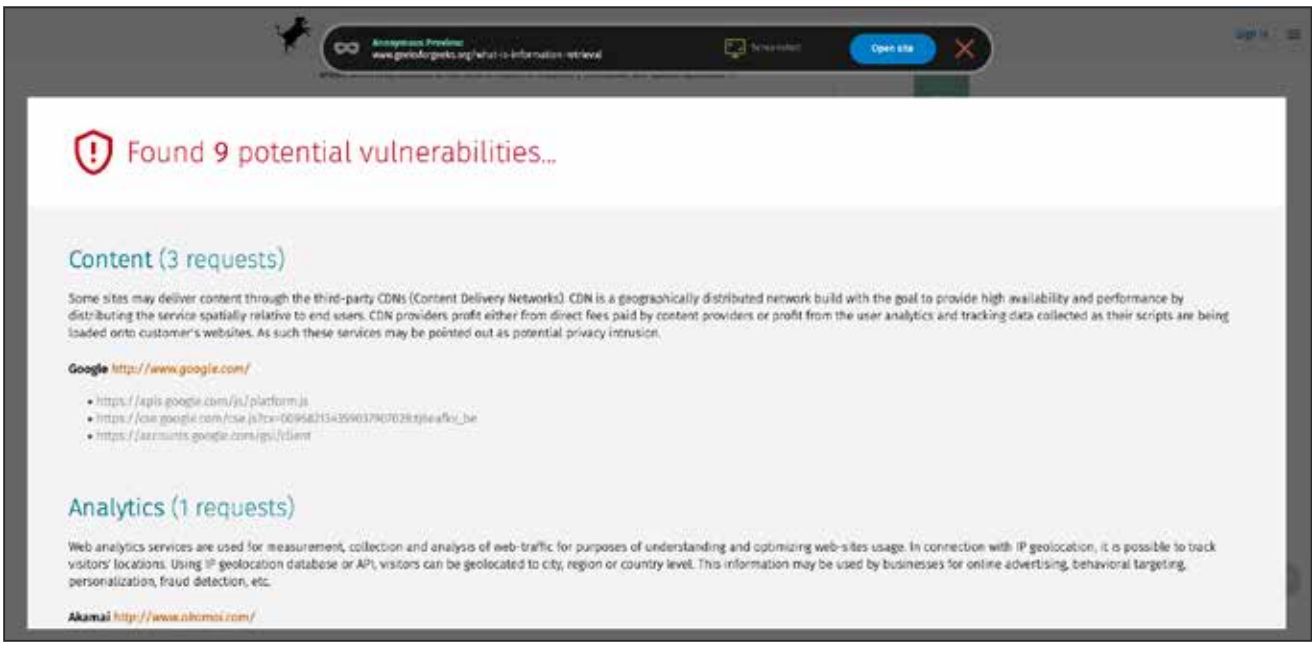
Švajčiarsky prieskumový stroj Swisscows nezbiera žiadne osobné informácie používateľov vrátane IP adresy, informácií z prehliadača či zariadenia. Takisto neukladá a následne neanalyzuje dotazy. Jediné, čo zbiera, je údaj o kumulatívnom počte požiadaviek na vyhľadávanie za daný deň.
Tento prieskumový stroj využíva vlastné servery umiestnené vo švajčiarskych Alpách, a tým pádom sa spolieha na infraštruktúru nezávislú od tretích strán.
Ich biznis model je založený na zobrazovaní reklám na základe zadaných kľúčových slov bez ďalšej personalizácie. Vyhľadané výsledky sa podobne ako to



Obr. 1 Startpage – výber možnosti anonymného náhľadu



Obr. 2 Startpage – anonymné prezeranie webovej stránky ako výsledku vyhľadávania



Obr. 5 Swisscows – identifikácia a zobrazenie potenciálnych hrozieb po kliknutí na anonymný náhľad

je pri StartPage dajú zobraziť v anonymnom náhľade a navyše aj identifikuje konkrétne hrozby a zoznam trackerov (obr. 5), ktoré môže používateľ ďalej analyzovať.

ZÁVER

EÚ má aktuálne najrozvinutejšie právo na ochranu súkromia vrátane známeho GDPR z roku 2016, ktoré efektívne reguluje bezpečnosť dát. Aj z toho dôvodu sa zlepšuje pozícia samotného používateľa, ktorý začína mať stále väčšiu kontrolu nad osobnými dátami. V kontexte legislatívy EÚ má každý občan právo na vymazanie výsledkov vyhľadávania, ktoré obsahujú osobné dáta a nie sú už aktuálne či relevantné. Toto právo sa nazýva ako „právo na zabudnutie". Takisto bol publikovaný názor európskych autorít v oblasti ochrany dát v kontexte anonymizačných techník (Opinion 2014), ktorý priamo nadväzuje na návrh algoritmov informačného prieskumu.

Na trhu je dostupných stále viac riešení na zabezpečenie súkromia a bezpečnosti osobných údajov na rôznych úrovniach. Spolu s novovznikajúcimi vyhľadávacími technológiami vznikajú aj nové príležitosti na ďalšiu legislatívnu reguláciu. Vidíme stále veľký priestor na edukáciu v rámci digitálnej a informačnej gramotnosti samotných používateľov, pretože aktuálne je kontrola nad údajmi aj v ich rukách. Používatelia aktuálne rozhodujú, či budú využívať konkrétne služby a súhlasiť s podmienkami ich používania, v rámci ktorých sú definované aj procesy monitorovania a spracovania osobných údajov či ich potenciálneho zdieľania s ďalšími stranami. Takisto je nevyhnutné, aby poznali nielen rozmanité nástroje na anonymizáciu dát, ale aj aktuálnu legislatívu v tejto oblasti. Len na základe pochopenia našich možností a existujúcich rámcov budeme vedieť robiť uvedomelé a racionálne rozhodnutia, ktoré majú vplyv na našu slobodu, ale aj na demokraciu v spoločnosti ako takej.

Zdroje

AHMAD W. U., CHANG K., WANG H. 2018. Intent-aware Query Obfuscation for Privacy Protection in Personalized Web Search. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18) [online]. [cit. 2021-03-02]. Association for Computing Machinery, New York, NY, USA, 285–294. Dostupné na: https://doi.org/10.1145/3209978.3209983

APP Tracking Transparency. In: Apple Developers [online]. [cit. 2021-03-08]. Dostupné na: https://developer.apple.com/documentation/apptrackingtransparency

ARRINGTON, M. 2006. AOL Proudly Releases Massive Amounts of Private Data. In: TechCrunch [online]. [cit. 2021-03-03]. Dostupné na: https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/

AUXIER B., RAINIE L., ANDERSON M., PERRIN A., KUMAR, M., TURNER E. 2019. Confused and Feeling Lack of Control Over Their Personal Information. In: Pew Internet Research Center: Internet & Technology [online]. [cit. 2021-03-03]. Dostupné na: https://www.pewresearch.org/internet/2019/11/15/ameri-

cans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/
BOUTET A., PETIT, A., MOKHTAR, S. B., LAPORTE, L. 2016. Leveraging Query Sensitivity for Practical Private Web Search. In: Proceedings of the Posters and Demos Session of the 17th International Middleware Conference (Middleware Posters and Demos '16) [online]. [cit. 2021-03-03]. Association for Computing Machinery, New York, NY, USA, 5–6. Dostupné na: https://doi.org/10.1145/3007592.3007595
COMBETTE, C. 2021. Preparing our partners for Apple's iOS 14 policy updates. In: Google: Ads&Commerce Blog [online]. [cit. 2021-03-08]. Dostupné na: https://blog.google/products/ads-commerce/preparing-developers-and-advertisers-for-policy-updates/
DUCKDUCKGO. 2021a. Homepage [online]. [cit. 2021-03-15]. Dostupné na: https://duckduckgo.com
DuckDuckGo Traffic. 2021b. In: DuckDuckGo.com [online]. [cit. 2021-03-15]. Dostupné na: https://duckduckgo.com/traffic
FEIGENBAUM, J., 2012. Privacy, Anonymity, and Accountability in Ad-Supported Services. In: IEEE Symposium on Logic in Computer Science [online]. [cit. 2021-03-05]. Dostupné na: http://doi.org/10.1109/lics.2012.10
HONG, Y., VAIDYA, J., LU, H., WU M. 2012. Differentially private search log sanitization with optimal output utility. In: Proceedings of the 15th International Conference on Extending Database Technology (EDBT '12) [online]. [cit. 2021-03-03]. Association for Computing Machinery, New York, NY, USA, 50–61. DOI:https://doi.org/10.1145/2247596.2247604
Opinion 05/2014 on Anonymisation Techniques. 2014. In: European Commission [online]. [cit. 2021-03-18]. Dostupné na: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
Protecting your personal data has never been this easy. 2021. In: Spread Privacy [online]. [cit. 2021-03-15]. Dostupné na: https://spreadprivacy.com/privacy-simplified/
PURCELL K. et al. 2012. Search Engine Use 2012. In: Pew Internet Research Center: Internet & Technology [online]. [cit. 2021-03-02]. Dostupné na: https://www.pewresearch.org/internet/2012/03/09/search-engine-use-2012/
STARTPAGE. 2021. Homepage [online]. [cit. 2021-03-10]. Dostupné na: https://www.startpage.com/
ŠUŠOL, J. 2019. Súkromie a jeho modifikácie v digitálnom prostredí. In: ITLib [online]. č. 4, roč. 23. [cit. 2021-03-05]. Dostupné na: https://itlib.cvtisr.sk/wp-content/uploads/docs/Šušol.pdf
Swisscows. c2020. Homepage [online]. [cit. 2021-03-17]. Dostupné na: https://swisscows.com/
TODSR. 2021. Homepage [online]. [cit. 2021-03-15]. Dostupné na: https://tosdr.org
TOR. 2021. Homepage [online]. [cit. 2021-03-03]. Dostupné na: https://www.torproject.org/download/
We don't collect or share personal information. [bez dátumu]. In: DuckDuckGo [online]. [cit. 2021-03-15]. Dostupné na: https://duckduckgo.com/privacy

Príspevok bol spracovaný v rámci riešenia projektu VEGA 1/0360/21 Sociálne reprezentácie etických výziev digitálnej informačnej revolúcie.