Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra managementu

Analýza sběru osobních údajů na internetu

Bakalářská práce

Autor: Martin Soukup
Studijní obor: Informační management

Vedoucí práce: prof. PhDr. Marek Franěk, CSc., Ph.D.

Hradec Králové duben 2023

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 21.4.2023 Martin Soukup

Poděkování:

Děkuji vedoucímu bakalářské práce prof. PhDr. Marku Fraňkovi, CSc., Ph.D. za metodické vedení práce a inspirování ke zvolení tohoto tématu. Rovněž děkuji Oddělení propagace a komunikace UHK za odbornou konzultaci.

Anotace

Bakalářská práce se věnuje problematice sběru osobních údajů uživatelů webových stránek. Práce nabízí pohled na vývoj webových technologií, od začátků této technologie až po současnost, v souvislosti se zpracováním uživatelských údajů. Dále představuje opatření, omezení a nástroje, které mají za účel chránit uživatele před subjekty nadměrně sbírající a využívající tyto údaje. Analýza je provedena formou inspekce sbíraných osobních údajů a třetích stran přítomných na webových stránkách a následným porovnáním s prohlášeními společností. Analýza je provedena na vzorku vybraných stránek pokrývajících více různých typů obsahu. Analyzovaná data ukazují na přítomné nedostatky v ochraně osobních údajů, na které lze aplikovat řešení pomocí dostupných technologií. Stejně tak již existují nástroje umožňující silnější ochranu osobních údajů uživatelů.

Annotation

Title: Analysis of Personal Data Collection on the Internet

The bachelor thesis aims to investigate collection of personal user data on websites. The thesis firstly looks at the development of web technologies in relation to data collection since its invention up until now. Following up with evaluation of measures, restrictions and tools which are used in defense against excessive data collection and usage for profit. The analysis consists of inspecting personal data collection and invited third parties and comparing it against the companies' privacy policies and statements. The analysis is performed on selected number of websites spanning different types of content. The results suggest that the current data protection measures are insufficient in some ways that could be solved with application of programmatic solution. The users, however, do already have the means available to them to enhance their privacy control.

Obsah

Seznam obrázků

Seznam tabulek

1 Úvod

V současné době probíhá celosvětový rozvoj neuvěřitelně rychlým tempem, a to především v oblastech týkajících se technologií na internetu. Máme k dispozici velmi efektivní a dostupné vyhodnocovací prostředky, umožňující provoz mnohých dnešních aplikací a webů, které by bez nich nebyly možné. V mnohém nástup těchto nástrojů ulehčuje každodenní život a fungování společnosti, může však působit i opačným účinkem. Zcela jisté je, že příchod těchto nových technologií vyhodnocujících každý údaj přinesl změnu.

Díky rapidnímu vývoji internetových technologií, obzvláště v citlivé oblasti jako je zpracování osobních údajů, nestačí mnohdy relativně pomalu reagující vlády, úřady a regulátoři reagovat a nastavit účinná pravidla, která by nastavila rovnováhu mezi oběma stranami, jak je tomu v jiných oblastech.

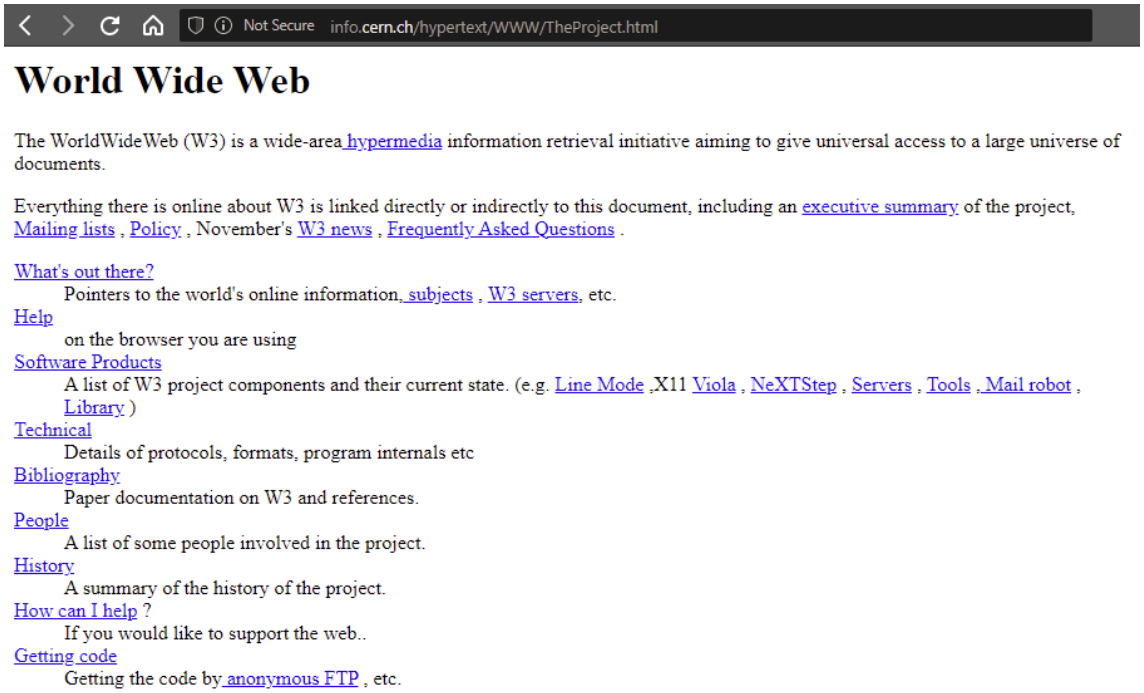
Zaměřením této práce je ověřit, zdali jsou současné opatření účinné a přináší pro uživatele prospěch v rámci ochrany soukromí. Nehledě na konkrétní definice jednotlivých státních či unijních opatření, uživatelé nemusí být dostatečně chráněni. Případně, nejsou-li opatření příliš přísná, natolik, že dávají společnostem možnost obrátit pozitivní dopad v negativní v očích veřejnosti jako je to v případě únavy z cookies bannerů. Pakliže jsou tato opatření dobře nastavena, jak fungují v praxi, a jestliže nejsou, existují cesty, jak se mohou uživatelé chránit proti nezměrnému množství sbíraných údajů?

2 Historie webových stránek

Webové stránky, technologie a internet se neustále vyvíjí. Pro lepší pochopení současné mentality okolo webových technologií a běžně využívaných technik je podstatné se podívat na vývoj této oblasti v širším kontextu. Podstatným faktorem, přispívajícím k současné podobě webu jsou totiž jeho počátky. Způsob, jakým využíváme webové stránky se v mnohém změnil a spolu s tím i funkce, které web nabízí.

2.1 Počátky webu

V roce 1989 přichází Tim Berners-Lee s novým systémem pro správu informací WorldWideWeb [1]. Berners-Lee vytvořil prohlížeč a zároveň editor webových stránek postavený na novém značkovacím jazyku HTML, ten vycházel z již používaného stylovacího formátu dokumentů v CERNu SGML [2]. Přenos dat obstarával komunikační protokol HTTP, který je s úpravami používán dodnes. První prohlížeč WorldWideWeb byl úspěšný, ale pouze v rámci CERNu, neboť fungoval jen na operačním systému NeXT [3]. Ačkoliv byl první, byl to až prohlížeč internisty CERNu Nicoly Pellowa, který pomohl adopci webu na jiných systémech, a i mimo CERN. Jeho prohlížeč běžel jen v příkazovém řádku a měl jen minimum funkcí oproti výtvoru Bernerse-Lee, ale byla to jeho přístupnost, která vyhrála nad sice robustnějším ale nedostupným WorldWideWebem. Po vzoru tohoto jednoduchého výtvoru, začali vznikat ostatní nezávislé prohlížeče, které pomáhali web rozšiřovat dále. Každý z nich také přinesl nějakou novou funkci či odlišné chování. Vše přitom stavělo na základech položených Bernersem-Lee HMTL a http protokolu [4].



Obr. 1 První webová stránka (kopie z roku 1992).
Zdroj: http://info.cern.ch/hypertext/WWW/TheProject.html.

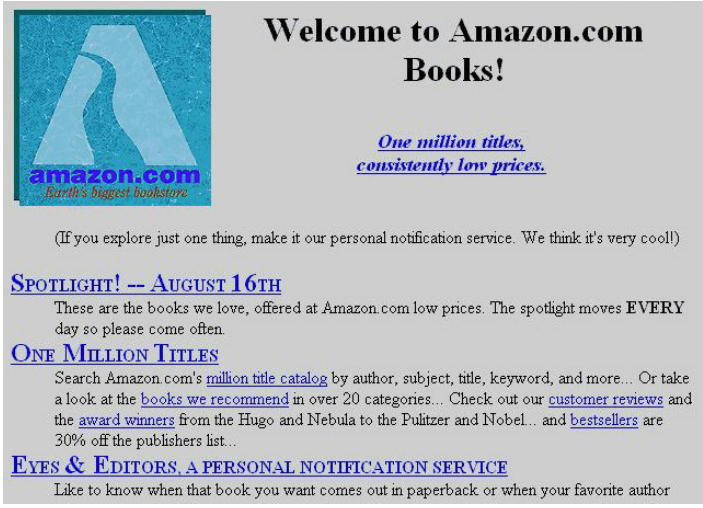
Rané webové stránky byly vytvořené na jednoduchých principech a za poměrně jednoduchým účelem – původně tomu byla správa dokumentů. Za jednoho z průkopníků se označuje prohlížeč Mosaic, který krom toho, že běžel v pouze jediném okně, přinesl do dokumentů vnoření obrázků přímo mezi text. Tento krok a další nové funkce se jednoznačně podílely na směru, kterým se webové prohlížeče a technologie na jejich pozadí ubíraly [4]. S příchodem velkého pozornosti ze stran obyčejných lidí se web začal měnit, až nakonec účel, kvůli kterému byl původně stvořen téměř vymizel. U navrhování systémů takového rozsahu nelze pevně určit směr, kterým se vývoj požene, protože směr určují jeho uživatelé. Jakmile se vnímání webových stránek změnilo z pouhého čerpání statických informací na interakci se stránkami, kde je možné nakoupit zboží, číst novinky nebo poslat zprávy, celé prostředí se změnilo [5].

Začátky moderního webu s sebou, na jedné straně, nesou ohromný přísun uživatelů bez větších technických znalostí, a na druhé straně růst komercionalizace webu. Tyto dvě strany jdou v některých ohledech ruku v ruce, a v jiných se ale rozchází. Právě rozložení mezi těmito dvěma stranami utváří směr postupu webu [6].

Okolo začátku nového milénia byla nově se objevující komerční sféra značně zaměřena na využití nového média založeného na tvorbě obsahu od uživatelů. Tento nový směr dal vznik službám jako je YouTube, Wikipedia či mnohým blogům. Tyto typy stránek byly přímo závislé na obsahu od jejich uživatelů, velká změna od jednostranného příjmu informací na původním webu. Přechod z *„read-only“* (pouze pro čtení) webů na dynamické *„read-write“* (pro čtení i zápis) weby interagující s uživatelem a využívající uživatelský obsah se označuje jako přechod z Webu 1.0 na dosud trvající éru označovanou Web 2.0 [5].

Tento postupný přechod s sebou logicky přinesl mnoho změn, jelikož byl obsah tvořen uživateli a webová stránka byla pouze jakousi platformou na platformě, bylo potřeba uchovávat tyto uživatelské údaje ze stránek. V začátcích se jednalo hlavně o obsah vytvořený uživateli při používání stránky, a anonymní agregované údaje k analytickým účelům jako třeba jednoduchá počítadla návštěv webové stránky [5].

Jako jeden z prvních, kteří využili všech nově dostupných nástrojů na internetu byl internetový obchod s knihami Amazon. V té době byli webové stránky Amazonu primitivní, ale již umožňovali uživatelům založení účtu, díky čemuž se jim mohla zobrazit personalizovaná nabídka. Aby Amazon na nové, ne moc rozšířené platformě uspěl, musel naplno využít všech výhod oproti tradičním obchodům, které web nabízí. Tou byly například historie objednávek, snadné vyhledávání, nebo třeba online recenze od ostatních uživatelů [9].



Obr. 2 Úvodní stránka Amazon.com z roku 1995.
Zdroj: https://www.webdesignmuseum.org

Vedení internetové společnosti spolu nese i velký risk, který na rozdíl od Amazonu, ne každému vyšel. Od roku 1996 vznikaly různé internetové společnosti provozující prodej skrze webové stránky. Ačkoliv byly zaměřením odlišné, většinu potkal na přelomu milénia stejný osud; bankrot. Společným prvkem těchto společností byl nesprávně zvolený směr ve vztahu k cílové skupině. Ačkoliv některé společnosti byly v určitých ohledech velice inovativní, avšak neměli udržitelný model financování, a tak nedokázali přežít hromadný odchod investorů na přelomu tisíciletí [10].

Právě kolem roku 2000 můžeme pozorovat větší obezřetnost a ústup od webových služeb založených pouze na nápadu, rozvíjeného bez ohledu na skutečný dosah a udržitelnost společnosti. Amazon, eBay a jemu podobní přežili pravděpodobně díky zkoumání uživatelských dat a zjišťování, které funkce, obsah, principy fungují, na základě čehož přizpůsobovali své zaměření [8][10].

2.1.1 Web vybudovaný na reklamách

Jeden z prvních webů, který přišel na potenciál modelu pokrytí nákladů na provoz za pomocí webové reklamy formou bannerů umístěných na stránce byl roku 1994 internetový magazín *Hotwired,* tehdy byla tato reklama nesmírně úspěšná, co se týče míry prokliku. To však netrvalo dlouho, neboť bannerovou reklamu začalo využívat mnoho ostatních webů. Bannerová reklama se rychle rozmohla a začala být pro uživatele otravná, protože byla uživatelům dokola zobrazována tatáž reklama, efektivita a cena tak šla prudce dolů [11]. Na to odpověděly v roce 1995 a 1996 svým produktem firmy WebConnect a DoubleClick. Ta první se zaměřila na cílení reklamy podle typů návštěvníků určitých webů. Zatímco DoubleClick, nyní vlastněn Googlem, byl jedním z prvních reklamních systémů, které se v té době objevily, a umožňoval inzerentům měřit účinnost jejich reklamy napříč několika weby [17][18][19].

Jak množství obsahu na webu rostlo, zvyšoval se i význam webových vyhledávačů, neboť právě oni byly častokrát tím neviditelným prostředníkem mezi uživatelem a kýženým obsahem. Protože webové vyhledávače vyžadují vstup od uživatele, aby vůbec plnily svou funkci. Netrvalo dlouho a na přelomu milénia se uživatelská vyhledávání začala zpracovávat a využívat k zobrazení reklamy související

s hledaným výrazem. K uživatelům se takto dostala poměrně relevantní reklama a webový vyhledávač se prodejem takovýchto placených umístění mohl uživit. Tento model se ukázal jako výhodný, neboť umožňuje „bezplatný“ provoz pro uživatele a zároveň uspokojí jak provozovatele stránky, tak i inzerenty. Není s podivem, že na tomto principu funguje vyhledávání dodnes. Samozřejmě, nic není bez kompromisů a tento model zpeněžení výsledků hledání s sebou přinesl řadu obav ohledně kompromitace kvality výsledků. Střetávají se tu totiž protilehlé strany, kdy na jedné straně je v zájmu vyhledávače zobrazovat výsledky, které mu přináší větší zisk, na straně uživatele jsou mnohdy tyto výsledky nekvalitní a irelevantní. Tento střet zájmů zmiňuje bývalý ředitel, dnes nejpoužívanějšího vyhledávače, Googlu Lawrence Page ve studii z roku 2010 [7][17].

Možná na první pohled překvapivé tvrzení právě od takovéto firmy. Je třeba se ale podívat na jejich tehdejšího konkurenta Goto.com. Byla to první firma nabízející placenou pozici ve vyhledávání na bázi *pay-per-click* (placení za kliknutí uživatele). Tento krok šel logicky přímo proti užitečnosti vyhledávání pro uživatele. V Googlu si toto uvědomily a vyvinuly systém AdWords, který sponzorované výsledky vyhodnocoval nejen s návazností na to, kdo zaplatil nejvíce ale velkou váhu dal také relevantnosti k hledanému výrazu. Na sponzorované výsledky tak lidé klikali více na Googlu než jinde. V roce 2003 Google zavedl AdSense, který se již zabýval umístěním reklam přímo na stránkách jiných společností. U nich opět využili kontextu a množství informací o návštěvníkovi, čímž umožňovali inzerentům cílit reklamu s dosažením větší účinnosti než konkurence. Stylem, jakým společnost Google dosáhla dominance na poli reklamy navždy proměnila způsob monetizace webu. Všem společnostem ukázali cestu, která je postavena na neustálém zlepšování cílení obsahu, reklam, služeb [7][17].

Zajímavé poznatky o situaci v té době nabízí studie J. D. Greerové z roku 2004. Podle které byly veškeré reklamy placeny z jedné čtvrtiny ostatními webovými společnostmi. Dále ve své analýze zjistila, že reklamu si platily spíše lokální a menší firmy, nežli velké společnosti [12]. Ze vzorku testovaných webů zpravodajských a televizních stanic se ukazuje, že v té době mnohé velké společnosti ještě neviděly plný potenciál webové inzerce a jeho benefity oproti té tradiční [13].

To se mělo již brzy změnit. Neboť na webovém prostředí se právě objevil nový typ webových stránek. Sociální sítě jsou stránky, ve kterých je uživatel motivován k zveřejňování osobních údajů, zážitků, myšlenek, obsahu s ostatními uživateli ve své „sociální síti“ či veřejně. Navazuje kontakt, buduje komunity a přispívá v nich. Všechny tyto akce produkují také množství informací o uživateli pro provozovatele sociální sítě. Provozování jakékoliv platformy na internetu není zadarmo, a proto musí společnosti hledat zdroje, jak náklady pokrýt, ideálně zásahu do použitelnosti webu. Facebook ihned od svého počátku v roce 2004 pokrýval náklady bannerovou reklamou, bez většího cílení. Tak tomu bylo až do roku 2007, kdy Facebook zavedl Facebook Ads, od té doby jsou jejich reklamy cílené. S postupem času přicházely ze stran sociálních sítí, ale i jiných typů stránek, nové způsoby, jak využít informací a integrovat reklamy co nejvíce mezi funkčnost stránek pomocí „nativních reklam“ [18]. Zjistili, že čím více je reklama „užitečná“ nebo skrytá, tím méně se budou uživatelé zajímat o negativní dopady umožňující takovéto reklamy. Facebook, Twitter, YouTube a další jsou motivováni ke zlepšování práce s údaji uživatelů, aby mohly poskytovat lepší cílení pro inzerenty nebo jiné subjekty [14]. Takovéto jednání, kdy se společnosti opírají o vytěžení hodnoty z osobních údajů uživatelů však může vést ke kompromitaci soukromí. Pokud se rovnováha nenastaví s ohledem na obě strany, jako tomu bylo například v případě Google AdWords, kdy užitečnost se setkala se ziskem, aniž by jedna strana výrazně utrpěla [19].

Ve své podstatě jsou sociální sítě pravým opakem webových stránek v době Webu 1.0, samy o sobě totiž žádný obsah nemají, ten vzniká až od uživatelů [5]. Tato kompletní změna principu fungování webu se nemohla udát, bez technologických pokroků v oblasti webových technologií.

3 Technologický vývoj

Tato kapitola se věnuje technologické stránce vývoje webových stránek, a jak tyto nové technologie skončily častokrát použity pro jiný účel, než pro který byly původně zamýšleny.

3.1 Identifikace pomocí cookies

Technologie takzvaných *HTTP cookies* byla prvně vydána už v roce 1994 v prohlížeči Netscape Navigator. Úkol cookies je velmi jednoduchý, mají za cíl uchovávat stav proměnných na zařízení uživatele. Důvodem ke vzniku bylo integrovat do http technologii, která by umožnila na straně uživatele uchovávat data o obsahu nákupního košíku pro stránky internetových obchodů. Cookies neobsahovaly žádné vestavěné údaje ani identifikátory o uživateli, u kterého jsou uloženy. Trvalo pár let, než se začala hojně využívat vývojáři webových stránek. To bylo také zapříčiněno různými standardy používání cookies prohlížeči [15]. Standardizace zápisu a chování cookies se napříč všemi hlavními prohlížeči dostalo až v roce 2011 [16].

Použití cookies s sebou také nese rizika porušení soukromí, na tyto rizika poukázala skupina *Internet Engineering Task Force* již 3 roky po uvedení cookies do Netscape Navigatoru. IETF navrhovala, aby byly cookies v původním nastavení vypnutá, nebo dokonce kompletně zakázaná, a uživatel by si je tak musel zapnout. Uživatel by tak musel být srozuměn s tím, že může být sledován skrze tuto technologii a explicitně ji povolit. V té době byly však cookies už hojně používány právě pro reklamní účely zasahující do soukromí uživatelů jejich sledováním skrze cookies [20].

3.2 Frontend a layout

Ačkoliv to tak nemusí na první pohled působit, i frontend – tedy pro uživatele viditelná část webu, hraje podstatnou roli ve vývoji webových stránek. Tím, že webové stránky dokázaly zobrazit a poskytnout více funkcí, chtěli tvůrci webů nabízet stále něco nového. Logickým krokem bylo využití údajů uživatelů pro interakci nebo k analytickým účelům.

3.2.1 Webový design

V úvodních letech webu nebyl design ani vizuální úprava příliš rozvinutým oborem. Důvodem byl původní účel jazyku HTML, který se měl starat pouze o strukturu dokumentů/webových stránek, u kterých nebylo příliš co vizuálně upravovat. Směr webu se však vyvíjel právě cestou vizuálně komplexnější prezentace a interakce s uživatelem a pro tento účel bylo pouhé HTML nedostačující [21].

Před příchodem jakéhokoliv dedikovaného nástroje bylo nutné pro vizuální úpravu využít kliček a používat části HTML pro věci, pro které nebyly vytvořeny. Příkladem může být umístění prvků na stránce pomocí rozložení celé stránky do několika vnořených tabulek. Situaci komplikovala také roztříštěná vzájemná podpora funkcionalit v nejpoužívanějších prohlížečích. Každý prohlížeč měl nějakou funkci navíc, kterou ostatní nepodporovaly, to byl značný nápor na programátory, neboť museli vybalancovat podporu jednoho a zároveň funkčnost na druhém [21].

Očekávanou změnu přináší představení CSS (Cascading Style Sheets), díky tomuto jazyku bylo možné stránky „stylovat“ stránky bez zásahu do HTML struktury. První verze CSS přišla v roce 1996, ačkoliv funkcí nebylo moc, otevřela do té doby nevídané možnosti pro webové stránky. Zanedlouho přišla v roce 1998 verze CSS2, která umožnila tvořit komplexnější rozložení a stylování. V době vydání druhé verze CSS byl tento jazyk již hojně využíván ze stran vývojářů [22]. Princip oddělení struktury od vizuálních prostředků je používán dodnes.

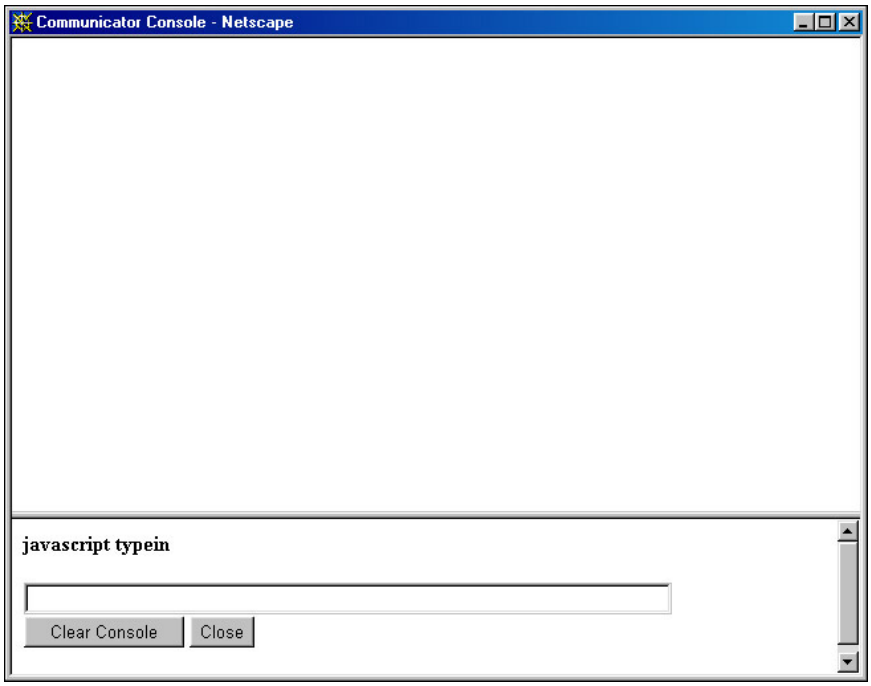
Přechodem z Webu 1.0 na 2.0 se všichni snažili přijít na nejlepší principy pro design a rozložení webových stránek. Výsledkem toho bylo mnoho nepřehledných webů, které se snažily na své stránce hlavně zaujmout. Webové stránky byly zcela nový formát, na který nebyl nikdo připravený. Někteří se snažili aplikovat principy z tisku na webový formát, mnohdy však neúspěšně, neboť obsah na webu je zobrazením variabilní, (různé rozlišení) avšak tištěná média mají pevně dané rozměry. Dalším aspektem, který musí vývojáři při tvorbě webu zvážit je náročnost na načtení. V začátcích webu byla průměrná rychlost internetového připojení velmi malá (zpravidla přes vytáčené připojení). Kvůli tomu bylo nutné optimalizovat velikost dat na webu, aby se načetl alespoň do 10 sekund. Nejúčinnějším způsobem je omezení použití multimediálního obsahu jako jsou obrázky (či dnes už videa)

ve vysoké kvalitě. Rovněž se prokázalo, že stránky působící přehlednějším a jednodušším dojmem mají větší šanci návratu návštěvníka stránky [24][51].

CSS má jasný účel a ten plní – upravit vzhled stránek. Nedokáže však interagovat s uživateli, k tomuto účelu bylo navrženo mnoho různých skriptovacích jazyků, zdaleka největší adopce se dočkal jazyk *Javascript*. Vytvoření jazyka působícího na přední uživatelské straně zobrazení stránky (a později také nepřímo na pozadí) bylo zcela klíčové pro rozvoj dynamicky se měnících webů [25].

3.2.2 Interaktivita a JavaScript

Kolem roku 1995 kolovala mezi vývojáři webových prohlížečů (převážně Internet Explorer od Microsoftu a Netscape Navigator od společnosti Netscape) idea o implementaci skriptovacího jazyka do prohlížečů. Tradičně používané jazyky jako C++ či Visual Basic byly pro zamýšlený účel vkládání drobných dynamických prvků a přidávání interaktivity na web zbytečně komplexní. Proto bylo cílem vytvořit jednodušší dynamický skriptovací jazyk, který by se hodil pro spojování HTML prvků a vnášení interaktivity. Tento krok šel proti původnímu návrhu webu od Tima Bernerse-Leeho z roku 1991. Podle Bernerse-Leeho a W3C (World Wide Web Consortium) má být web postaven pouze na deklarativním HTML [25].



Obr. 3 Raná verze JavaScriptové konzole.
Zdroj: http://www.yaldex.com/javascript_tutorial_3/LiB0072.html

Tento názor nebyl mezi ostatními vývojáři sdílen ani vyslyšen, neboť věděli, že web směřuje směrem k interaktivitě mezi webem a uživateli. Společnost Netscape proto roku 1995 implementovala svůj skriptovací jazyk Mocha. První prototyp vytvořil Brendan Eich za pouhých 10 dní. Původní schéma bylo vnést do HTML možnost drobných interakcí pomocí skriptovacího jazyka a pro implementaci větších aplikací by byl jazyk kompatibilní s „velkým“ a tradičním jazykem Java. Mocha bylo jen dočasné jméno s plánem vydání jazyka pod názvem JavaScript [25].

Co byl jen malý skriptovací jazyk se rychle vyvinulo v praktický standard, a i ostatní prohlížeče byly nuceni jazyk podporovat v zájmu interoperability. Jelikož jazyk vznikl pod křídly pouze jednoho webového prohlížeče, a neměl ani specifikaci, musely ostatní prohlížeče implementovat JavaScript formou reverzního inženýrství. Konkurenční prohlížeče byly nespokojené a formálního standardizování se JavaScript dočkal roku 1997 pod organizací ECMA a názvem ECMAScript [25].

JavaScript brzy přerostl svůj původní cíl spočívající v přidání drobných interaktivních prvků na stránky. Zdaleka největším milníkem pro JavaScript byla implementace AJAX (Asynchronní JavaScript a XML). AJAX umožňoval programátorům dynamicky načíst obsah do již načtené stránky pomocí nového rozhraní nazvaného XMLHTTP. Do té doby bylo možné přidávat obsah na webové stránce pouze odesláním požadavku na server a znovunačtením stránky s novými daty. Díky této vlastnosti se z JavaScriptu stal skriptovací jazyk schopný vybudování bohatých interaktivních webových aplikací, které doposud nebyly možné [25].

Na principech AJAXu vzniklo mnoho frameworků pro JavaScript. Framework je prostředníkem mezi jazykem (v tomto případě JavaScript) a programátorem, účelem je poskytnout nástroje pro rychlejší vývoj a přehlednější správu webové aplikace. JavaScriptové frameworky typicky integrují práci s AJAXem pro jednostránkovou aplikaci, která se chová jako by byla vícestránkovým webem. Nemusí však načítat celou novou stránku při přechodu mezi stránkami, místo toho se pouze dynamicky přidává nový obsah. Typicky jsou tyto frameworky využívány pro budování uživatelských rozhraní, kde by bylo znovunačtení stránky kazilo uživatelský zážitek, místo toho se ponechávají například jen ovládací prvky a obsah se vymění [26]. Díky plynulému načítaní obsahu, odstranění stránkování je pro

uživatele mnohem lehčí stát se pohlcen bezmyšlenkovým přísunem obsahu. Který může vést k nadměrnému používání dané služby. To může být zprvu pro provozovatele stránky pozitivní, ale dopady na celou společnost, pokud je tato stránka celosvětově rozšířená pozitivní nejsou. jako úbytek pozornosti, snížení odolnosti vůči dezinformacím, kvůli nadměrnému přísunu informací, a další [23][34].

3.3 Backend a databáze

Vývoj na přední straně (frontend) i za oponou (backend) webu musel jít ruku v ruce, aby se mohl web neustále posouvat kupředu. Vývojáři webových stránek potřebovali pro tvorbu dynamických webů nástroj, který dokáže zprostředkovat bilaterální komunikaci mezi databází/serverem a webovým uživatelským prostředím.

3.3.1 Skriptovací jazyky

Jako první byly vytvořeny nástroje umožňující tyto funkce založené na tradičních jazycích hlavního proudu jako je C, C++ nebo Java. Tyto jazyky nebyly vytvořeny pro web, potřebovaly tedy prostředníka, který by data (zpravidla z formuláře) odeslal na server pro zpracování. Díky *Common Gateway Interface* se data poslala ke zpracování v daném jazyku a výsledek byl poté zpět odeslán k uživateli [27]. S postupem času byly vytvořeny „server-side“ skriptovací jazyky, které uměly prakticky vše, co jejich staticky typovaní předchůdci, avšak značně usnadňovaly a zrychlovaly vývoj. Mohla za to jejich větší přehlednost a zaměření přímo na webové stránky. Kdežto staticky typované jazyky vyžadovaly většinou delší syntax a komplexnější vývoj, skriptovací jazyky tyto prvky vypustily a poskytly více zaměřený a vhodnější jazyk pro rychlý vývoj webových stránek [28].

V dnešní době dominuje mezi server-side skriptovacími jazyky PHP, které je používané na více než 80 % webů [29]. Tento jazyk byl původně vystavěn jako balíček nástrojů pro CGI. O PHP jako o skutečném skriptovacím jazyku lze hovořit od verze PHP 3, kdy bylo PHP kompletně přepsáno [30]. PHP a jemu podobné jazyky umožňují vybudování rozsáhlých stránek, které se dynamicky mění na základě

interakce s uživateli a jejich údaji. Příkladem mohou být sociální sítě, e-shopy nebo například e-learningové stránky [31].

V průběhu používání jazyka často vznikají frameworky, které přináší zjednodušení práce, většinou se specifickým zaměřením na určitý častý druh použití. V případě PHP je jedním z nejpoužívanějších Laravel. Díky užšímu zaměření frameworků na konkrétní oblasti webového vývoje je možné dosáhnout agilnějšího a usměrněnějšího vývoj nežli bez něj, byť bychom brali v potaz už jen určenou strukturu projektu, kterou mnohdy použití frameworku diktuje. Mnohokrát se hledí pouze na ty nejlepší a nejrychlejší (z hlediska běhu aplikace) nástroje. Přístupnost pro většinu je ale to, co posouvá web kupředu, neboť menší subjekty nemají čas ani prostředky na vývoj ve rychlejších, ale z pohledu vývoje časově náročnějších konvenčních jazycích. Použití frameworku si častokrát vybírá daň právě na rychlosti a Laravel není výjimkou, výměnou se vývojáři dostane zkvalitnění práce a mnohdy i nových, jinak těžko implementovaných funkcí [32].

PHP samozřejmě není jediným používaným server-side jazykem. Paralelně bylo vyvíjeno mnoho skriptovacích jazyků, například Node.js, Ruby on Rails, či C#. Důvodem pro použití jiného jazyku může být několik: Preference vývojáře, specifické podmínky vývoje diktující jazyk nebo nedostatečná rychlost v určitých případech použití. Právě poslední z důvodů může být důvodem sáhnout po Node.js, technologii spojující webový server se skriptovacím jazykem. Hlavní výhodou oproti rozšířenému PHP je velice rychlé zpracovávání drobných požadavků a s tím spojená vysoká škálovatelnost [33]. Za výhodu se dá považovat také zavedení jednotného jazyku JavaScript na frontend i backend.

3.3.2 Databáze

Tyto robustní skriptovací jazyky pro práci s údaji by byly zbytečné, kdyby neexistovali servery a databáze schopné s nimi držet krok.

Od počátku byly s webovými stránkami používány relační databáze pro ukládání a manipulaci s daty. Osvědčenou používanou kombinací je LAMP (Linux – Operační systém, Apache webový server, MySQL – relační databázový model, PHP/Perl/Python – server-side jazyky), tento akronym je používaný i dnes s tím

rozdílem, že již neoznačuje konkrétní technologie ale obecnou strukturu [35][52]. Typickou volbou relačního databázového modelu pro webovou stránku jsou modely založené na dotazovacím jazyku SQL (MySQL, PostgreSQL, MS SQL server, …) nebo relační databáze Oracle. Pro mnoho současných webů, je tato struktura používána dodnes [39].

Jak se webové stránky rozvíjely a rozšiřovaly, spolu s tím u velkých společností na webu narůstal i objem sbíraných údajů. Nároky na rychlost a množství přenosu dat se neustále zvyšují. V některých případech proto nemusí být klasické relační databáze vhodným řešením. Jejich robustnost totiž komplikovala škálovatelnost napříč celou zemí. Platí to především u aplikací s velkým objemem dat, které se v průběhu mění. U těchto případů může být relační databáze nedostačující, jelikož se velmi obtížně adaptuje na změnu struktury za chodu. Odpovědí na tuto limitaci přináší distribuované databáze typu NoSQL (Not only SQL). Tyto databáze jsou flexibilní a velmi dobře škálovatelné. Použití NoSQL je vhodné, pokud se jedná o často se měnící a rychle narůstající množství dat, které však nejsou kriticky důležité, neboť NoSQL databáze ztrácí na SQL v konzistenci a bezpečnosti. Tyto limitace vyplívají z lepší horizontální škálovatelnosti NoSQL, kde není možné zajistit takové zabezpečení dat jako na vertikálním monolitickém serveru založeném na relační databázi SQL [40][41].

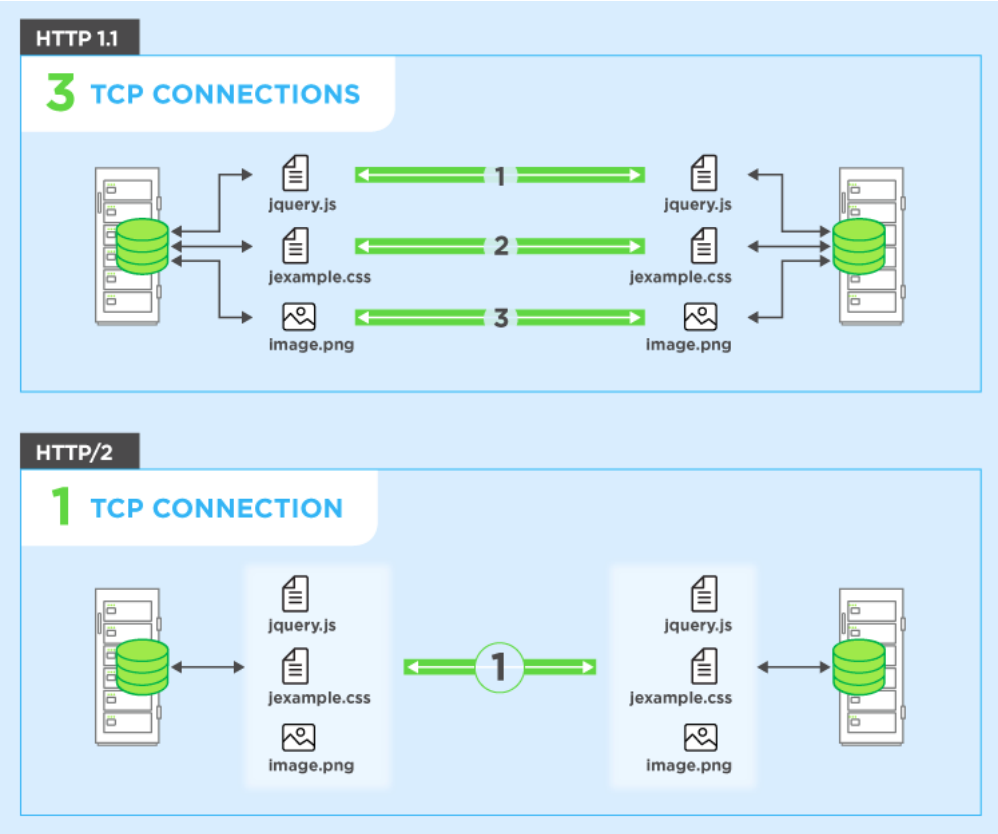
Databáze typu NoSQL zahrnují mnoho různých podkategorií databází, každá se specializuje na vyřešení jiného problému. Pro zrychlení zápisu a přístupu k často používaným informacím vznikly in-memory databáze jako je Redis. Pro pokrytí oblastí, kde pouhé vazby relačních paradigmat nestačí mohou slepé místo zacelit grafové databáze, například Neo4j. V těchto databázích jsou data místo tabulek umístěny volně a navzájem propojeny vztahy. Někdy je naopak potřeba mít volnou strukturu informací pro případy, kdy nevíme typ příchozích informací nebo se typy značně liší. V takové situaci je na místě použít dokumentové orientovanou populární MongoDB či ekvivalent [36][37]. Flexibilita nových typů databází je prakticky neomezená. Důkazem mohou být databáze označované jako NewSQL slibující spojení obou předchozích v nový model, který si bere škálovatelnost z NoSQL, ale zároveň ponechává konzistenci dat z SQL [38].

3.3.3 Servery a komunikace

Ne tak radikálním, ale přesto neméně důležitým vývojem prošly i webové servery a protokol http, bez kterého by web nemohl existovat. Na počátku milénia, kdy byl web stále ještě nový a neprozkoumaný, přišly internetové společnosti na nové využití webu. Mimo webové stránky se začaly vznikat webové API. *Application Programing Interface* je v kontextu webu používáno jako přístupový bod (skrze URL adresu) pro vývojáře umožňující sdílení informací a interakci s cizími systémy. Webové API, ačkoliv se v podstatě nejednalo o novou technologii, jen jiné využití webových serverů, otevřely cestu většímu propojení webových stránek mezi sebou. Nejčastěji to můžeme pozorovat u zpravodajských stránek, kdy je v článku například video z YouTube, příspěvek z Twitteru, či reklamy od třetí společnosti [54].

Důležité změny z pohledu manipulace s informacemi se odehráli v roce 2011, kdy byl standardizován nový komunikační protokol WebSocket operující s http. Tento protokol umožnil poprvé navázat oboustrannou komunikaci mezi webovým serverem a klientem bez použití jakýchkoliv kliček. Jako byl do té doby používaný *polling* – klient naváže spojení se serverem a čeká na odpověď, server tak může odpověď pozdržet, dokud nemá nějakou zprávu, pro zprávy od klienta se však muselo otevřít další spojení [55]. Očividným použitím toho nového protokolu byly chatovací či herní webové aplikace, ale možnosti použití této oboustranné komunikace nejsou nijak omezeny [57].

WebSocket řešil problém, který zbytečně zahlcoval přenos skrze http mezi serverem a klientem, ale webové stránky přenášely čím dál tím více údajů, jak se stávaly komplexnější. Dosavadní verze http/1.1 začala ukazovat své limitace a bylo potřeba toto omezení odstranit. Řešením se stala v roce 2015 představená nová verze http/2. Klíčovou změnou oproti předchozí verzi bylo odstranění limitu na počet souběžných aktivních http připojení klienta se serverem, díky paralelním požadavkům. Tato změna otevřela možnosti přenosu mnohem většímu objemu údajů na mnoho různých míst současně v kratším čase, limitací bytí jen rychlost připojení obou stran. Další úspory na složitosti přenášeních dat díky kompresi hlaviček požadavků, neboť se často nemění mezi požadavky [56].



Obr. 4 Paralelní požadavky pomocí http/2
Zdroj: https://factory.dev/blog/http2-difference-from-http1

Se zavedením http/2 se předchozí verze nepřestaly používat, aktuálně je http/2 používáno na 39.5 % všech webových stránek [58]. Proč je tento podíl tak nízký? Většina menších či středních webů nevyžaduje takové množství přenesených údajů či nové funkce, aby narazila na limitace http/1.1 a tak například nikdy nepřešly na novou verzi [59].

Plynulým navázáním na popularitu webových API se od největších společností zrodily cloudové servery. Cloudové servery nepředstavují novou technologii, ale jsou ve své podstatě webové API ve velkém měřítku. Do té doby neproveditelné analýzy, zpracování, či velikosti úložiště údajů se najednou stali dostupné pro větší klientelu. Na základě cloudu prošel, a nadále prochází, web transformací, kde prakticky kdokoliv může využít prostředků, dříve dostupných pouze v rámci těchto společností [54].

4 Současný web a budoucnost

Webové technologie a celkový technologický rozvoj zaznamenaly v krátkém časovém intervalu nesmírný rozvoj. Mnoho lidí a společností je zcela závislých na určitých webových stránkách/službách. Tuto závislost mohou chtít velké společnosti využít ve svůj prospěch skrze monopolizaci. Takovýmto krokem často utrpí uživatelé, a tak se tyto dvě strany nestále pokouší najít rovnováhu mezi vzájemnými potřebami.

4.1 Internet korporací

Jestliže na začátku měl web sloužit k akademickým účelům, a později veřejnosti, tak dnes existuje primárně komerční web. Za počátek komerčního webu (Web 2.0) jsou často symbolicky považovány sociální sítě, a to především příchod Facebooku v roce 2004 [42]. Důvodem je to, že takovýto typ stránek je založen čistě na uživatelské interakci, bez uživatelů by takovýto web nemohl existovat. Vznikem a rozšířením sociálních sítí se i principy strukturalizace webu ubraly směrem komponentů s variabilním obsahem, zpravidla založeným na údajích o uživateli. Tento styl webového obsahu se stal praktickým standardem při budování menších i rozsáhlejších webových stránek [63].

Na základě těchto nových typů stránek se reklamy začaly vyvíjet ve stejném směru. Nové informace o uživatelích otevřely dveře k větší integraci inzerce do používání webů. Roky 2010-12 odstartovaly éru takzvaných sponzorovaných příspěvků. Tyto příspěvky měly působit méně intruzivně, neboť méně vyčnívaly, a to nejen podobou ale také obsahem, neboť byly cílené na lidi s podobnými zájmy či potřebami jaké reklama poskytovala [14][18]. To samozřejmě neznamenalo, že by vymizely mnohdy dotěrné reklamy, ale spíše byly doplněny o další. Praktickým výsledkem je web běžící na zpeněžení ukořistění lidské pozornosti.

Nelze opomenout významný vliv chytrých telefonů na vývoj webu. Pro zpřístupnění telefonům musel web projít změnami nejen na povrchu – nástup responzivního designu přizpůsobující se více rozlišením. Chytré telefony ovlivnily také pozadí, a to v ohledu poskytování dat. Neustálá konektivita, množství senzorů a větší vnímaná intimita mobilních zařízení vedlo k mnohonásobnému zvýšení odesílaného

objemu osobních údajů [60][61][62]. Kupříkladu údaje o poloze v době před chytrými telefony prozradily pouze kde se nachází počítač, často neměnná informace. U mobilního zařízení může údaj o poloze v reálném čase prozradit kde se uživatel pohyboval, neboť telefony má dnes již mnoho lidí neustále u sebe.

V posledních letech lze pozorovat rozvoj mikrointerakcí a navádění uživatele pomocí animací. Vývojáři vykonávají značnou snahu, aby uživatele grafickým prostředím zaujali a vyvolali v něm pozitivní emoční reakci při používání webu [43]. Animace nejsou jedinou pomůckou k ovlivnění uživatele v prospěch provozovatele stránek. Stále častěji se využívá takzvaných psychologických *„dark patterns“*, tyto modely jsou speciálně navrženy k manipulaci uživatele do cílové interakce. Často to bývá nátlak ke koupi produktu, nebo udržení uživatele déle na webové stránce [44].

S tím, jak se zlepšovaly technologie pro přímou interakci s uživatelem, se zároveň rozvíjely technologie umožňující tuto interakci zachytit a analyzovat. Údajů přibývá a stávají se čím dál cennější, za pomoci vyspělých analytických nástrojů lze údaje v různých formátech propojit dohromady k vytvoření detailního profilu o uživateli [45]. Z nestrukturovaných dat je možné získat užitečné informace o kvalitě služeb, mentalitě uživatelů, způsobu užití a dalších [46]. Jak už bylo zmíněno v předešlých kapitolách, dominantním formou příjmu na webu je cílená reklama, čím více informací mohou společnosti zjistit o chování uživatelů, tím z jejich pohledu lépe.

4.2 Společnost a web

Internet ani web neexistují ve vakuu, jsou denně využívány lidmi k tvorbě, konzumaci obsahu, hledání informací, plánování, kolaboraci a pro mnoho dalších aktivit. Je tedy logické, že se prostřednictví webu budou dít i pochybné a neetické aktivity. Posouzení etiky a legitimity akcí je subjektivní a strana porušující pravidla většinou nebude posuzovat své akce za nesprávné. Proto se v této práci za pochybné berou především praktiky, které neúměrně obohacují jednu stranu na úkor druhé, bez značného přínosu pro opačnou stranu či ostatní. Pro příklad neetického designu je možné uvést uvedení pop-up reklam, které intruzivním způsobem útočily na pozornost uživatelů za cílem zobrazení více reklam [18].

Lidé využívající služeb webových stránek jsou v zásadě vydáni na milost provozovatelům, kteří mnohdy pro zvýšení výdělků kompromitují integritu produktu, nebo hůře, prodají soukromí uživatelů. Případů bylo v krátké historii webu již mnoho, například uniklé interní studie Facebooku (nyní Meta), které potvrzují, že používání Instagramu u náctiletých má na jejich psychiku negativní dopad. Před zveřejněním však Meta nijak tento problém neřešila, neboť v jejich zájmu je mít více uživatelů, nikoliv jejich blaho [66].

Uživatelé mají jako celek moc provozovatele dané stránky přinutit k odvrácení zaváděné změny jako se tomu stalo například u crowdfundingové webu Kickstarter, který ohlásil přechod na systém blockchain, ale po velké negativní odezvě krok odložil na neurčito [64]. Podobným případem byl test nových formátů Instagramu inspirovaných popularitou konkurenční sociálně sítě TikTok. Po negativním ohlasu uživatelů Instagram změny neuvedl a testování pozastavil [65]. Operovat s předpokladem, že všichni uživatelé mají stejné zájmy a zásady je naivní. Častokrát jsou společnosti operující webové služby tak velké s širokou uživatelskou základnou po celé Zemi, že společný nátlak většiny uživatelů je neuskutečnitelná představa. Největší šanci na prosazení svých zájmů mají uživatelé v podobě regulací a zákonů svých zemí a vlád. V kontextu České republiky hrají největší roli České a Evropské právo, do jisté míry jsou Česko a jiné země ovlivněny i jinými zákony.

4.2.1 Evropská směrnice ePrivacy

První krok v rozšíření práv uživatelů ohledně osobních údaje učinila Evropská unie směrnicí Evropského parlamentu a Rady 2002/58/ES (často zkracována jako ePrivacy či EDP) vydanou roku 2002, v Česku byla směrnice zavedena v platnost zákonem č. 127/2005 Sb. [70]. Tato směrnice mimo jiné reagovala na, v té době neregulované, sledovací technologie na webu (mimo dalších úprav týkající se soukromí a telekomunikační oblasti, to je však pro zaměření této práce irelevantní). Webové stránky mají nově povinnost srozumitelně informovat uživatele s ukládáním a přistupování k informacím na zařízení uživatele a o účelu jejich zpracování, tyto informace byly často používány k sledování uživatelů na webu. Uživatelům musí být také poskytnuta možnost odmítnout veškeré zpracování informací z úložiště zařízení kromě takového, které nezbytně nutných

pro provoz webu. Tento systém byl tedy založen na aktivním odmítnutí uživatelem – *opt-out* systém. Kvůli tomu, že společnosti mohly využívat úložiště i bez explicitního souhlasu od uživatele, směrnice příliš nezvýšila transparentnost ani informovanost uživatelů o informacích ukládaných na jejich zařízeních [69].

Důležitá úprava této směrnice přišla v roce 2009 pomocí směrnice 2009/136/ES označované jako „zákon o cookies". Touto změnou se systém změnil na *opt-in*, web musel pro ukládání a používání informací na zařízení obdržet souhlas od uživatele. V reakci na to byly na weby nasazeny bannery, které se zobrazí novým návštěvníkům a informují je o ukládání a používání informací na jejich zařízení spolu s účelem za kterým jsou zpracovány a vyžadují od uživatele souhlas. Výjimku stejně jako v původní úpravě tvořili informace nezbytně nutné k základní funkčnosti webu [71].

Poměrně malá změna byla zásadní, nyní totiž společnosti musely získat souhlas uživatele a toho docílily pomocí různých pochybných praktik. Na jedné straně implementace byly weby, kde byl na spodní straně stránky malý pruh informující uživatele o cookies, kde uživatel má pouze možnost potvrdit a po potvrzení není souhlas možné vrátit. Druhý extrém bylo skrytí obsahu webu za okno překrývající obrazovku, informující uživatele o zpracování informací z úložiště zařízení s jedinou možností potvrdit, nebo se uživatel k obsahu nedostane. Tyto a všechny typy mezi tím mají společné téma, vnucování souhlasu uživateli, který následkem toho nemá svobodnou volbu, zdali chce cookies ukládat či ne. Častým úkazem bylo také automatické ukládání souborů bez ohledu na souhlas. Uživatelsky nepřátelské implementace byly zapříčiněny velkou obecností a nespecifičností pravidel směrnice ohledně implementace získávání souhlasu, které společnosti, pro které je sledování uživatelů důležité, v zásadě obrátily proti nim [72].

Uživatelsky nepřátelské bannery a vyskakovací okna značně rozhořčily veřejnost a směrnice byla kritizována za neúspěch v ochraně osobních údajů na internetu. Všudypřítomné bannery sice rozšířily povědomí o cookies a sledování, nicméně nedávaly uživatelům příliš kontroly nad svými údaji. I za předpokladu, že společnosti nesouhlas respektovaly, uživatelé měli pouze nástroj, jak zamezit ukládání trackerů na svých zařízeních, neměli však jakoukoliv kontrolu nad sběrem

různých typů svých osobních údajů [72]. V návaznosti na tuto směrnici již Evropská Unie projednávala aktualizaci obecného nařízení o ochraně osobních údajů, které na některou kritiku ePrivacy přímo odpovídá.

4.2.2 General Data Protection Regulation

Doposud největším zásahem do volnosti společností operujícími s osobními údaji je nařízení GDPR Evropské Unie, neboť se vztahuje nejen na jednu zemi ale na uskupení mnoha, vyžaduje větší pozornost od nadnárodních společností. Tato směrnice byla schválena evropskou unií již roku 2016 avšak v platnost vešla až roku 2018, kdy každý členský stát musel nařízení přijmout do svých zákonů. Toto nařízení nahradilo směrnici Evropského parlamentu a Rady 95/46/ES, která platila do té doby a nereflektovala změny ve sběru osobních údajů na internetu [67].

Hlavními cíli evropského nařízení je zvýšit transparentnost sbíraných osobních údajů a vytvořit nástroje pro ochranu občanů EU, pomocí posílení konkrétních práv občanů. Shrnutí hlavních bodů nařízení týkajících se internetu a webu:

- Strana sbírající osobní údajů musí od uživatele, jehož údaje zpracovává dostat jeho souhlas ke konkrétnímu účelu zpracování (ve srozumitelném jazyce!), pro jiné účely údaje subjektu nemůže použít. Uživatel webových stránek, tak nyní může teoreticky zjistit jaké všechny údaje o vás kdo sbírá.
- S kým společnost sbírající osobní údaje uživatele tyto údaje sdílí, kdo je bude zpracovávat či k nim mít přístup.
- Právo uživatele na poskytnutí veškerých svých osobních údajů u společnosti, která je sbírá/zpracovává.
- Právo na přenositelnost údajů, uživatel může zažádat o údaje a ty poskytnout jiné společnosti.
- Právo na to být zapomenut, tudíž společnost musí smazat veškeré údaje o uživateli, pokud tomu něco nebrání (smluvní závazek, trestní stíhání atd.).
- Povinnost mít pověřence pro ochranu osobních údajů u velkých společností s rozsáhlým monitorováním osobních údajů uživatelů, a u veřejných subjektů.

- V případě porušení zabezpečení osobních údajů uživatelů musí společnosti informovat dozorové úřady příslušného členského státu do 72 hodin od zjištění porušení, tímto by se měla zlepšit informovanost uživatelů o únicích svých osobních údajů.

GDPR tak dává uživatelům nástroje, jak aktivně své údaje spravovat u společností nad nimi nemá jinak kontrolu, uživatelé by tak měli mít možnost kdykoliv požádat o své osobní údaje [67]. Všechny tyto opatření mají nepochybně pozitivní dopad na uživatele, ale jak se pravidla přenáší do praxe?

Odpovědí mnoha webů na nové požadavky skrze GDPR byla implementace vyskakovacího okna, které se zobrazí při první návštěvě webu, ne nepodobně jako u směrnice ePrivacy. Důvodem je, že mnoho webových stránek potřebuje ke své základní funkčnosti sbírat osobní údaje. Pomocí těchto oken má uživatel možnost zvolit souhlas se zpracování svých údajů rozdělených podle účelů zpracování. Patrně neočekávaným následkem takovéhoto dodržení GDPR, vznikl problém „únavy z vyskakovacích oken“. Neboť tyto okna byla na velkém množstvím webů, u uživatelů se rychle projevila únava a začali jen „odklikávat“ sběr veškerých údajů, aby se dostali k obsahu, jak tomu ostatně je u cookies bannerů [68].

Společnosti profitující z osobních údajů uživatelů využívají skuliny v GDPR ve svůj prospěch. Nařízení není v určitých definicích dostatečně specifické podrobnost informování o účelech sběru dat se může značně lišit. Mnohdy jsou společnosti záměrně co nejvíce obecné a vágní ve svých zásadách ochrany osobních údajů, že běžný uživatel jen obtížně zjišťuje, jaké jeho údaje jsou sbírány či jak a kým zpracovávány. Dalším způsobem ovlivnění uživatelů ve svůj prospěch je využívání *dark patterns* v rozhraních souhlasových oken, kdy kladná odezva (povolení všeho sběru) je často zvýrazněna a odmítnutí je málo prominentní, ne-li schované za nabídkou či odkazem. Některé weby ani neumožňují granulární selekci povoleného zpracování údajů a poskytují pouze souhlas se vším. S odkazem na fakt, že granularitu může uživatel docílit nastavením prohlížeče, které ale velmi často chybí na mobilních prohlížečích. Zákonná povinnost je tímto technicky splněna, ale reálný ochranný účinek pro uživatele je minimální [73][74].

Nehledě na to jde GDPR správným směrem v boji proti nadměrnému zneužívání osobních údajů. Důkazem mohou být nově vznikající regulace vlád ostatních zemí, často začleňující stejná práva a ochrany uživatelů jako GDPR. O globálním nárustu zájmu po větší ochraně osobních údajů vypovídá zavádění obdobných regulací i v USA, v podobě The California Consumer Privacy Act. Neboť Spojené státy mají mnohdy slabé regulace a větší volnost korporací [75].

4.2.3 The California Consumer Privacy Act

Po vzoru GDPR zavedla Kalifornie svůj The California Consumer Privacy Act of 2018 (CCPA) který začal platit od roku 2020. Tento zákon je podstatným krokem nejen pro občany Kalifornie ale pro celý svět, neboť Kalifornie je pátým největším trhem na světě a místem vzniku mnoha celosvětově používaných služeb a webů. V mnohém vychází, či je v souladu s GDPR, ale liší se v několika významných věcech, zde je výčet těch podstatných [75]:

- Vztahuje se na společnosti obchodující s občany Kalifornie, pouze pokud vydělávají nad 25 milionů USD ročně; obchodují s osobními údaji více než 50 tisíc uživatelů či zařízení pro komerční účely; anebo tvoří podíl z ročních příjmů z prodeje osobních údajů uživatelů více než 50 % celkového ročního příjmu. Z toho je jasně patrné zaměření zákona na větší společnosti, nebo specializované na osobní údaje.
- CCPA neaplikuje pravidla na osobní údaje typu zdravotnických informací, veřejně dostupných informací, pro spotřebitelské agentury a jiné typy údajů již upravované specifickými zákony.
- Uživatelé mají možnosti smazat, stáhnout a přesunout své údaje, ale k jejich sběru společnosti musí pouze informovat, nepotřebují, výjimkou osobní údaje osob mladších 16 let, kde souhlas musí být poskytnut předem. Navíc oproti GDPR mají uživatelé právo nebýt diskriminování či omezováni za použití svých práv a také právo zakázat prodej svých údajů, který musí být absolutně respektován. Tyto práva jsou aplikovatelná pouze na údaje za posledních 12 měsíců.

- Pokuty za porušení mohou být maximálně 2 500 USD za každé neúmyslné a 7 500 USD za každé úmyslné porušení. To je oproti GDPR menší s pokutou až ve výši 4 % z ročního celosvětového obratu z předchozího roku nebo 20 milionů EUR, podle toho, co je vyšší.

Zatímco GDPR se zaměřuje na práva z pohledu uživatelů, CCPA je přímou regulací společností, speciálně těch velkých. Mimo zmíněné rozdíly jsou si regulace podobné, což dokazuje, že je možné nastavit pravidla pro velké korporace, aniž by to vedlo ke globálnímu kolapsu internetové ekonomie [75].

4.3 Budoucnost webu

Do budoucna je velice pravděpodobné další přitvrzování ochranných zákonů. Nasvědčuje tomu rozpracované nařízení EU nahrazující dosavadní směrnici ePrivacy, které mělo vyjít spolu s nařízením GDPR, ale nepanovala mezi členskými státy shoda na tom, jak by mělo vypadat. Debatovanými opatřeními jsou například povinné do-not-track nastavení v prohlížečích a smartphonech, či zákaz podmínění souhlasu s cookies pro přístup k obsahu [74]. CPPA má být také rozšířeno pomocí The California Privacy Rights Act (CPRA), kdy Kaliforňané nově budou mít právo po vzoru GDPR k opravě nepravdivých osobních údajů, omezení použití citlivých údajů, nebo odmítnutí zpracování k behaviorálnímu marketingu. Rovněž se právo být informován o svých údajích zbaví limitu z posledních 12 měsíců [76].

Mimo oblast nekonečného boje za ochranu soukromí je již dlouhou dobu pro budoucnost webu prosazována idea sémantického webu, někdy též označeného jako Web 3.0. Idea sémantického webu je založená na vzájemné provázanosti informací a odvozování významu dle kontextu ve kterém jsou použity. Sémantický web můžeme pozorovat již dnes v podobě personalizovaného obsahu a profilování uživatelů hojně využívané především k cílení reklamy. Ten je totiž založený na vazbách mezi údaji, díky čemuž může uživateli nabídnout podobný obsah. Princip, který se v lokálních měřítkách jednotlivých firem aplikuje již dnes skrze grafové databáze [36][45][46].

Existují jiné názory, podle kterých bude web 3.0 revoluční změnou podobně jako web 2.0. Tento nový web bude založen na decentralizované síti [47].

Rozvoj v kryptografii, distribuovaných databázích a celkové významné zvýšení výkonu jednotlivých zařízení má umožnit přechod na novou síť [48]. Technologie pro vytvoření ekvivalentů některých služeb do decentralizovaného webu jsou dostupné již dnes [80]. Vize takovéhoto webu 3.0 je často spojována s kryptoměnami, jakožto model monetizace. V posledních letech cena a reputace kryptoměn utrpěla značnou ztrátu a spolu s ním se pomalu rozplývá i vize decentralizovaného webu. Ukazuje se však, že blockchain a přidružené technologie ve skutečnosti neposkytují žádné výhody oproti těm, které chtějí nahradit, v mnohých případech jsou i horší – především nejsou použitelné ve velkém měřítku [78][79]. Jen samotný bitcoin, který má údajně nahradit tradiční peníze a finance je energeticky neskutečně neefektivní. Jedna transakce bitcoinu spotřebuje elektřiny jako zhruba 1 milion VISA transakcí. Jasně se ukazují trhliny v decentralizovaném webu, když i tak jednoduchá věc jako transakce, je neúměrně náročná [77].

Jakožto více pravděpodobný směr vývoje webových technologií se jeví významné rozšíření a integrace virtuální, případně rozšířené reality. Web s virtuální technologií by nejvíce změnil interakci s webem a také formát, či účel obsahu [49]. Náznaky vydání se touto cestu můžeme pozorovat už nyní u velkých společností jakou je Facebook (nově Meta), která považuje virtuální svět za budoucnost [50]. Otázkou je, zdali by se vůbec stále jednalo o pokračovaní webu, či o něco zcela jiného.

Nicméně v případě odvození ze současných trendů se jeví jako budoucnost pouhá evoluce současného webu 2.0. Především v zapojení služeb poháněných či jinak spolupracujících s umělou inteligencí. Ukázkovým případem je bleskový vzrůst chatbotu ChatGPT postaveném na jazykovém modelu GPT-3.5, kdy chatbot dosáhl milionu uživatelů za pouhých pět dní. Pro srovnání Facebook dosáhl stejného limitu za 10 měsíců a Spotify za pět měsíců [81]. Nejsou to pouze chatboty, vznikají generativní AI programy produkující obrázky, kód, či videa podle zadání. Potenciál je obrovský, neboť umělá inteligence poskytuje nové nástroje, které dříve možné nebyly [82]. Spolu s rozšířením nových technologií přichází také spoustu problémů, které je třeba vyřešit. Mezi ty nejpalčivější patří problém se prezentace zavádějících výsledků, protože vychází z nekvalitních dat. Nebo nekonečná debata ohledně autorského práva a plagiátorství pomocí umělé inteligence [83][82].

5 Cíle a metodika analýzy

Praktickou částí této práce je, jak už název napovídá, analýza sběru osobních údajů. Skrze analýzu na skutečných webových stránkách je možné získat vhled pod povrch teoretických poznatků a objevit skutečný aktuální stav v oblasti sběru a ochrany osobních údajů. Cílem této analýzy je poskytnout odpovědi na otázky jako:

Přinášejí regulace vlád skutečné výsledky pro uživatele?

Jsou uživatelé poctivě informování o sběru, a sdílení svých osobních údajů třetím stranám?

Mají uživatelé skutečně kontrolu nad svými údaji, případně existují způsoby, jak se mohou chránit před nechtěným sběrem?

Je možné pomocí běžně dostupných nástrojů ověřit sbírané údaje?

5.1 Související práce

Analyzovat webové stránky lze v mnoha ohledech se zaměřením na různé aspekty a každý má svá pro a proti. Studie Englehardta a Narayanana z roku 2016 analyzuje jeden milion nejpopulárnějších webových stránek pomocí nástroje OpenWPM instruující prohlížeč Selenium založený také na jádru Chromium. Studie poskytuje dobrý průřez skutečným webovým prostředím, díky analýze nejen požadavků http ale také lokálně uložených souborů (cookies, localstorage). Demonstruje také efekt, jaký mají rozšíření blokující sledování či reklamu. Ačkoliv testují weby pomocí skutečného prohlížeče, a ne pomocí okleštěného headless prohlížeče, z principu automatického testu není možné simulovat realistické používání webové stránky. Taková forma automatické analýzy by dnes již nebyla možná, neboť je často vyžadováno udělení souhlasu se zpracováním údajů [85].

Výzkum Papadogiannakise et al. analyzuje, zdali stránky dodržují volbu souhlasů uživatele v rámci GDPR. Forma analýzy probíhá pomocí automatizované návštěvy pomocí prohlížeče založeném na jádru Chromium s rozšířením Consent-O-Matic 850 tisíc nejnavštěvovanějších webových stránek dle online seznamu. Za pomocí rozšíření se sbírají odesílaná data dle různých stavů potvrzení souhlasu uživatelem. Rozšíření detekovalo pouze okolo 27 tisíc stránek, kde mohlo manipulovat se

souhlasy, tedy zhruba 3,3 %. Studie má nepochybně široký záběr díky automatizaci a počtu vzorků, nicméně je kvůli možné detekci malého počtu CMP (Platforma pro správu souhlasu) pravděpodobné, že jsou údaje zkreslené. Mnoho webů si tvoří vlastní CMP, tudíž panují velké rozdíly mezi detekovatelnými prvky. Zkreslení by tedy postihlo především střední, nebo propracovanější menší weby s vlastními řešeními správy souhlasu [84].

Analýzy pomocí automatizovaných nástrojů jsou vystaveny riziku detekce pomocí webových stránek, která vede k snížení věrnosti výsledků. Spojení manuální analýzy s použitím obtížně detekovatelných nástrojů přináší zvýšenou kvalitu výsledků oproti čistě automatické analýzy. Výměnou za zvýšenou flexibilitu se bohužel snižuje kvantita sbíraných dat, neboť manuální vstup nelze tak jednoduše škálovat [86]. Analýza v této práci má primárně za cíl zjistit dopad regulací na uživatele, proto je hlavním kritériem při provádění analýzy co nejvíce se přiblížit tomu, jak běžný uživatel interaguje s webovými stránkami. Zvolená metodika analýzy se zaměřuje na menší počet kvalitních dat, získaných manuálním vyhodnocením za podpory specializovaných detekčních nástrojů.

5.2 Metodika analýzy

Pro účely analýzy je potřeba definovat, jak se v této práci používá pár klíčových pojmů.

Tím nejpodstatnějším je pojem **osobní údaje**. Podle Evropské unie jsou osobní údaje takové údaje, které poskytují informace o identifikované nebo identifikovatelné osobě. Na základě této definice jsou proto v této práci považovány za osobní údaje veškeré údaje, které poskytují informaci o uživateli. Důvodem pro použití této definice je takzvaná metoda *fingerprinting*, kdy spojením samostatně neidentifikujících údajů je možné po dosažení určité specificity implicitně přiřadit údaje k uživateli [87]. Neboť je mimo možnosti této práce zjistit, zdali každý údaj může vést k identifikaci osoby. Pro zpřehlednění získaných výsledků budou veškeré údaje, jakkoliv zpřesňující specifičnost uživatele/zařízení automaticky považovány za potencionálně identifikační, a tudíž zařazeny pod osobní údaje.

První strana, nejčastěji v použití *cookies první strany* jsou takové zdroje, které pochází přímo z domény, či subdomény provozovatele webové stránky. To znamená, pokud se uživatel nachází na stránce domena.cz a údaje jsou sdíleny s doménou api.domena.cz, která je prokazatelně vlastněna stejným subjektem, jedná se o první stranu v obou případech.

V kombinaci s první stranou je v práci využíván pojem **druhá strana**. Druhou stranu lze brát jako rozšíření první strany a spadají pod ní strany splňující alespoň jedno ze dvou kritérií:

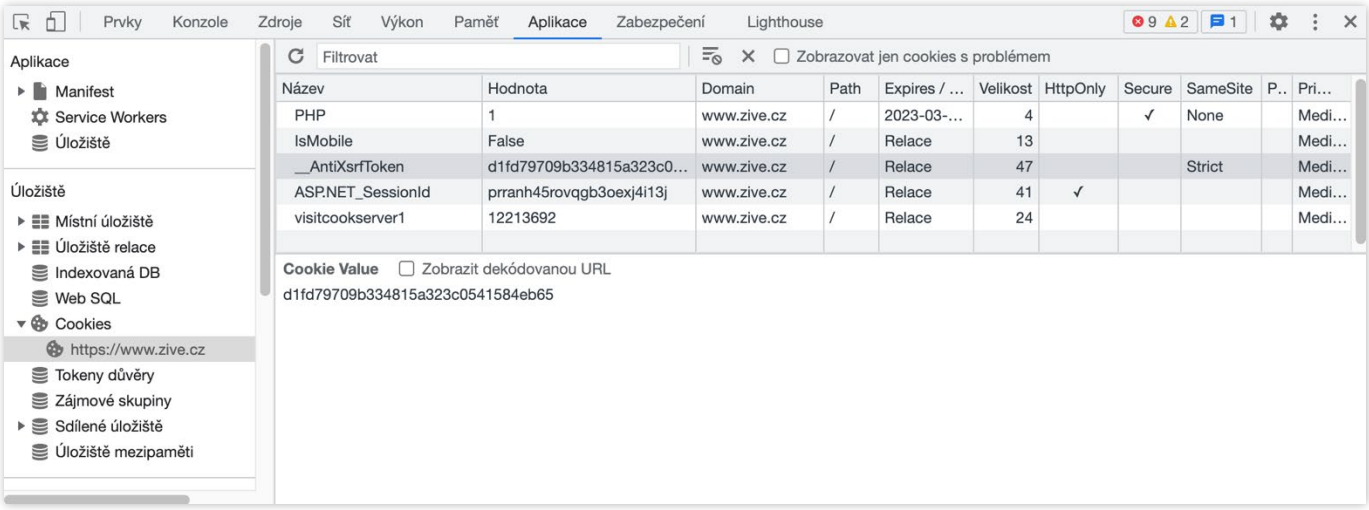
1. Pakliže opět přistupujeme na web domena.cz, který je vlastněn společností Doména s.r.o. a údaje jsou sdíleny s doménou top.domena.cz či top-domeny.cz, potom se jedná o druhou stranu, neboť tyto domény jsou vlastněny společností TopDomény a.s. která také vlastní společnost Doména s.r.o. V opačném případě, kdy z webu vlastníka jsou sdíleny údaje na dceřiné společnosti se opět jedná o druhé strany. Tato definice je nutná pro odlišení „nezávislých" třetích stran od těch, kdy údaje jsou technicky pod kontrolou stejné společnosti. Neberou se v potaz nuance, kdy dceřiné společnosti jsou prakticky nezávislé na mateřské, protože takovéto tvrzení nejsou prakticky ověřitelné.
2. Za druhou stranu se mohou brát i domény společností technicky provozující web. Uvedeno na příkladu: Existuje společnost Hynek s webem na doméně hynek.cz, web je ovšem vytvořený pomocí platformy na tvorbu stránek wp.com. Stránka hynek.cz tak získává některé zdroje z cdn.wp.com, které jsou neoddělitelné od obsahu hynek.cz. Pokud tedy web běží na nějaké takovéto platformě, není pro něj technicky možné vyvarovat se sdílení údajů s doménou platformy, a tak je platforma brána za druhou stranu.

Třetí strana je definována jednoduše jako nezávislá třetí strana, se kterou kromě dohodnuté (nebo jiné) interakce a sdílení údajů nemá na první stranu vazby. Nejčastějším příkladem v praxi jsou Content Delivery Network poskytovatelé obsahu nebo analytické zpracování nejčastěji Google Analytics či jiné.

Třetí strana poskytující obsah jsou takové strany, které pro stránku ze své domény doručují obsah, nebo nějakou část funkcionality, například vložené video z Youtube, hosting obrázků nebo knihovna pro Javascript, …

Třetí strana sledující uživatele (trackery) jsou strany, které nepřispívají žádným způsobem přímo k užitku návštěvníka stránky, ale pouze o něm sbírají údaje pro své účely.

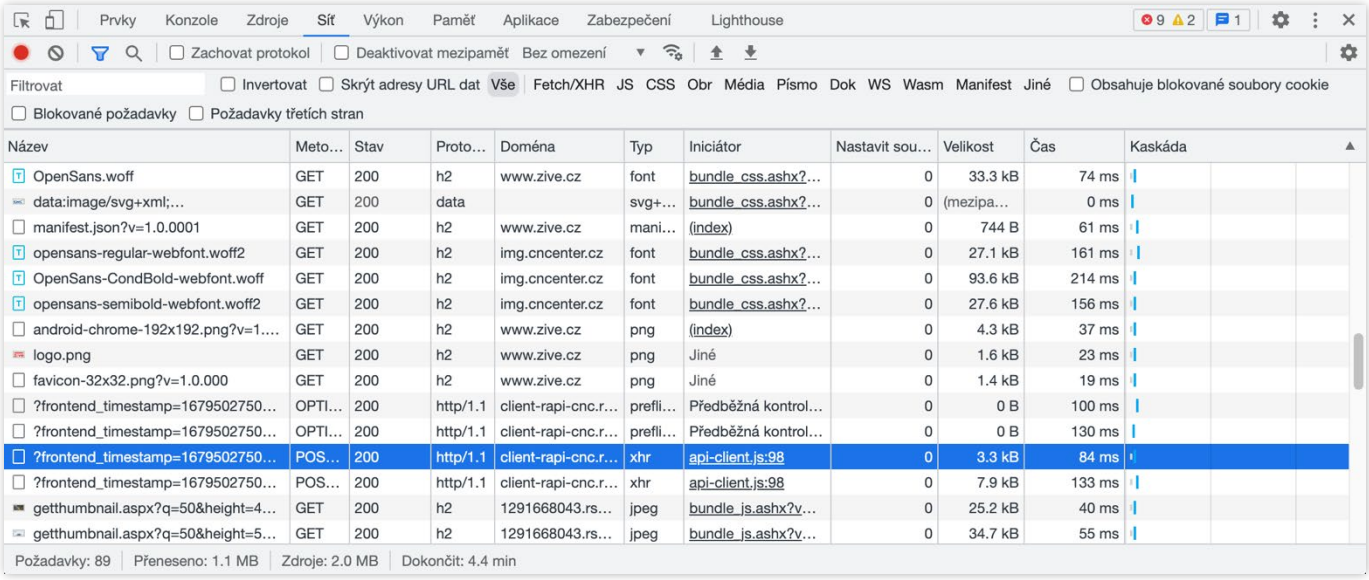
5.2.1 Použité nástroje



Obr. 5 – Chrome DevTools karta Aplikace, detail cookies (zive.cz)

Pro analýzu webových stránek byla klíčová sada nástrojů pro vývoj na webu zabudovaná v moderních prohlížečích – Chrome DevTools. Pro inspekci uložených souborů cookies a položek v lokálním úložišti sloužila karta Aplikace. Tato karta poskytovala informace o všech platných cookies, konkrétně pro kterou doménu platí, jejich životnost, hodnota a stav zabezpečení, všechny tyto údaje jsou užitečné pro správné vyhodnocení účelu konkrétních cookies. Na obrázku je vidět příklad z webu zive.cz před přijetím jakéhokoliv souhlasu. Parametr *HttpOnly* určuje, zdali je cookie soubor přístupný Javascriptu, tzn. pokud je pozitivní tak je cookies přístupné pouze skrze síťové http(s) požadavky. Vyplněný parametr *Secure* nám říká, že cookie se k požadavku přidá pouze pokud je přenos skrze https. Nakonec parametr *SameSite* má možné stavy: *None* – cookies budou odeslány i v požadavcích z jiných stránek; *Lax* a *Strict* odesílají cookies jen ze stejné domény jako byly cookies vytvořeny [88]. Nejen obsah cookies ale také tyto parametry prozrazují důležité informace ke správnému zařazení do kategorie účelu cookies.

Dalším klíčovým nástrojem v devtools je karta *Síť*. Pomocí této karty lze zaznamenat a uložit veškerou komunikaci mezi klientem (prohlížeč) a servery webových stránek. Protože se jedná o vývojový nástroj, zobrazuje také obsah zašifrované komunikace skrze https, který je pro vnější okolí nečitelný. Tato karta poskytuje mnoho podstatných informací pro analýzu. Zde je souhrn těch podstatných [89].



Obr. 6 – Chrome DevTools karta Síť (zive.cz)

Sloupec **metoda** poskytuje informaci o metodě požadavku. Typicky se používají metody GET a POST, liší se v tom, že GET je považována za „bezpečnou“ metodu, která nezpůsobují změnu na serveru. GET se používá pro získávání zdrojů nebo pro ohlášení, zatímco metoda POST zpravidla ukládá nebo jinak manipuluje se serverem. Hlavním rozdílem je také, že požadavek metody GET neobsahuje žádné tělo a vše je uvedeno v adrese URL formou parametrů. Další specifické metody jako DELETE, PUT, PATCH se používají v malém množství případů.

Sloupec **doména** ukazuje doménu, na kterou požadavek odchází, pokud je prázdný, požadavek se nepodařil nebo se jedná o datovou URL, která je lokální.

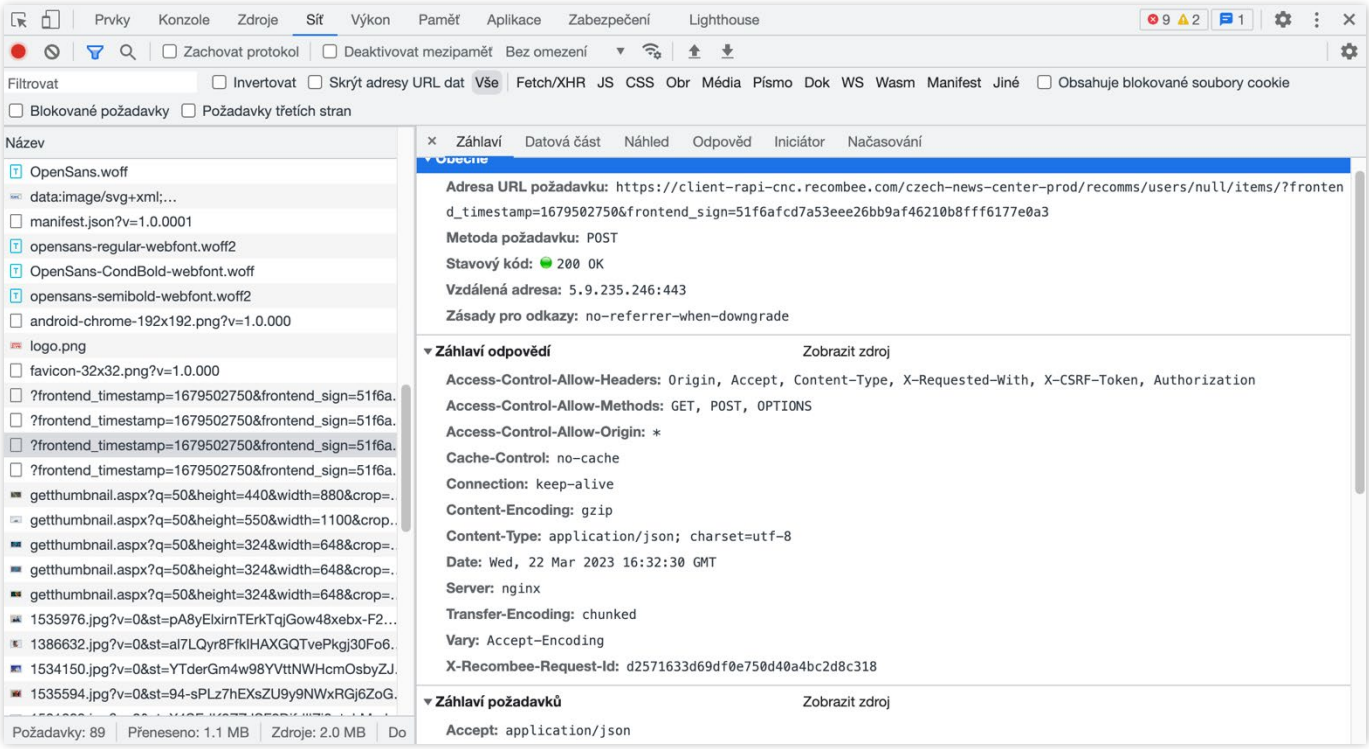
Sloupec **Stav** ukazuje stav požadavku, běžně je to 200, což je http kód pro úspěšný požadavek, kódy 3xx jsou přesměrování, 4xx je odpověď pro špatný požadavek ze strany klienta (například omezený přístup), 5xx je chyba serveru. Některé kódy plní speciální funkce například 101 značí změnu protokolu, ale nejčastěji iniciuje spojení WebSocket, 204 je kladná odezva bez těla odpovědi, 451 je obsah nedostupný z právních důvodů atd [90].

Sloupec **typ** značí typ těla odpovědi na požadavek, *script* značí že server posílá skript, *xhr/fetch* jsou požadavky vyvolané skrze JavaScript, *jpg/png/gif* je obrázek, *document* je odpověď vracející stránku.

Iniciátor je podstatný sloupec, protože ukazuje, jaké požadavky či volání skriptů požadavek vyvolali. Nesmírně důležitý nástroj pro sledování historie požadavku a reverzního inženýrství odesílaných informací v požadavku.

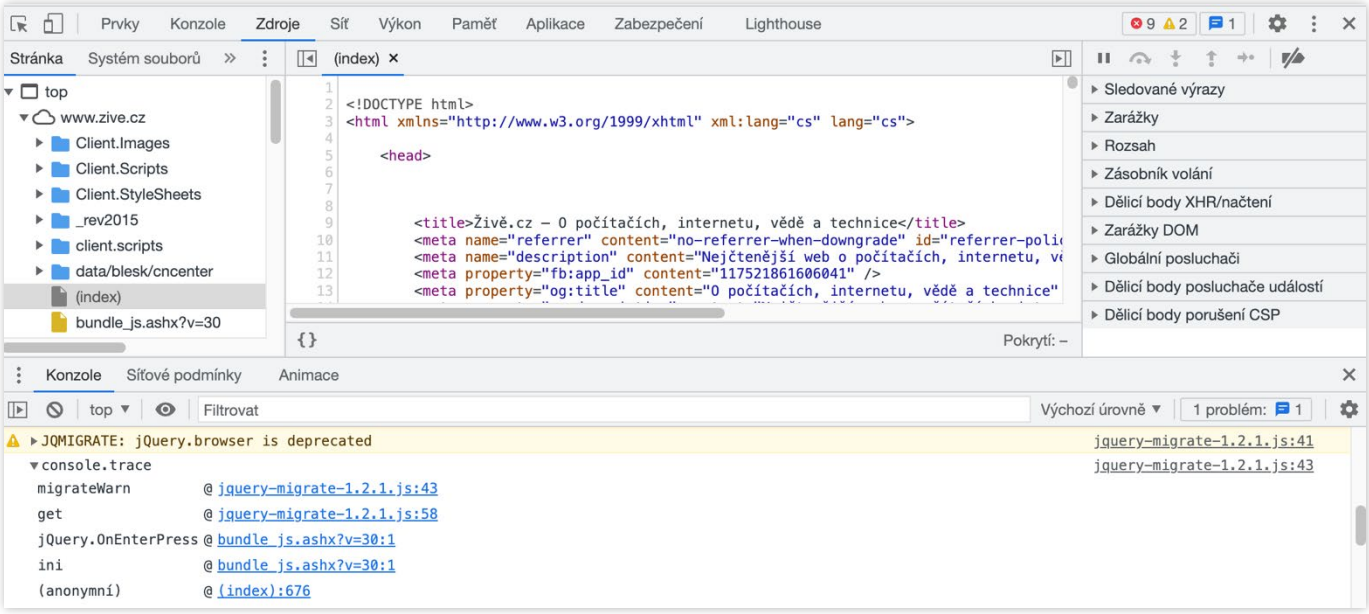
Sloupce **cookies** ukazují počet odeslaných souborů cookies s požadavkem a **nastavit soubory cookies** ukazuje počet nastavených cookies s odpovědí serveru.

Kaskáda vizuálně zobrazuje dobu načítání a čekání na odezvu požadavků. Pro tuto práci není příliš důležitý, kromě vizualizace chronologického řazení požadavků.



Obr. 7 – DevTools detail POST požadavku (zive.cz)

Po rozkliknutí jednotlivé položky je možné zjistit další podrobnosti o jednotlivých požadavcích. V kartě *Záhlaví* zjistíme veškeré hlavičky http požadavků a odpovědi serveru. Karta *Datová část* obsahuje odesílaná data jak v URL adrese (GET) tak v těle požadavku (POST). Karta *Odpověď* obsahuje odpověď serveru v surovém formátu a karta *Náhled* obsahuje upravenou odpověď, tzn. náhled obrázků, zformátovaný kód, náhled dokumentu, a další. *Iniciátor* nabízí detailnější rozpis iniciátorů z přehledu a *Načasování* je podrobnější kaskáda. Karta *Síť* mimo jiné poskytuje podrobné filtry dle typů, domén, domén odeslaných cookies a mnoho dalších usnadňující orientaci v případě velkých počtů požadavků. Celé záznamy síťové komunikace lze také ukládat do formátu HAR, který je možné zpět nahrát do devtools a pracovat s nimi, nicméně není možné ukládat historii iniciátorů [53].

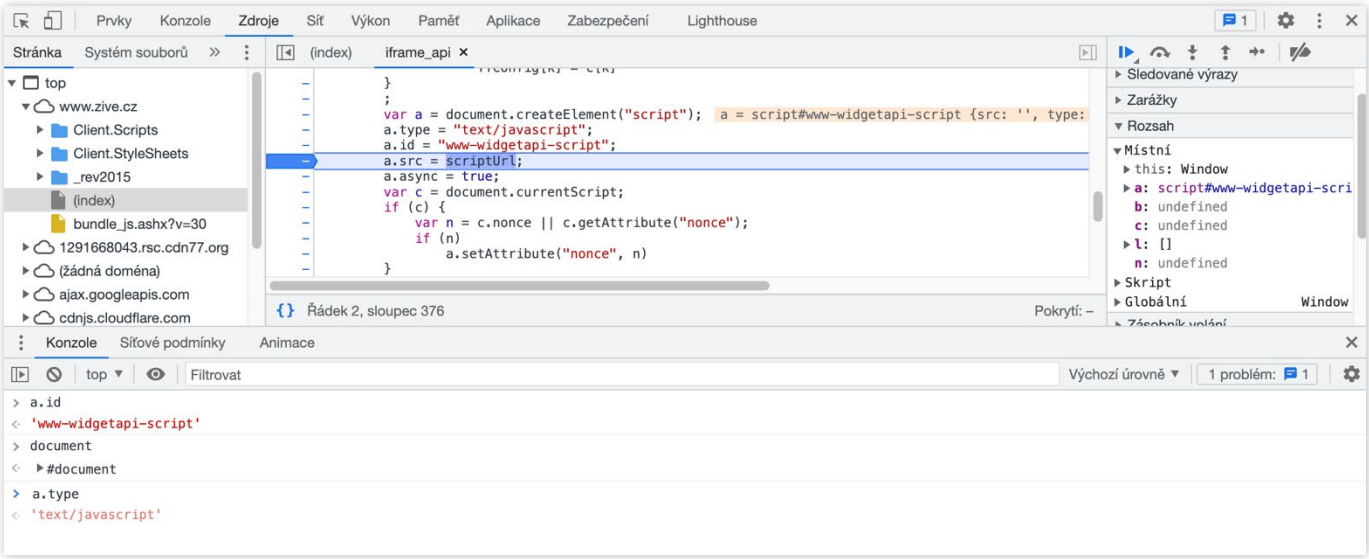


Obr. 8 – rozdělení DevTools, kombinace karty Zdroje a Konzole (zive.cz)

Poslední karty, které je třeba zmínit jsou karty *Zdroje* a *Konzole*. Jak už název napovídá karta *Zdroje* umožňuje prozkoumání veškerých načtených zdrojů na webové stránce. Zároveň umožňuje manipulaci se zdroji ukládáním *Zarážek*. Díky zarážkám je možné pozastavit načítání, či provádění kódu v daném bodě (standardně dle jednotlivých řádků kódu). To je velmi podstatné při provádění reverzního sledování tvorby odesílaných dat v těle požadavku [89].

Obě karty jsou na jednom snímku obrazovky, neboť devtools je možné vodorovně rozdělit a dolní část může nezávisle na horní zobrazovat údaje. Nejpoužívanějším

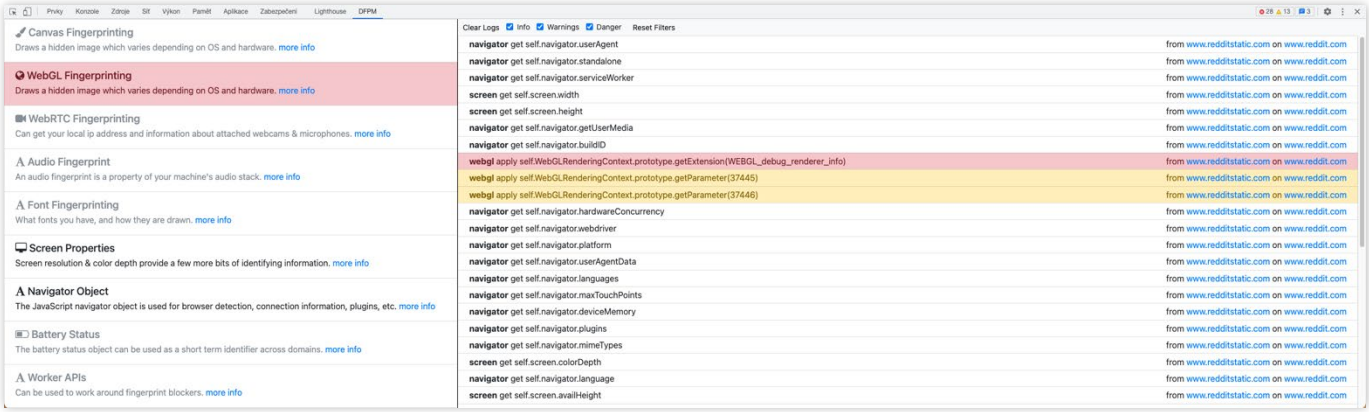
využitím dolní části je právě pro kartu *Konzole*. Konzole je rozhraní, do kterého může Javascriptový kód vypisovat jakékoliv údaje, případně slouží pro automatické vypisování varování a hlášení chyb prohlížeče. Pokud je pomocí zarážky nebo jiným nástrojem pozastavený kód tak lze pomocí konzole manipulovat s proměnnými v paměti. Jak je možné pozorovat na snímku, zarážky lze aplikovat i na řádky bez čísla. Neočíslovaný řádek je způsobený minifikací zdrojového kódu, ten lze zformátovat pomocí tlačítka {}. V horní pravé liště se nachází nástroje pro vstupování či vyskakování/přeskakování z funkcí v aktuálně běžícím prostředí [89].



Obr. 9 – pozastavený kód a výpis proměnných v konzoli (zive.cz)

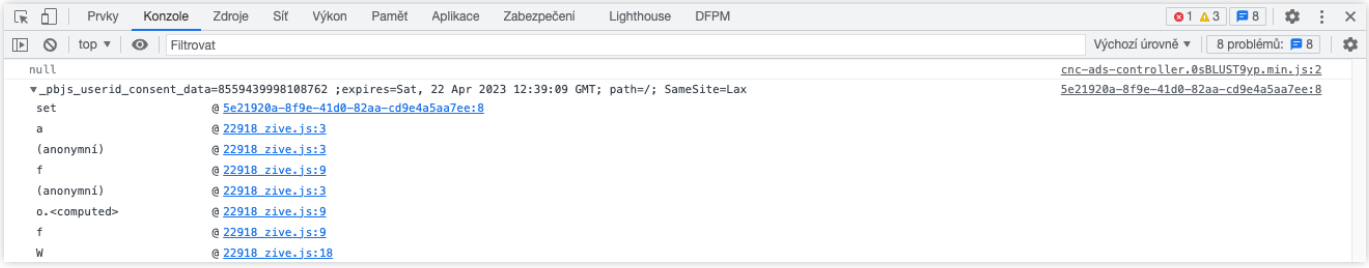
Mimo zabudované nástroje přímo v devtools byly pro analýzu využity také rozšíření. První z nich s názvem *Don't FingerPrint Me* poskytuje rozšíření devtools o novou kartu. Toto rozšíření poskytuje nástroj, jak efektivně zjistit, zdali webové stránky užívají identifikaci pomocí některých z technik generování unikátního otisku zařízení. Rozšíření funguje na jednoduchém principu vložení vlastního skriptu před načtením všech ostatních částí stránky. Tento skript poté hlídá a hlásí veškeré přístupy na určité proměnné prohlížeče jako je *navigator, screen, webrtc* atd. ze kterých lze získat detailní informace o prohlížeči či uživateli. Mimo hlášení těchto přístupů je také kontroluje oproti vzoru známých fingerprintingových technik. Do protokolu přijde varování, v případě podezřelého přístupu, který může i nemusí souviset s fingerprintingem. Pokud je shoda se vzorem, tak do protokolu ohlásí nebezpečí. Veškeré varování a nebezpečí byly dále manuálně zkontrolovány, zdali se nejedná o falešná pozitiva [91]. U nejčastější metody canvas fingerprinting,

kdy je využito vykreslení písem a komplikovaných obrazců na plátně canvas API a následně se na základě vykresleného vytvoří unikátní identifikátor, neboť každé zařízení bude prvky vykreslovat nepatrně jiným způsobem a jinak rychle [92]. Častým falešných pozitiv bylo vykreslení emodži pomocí doplňku platformy Wordpress.



Obr. 10 – Karta Don't FingerPrint Me detekující fingerprinting (reddit.com)

Druhým použitým rozšířením bylo jednoduché rozšíření *CookieLogger* vytvořené pro tuto analýzu. Toto rozšíření je reakcí na limitaci devtools v monitorování vzniku cookies. V standardním nástroji devtools je možné najít pouze odezvu serveru, která soubor cookie vytvořila. Některé soubory (především bez parametru *httpOnly*) jsou vytvářeny v průběhu interakce se stránkou pomocí Javascriptu. Obdobným principem jako u rozšíření DFPM je zde při načtení dokumentu vložen skript, který poslouchá na veškeré nastavení nových cookies pomocí `document.cookie` a odešle zprávu do konzole skrze `console.trace()`. Výsledkem je nejen zobrazení hodnot nastavovaných cookies ale také zobrazení historie volání, jež cookie nastavila. Rozšíření tedy zastává funkci iniciátoru pro nastavení cookies skrze Javascript.



Obr. 11 – ohlášení nastavovaného souboru cookies skrze CookieLogger (zive.cz)

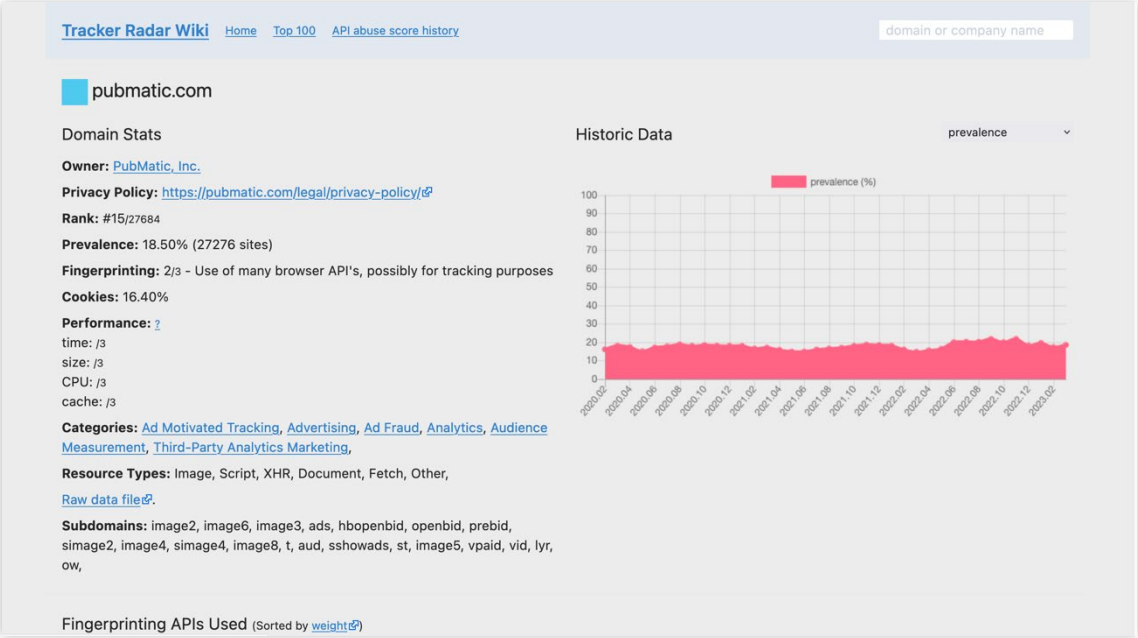
Mimo výše zmíněné konkrétní nástroje bylo využito i dalších nástrojů v rámci devtools jako je umělé zpomalení sítě či změna user-agent stringu prohlížeče pro

otestování variability odesílaných údajů. V některých případech bylo využito služeb virtuální privátní sítě od ProtonVPN, skrze které lze ověřit, zdali jsou odesílány údaje o poloze.

Pro zajištění správného přiřazení společnosti k jednotlivým doménám (mnoho adTech společností operuje pod více doménami) bylo kombinace využito veřejně dostupných nástrojů:

Registry domén Whois (CZ.NIC, ICANN) – domény musí společnosti a jedinci zaregistrovat u registrátorů, kde poskytují kontaktní informace, podle kterých se dá přiřadit daná doména ke společnosti. Mnohdy jsou však u domén trackerů tyto údaje vymazány „z důvodu ochrany soukromí“ nebo je doména registrována skrze zastupující společnost. Proto bylo využito dalších nástrojů.

Nástroje Better [94], Netify [95], Tracker Radar [96] a WhoTracksMe [97] obsahující rozsáhlé databáze známých domén trackerů. Ve většině případů alespoň jeden z nástrojů dokázal přiřadit k doméně společnost. Pakliže se výrazně lišila oznamovaná společnost domény mezi jednotlivými službami, bylo nutné pomocí vyhledání názvu společností zjistit komu skutečně patří. K tomu se bylo využito mnoho registrů firem a veřejných sledovačů aktivit a akvizic společností.



Obr. 12 – nástroj DuckDuckGo Radar Tracker pro doménu pubmatic.com

5.3 Průběh analýzy

Analýza probíhala v době od 3. do 25. února 2023 na vzorku 31 webových stránek z Česka i zahraničí. Webové stránky byly navštíveny pomocí prohlížeče Chrome verze 110 na zařízení Apple Macbook M1 s operačním systémem MacOS 13.2.1. Mezi každým testovaným pokusem bylo provedeno navrácení prohlížeče do čistého stavu (vyčištění mezipaměti, cache, cookies, …). Díky flexibilitě manuální inspekce a použití nejpoužívanějšího prohlížeče Chrome [98] (nikoliv Chromium) výsledky reflektují zkušenosti simulovaného skutečného používání průměrným uživatelem.

Analýza probíhala v třech etapách, mezi každým krokem nevyžadujícím kontinuitu předešlé relace, nebo nástroje v podobě rozšíření, bylo provedeno uvedení prohlížeče do čistého stavu. Každá stránka byla otestována v následovně:

5.3.1 První etapa – třetí strany

Pomocí čistého prohlížeče bez rozšíření byla navštívena webová stránka. Na této stránce se skrze devtools monitorovaly veškeré požadavky a odezvy serverů. Na každé stránce bylo provedeno několik akcí, simulující chování běžného uživatele (rozkliknutí příspěvků, navigace na různé stránky, interakce s obsahem). Každá stránka byla navštívena čtyřikrát, při každé návštěvě byly provedeny stejné akce pro zachování konzistence výsledků. První tři návštěvy se lišili stavem udělení souhlasu, konkrétně *před udělením souhlasu*, *minimální souhlas* – odmítnutí všeho, typicky kromě nezbytných údajů a cookies pro chod webu, *plný souhlas* – přijmutí všeho, toto chování je reprezentativní chování většiny uživatelů ať už z vlastní vůle nebo nedobrovolně [73][93]. Čtvrtá návštěva je opět s plným souhlasem, tentokrát však s nainstalovaným populárním rozšířením pro blokování reklamy a sledování *uBlock Origin*.

Při každé návštěvě byly zaznamenány veškeré úspěšné požadavky na domény třetích stran a pomocí výše zmíněných nástrojů došlo k zařazení domény ke společnosti (každá společnost = jedna třetí strana). Dále se určila pro každou třetí stranu jedna ze tří kategorií: *tracker*, *tracker/obsah*, *obsah*. Do kategorie *tracker* byly zařazeny mimo jiné takové třetí strany, které přijímaly údaje typu zahešovaný identifikátor, sdílely cookies, metoda 1x1 sledovacího pixelu a zároveň nijak

nepřispívaly k doručení obsahu, který je k užitku návštěvníka webu. Kategorie *obsah* je pro třetí strany, ze kterých je získáván obsah k užitku návštěvníka či funkčnosti webu. A poslední kategorie *tracker/obsah* je pro třetí strany spadající do obou předešlých kategorií zároveň.

5.3.2 Druhá etapa – kategorizace odesílaných údajů

Následoval test zjišťující kategorie sbíraných údajů od prvních a druhých stran. Prohlížeč byl v čistém stavu, pouze s nainstalovaným rozšířením CookieLogger. Opět byla simulována běžná interakce uživatele s webem. Skrze devtools byly prohlíženy všechny požadavky prvních a druhých stran, a hlavně jejich datové části. Mimo to, také inspekce veškerých cookies s cílovou doménou první či druhé strany. Je důležité poznamenat, že do sbíraných osobních údajů se braly pouze ty skryté, tedy takový sběr, který vzniká na pozadí. Pro představu je to sledování uživatelových kliknutí na odkazy oproti odeslání údajů z formuláře vyplněného uživatelem.

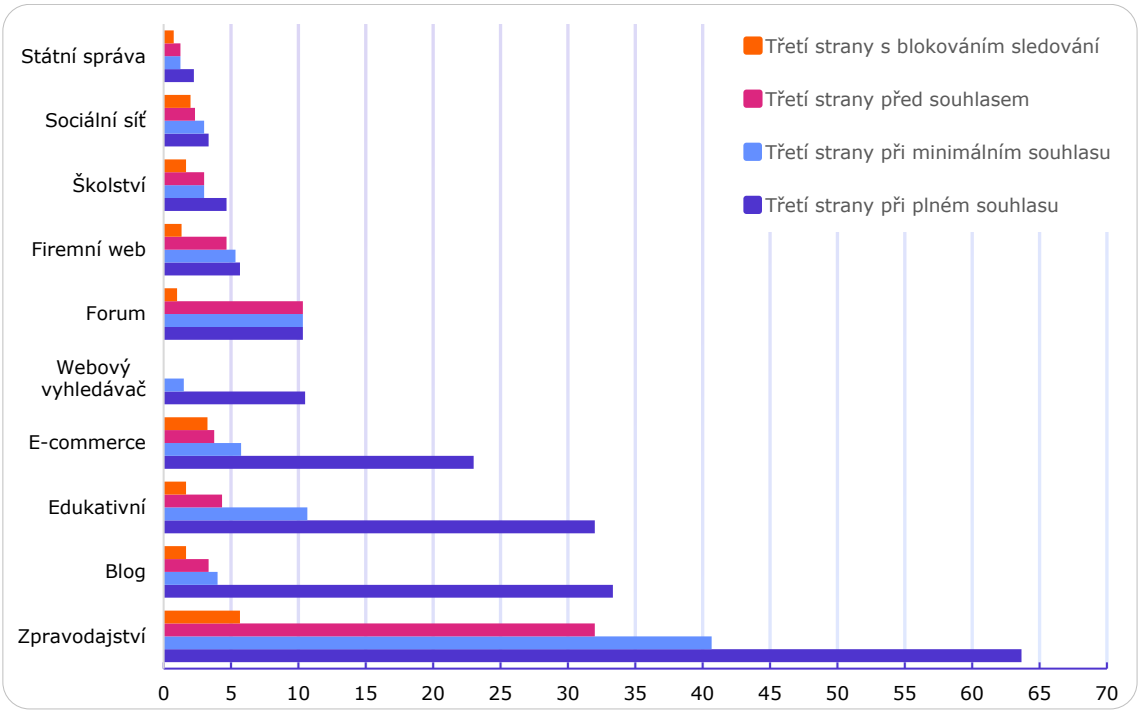
Pro zjištění vzniku a hodnoty se skrze historii volání cookie od CookieLoggeru prostupovalo řetězem volání až ke vzniku hodnoty cookie, pokud byla vytvořena na straně klienta. Obdobným postupem bylo postupováno k rozklíčování obsahu datových částí odesílaných požadavků za využití iniciátorů. Jednotlivé záznamy poté tvoří jednotky sbíraných osobních údajů, které již logicky není možné dělit na menší. Například dělení rozlišení na vertikální a horizontální produkující dvě jednotky by pouze zkreslovalo data.

5.3.3 Třetí etapa – doplnění o test na fingerprinting

Konečnou zvláštní návštěvou webu se měla zjistit přítomnost fingerprinting technologií. Prohlížeč byl tentokrát vybaven pouze rozšířením DFPM. Následně bylo projito několik podstránek webu a interagováno s obsahem. Každé detekování prošlo následnou manuální kontrolou zdroje. Případy, kdy bylo detekován fingerprinting skrze obsah třetí strany (typicky vložený iframe) nebyl případ zaznamenán jako fingerprinting pro danou webovou stránku.

6 Shrnutí výsledků

Výsledky analýzy 31 webových stránek ukázaly na několik společných vzorů mezi webovými stránkami ze stejných kategorií, ale také několik nepříjemných zjištění z pohledu uživatelů.



Graf 1 – Průměrný počet třetích stran v kategoriích webů

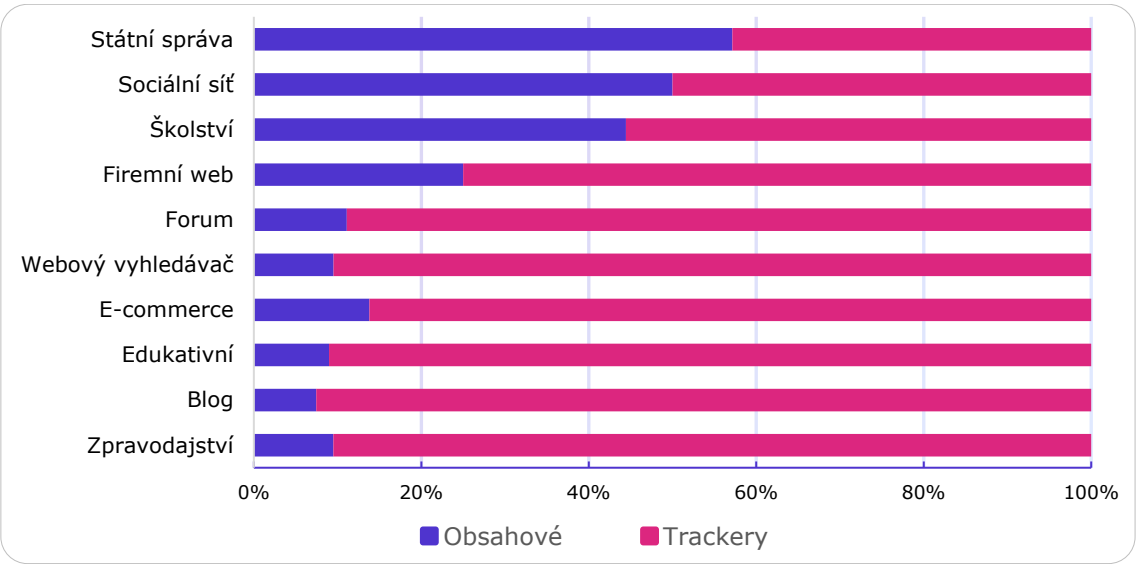
Začneme-li u pozorování sdílení osobních údajů s třetími stranami. Nejvíce zapojují třetí strany webové stránky z kategorií *zpravodajství*, *blog*, *e-commerce* a *edukativní*. Důvodem takto vysokého počtu jsou vložené reklamy a jejich systém aukce v reálném čase, který kontaktuje mnoho reklamních agentur, nabízejících reklamy od inzerentů podle údajů o uživateli (skrze sdílené unikátní ID uživatele) [99]. U mnohých stránek jsou třetí strany (jak obsahové, tak sledovací) kontaktovány již před poskytnutím souhlasu, po odmítnutí všeho se počet třetích stran navýší minimálně. To neznamená, že trackery byly aktivovány již před udělením souhlasu, většinou se jednalo o pouhé stáhnutí sledovacích skriptů, které do té doby byly nečinné. Rozdíl mezi minimální souhlasem a plným byl buďto velmi významný nebo nepatrný. Nejmenšího počtu třetích stran bylo pozorováno se zapojením rozšíření *uBlock Origin* pro blokování trackerů a reklam. Weby v kategorii *forum* kontaktovaly prakticky jen minimum třetích stran, výjimkou byl web DPReview, který

kontaktoval 27 třetích stran (viz příloha č. 1), web totiž slouží i jako zpravodajský a komunitní s podporou pomocí reklam. Deset webových stránek, neumožňují nijak nastavit úroveň souhlasu, z toho pouze Hacker News neodesílá žádné údaje třetím stranám. Ze sesbíraných dat lze pozorovat, že velké společnosti typu Google, Facebook, Twitter, Amazon přímo na svých stránkách nevyužívají mnoho třetích stran.

Webová stránka	**Země původu**	**Počet třetích stran**
Linus Tech Tips (linustechtips.com)	Kanada	4
Hacker News (news.ycombinator.com)	USA	0
DPReview (dpreview.com)	USA	27
TechArena (techarena.cz)	Česká republika	36
Vilnius University of Applied Sciences (viko.lt)	Litva	7
Ministerstvo zahraničí ČR (mzv.cz)	Česká republika	4
Blueghost (blueghost.cz)	Česká republika	9
Evropské spotř. centrum (evropskyspotrebitel.cz)	Česká republika	3
IMAGINARY (imaginary.org)	Německo	5
Bartosz Ciechanowski (ciechanow.ski)	USA	2

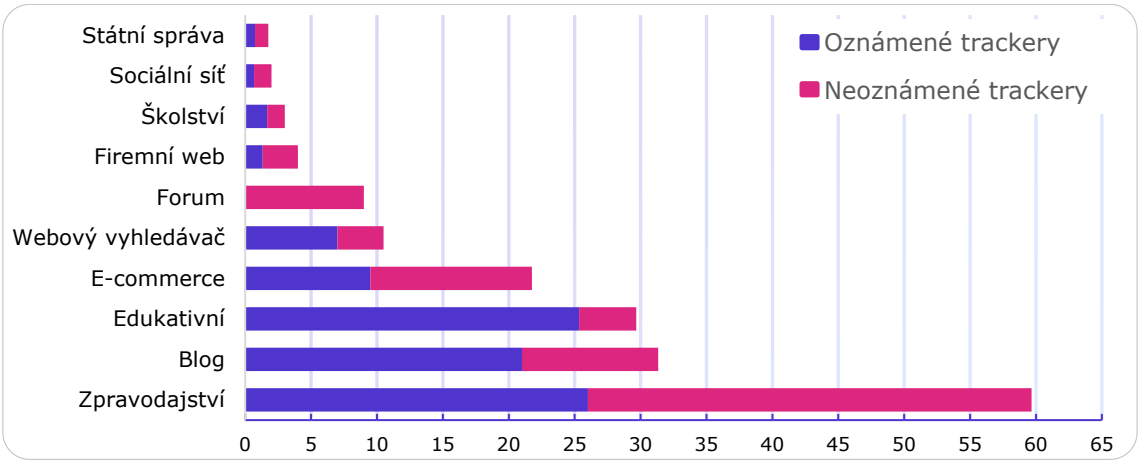
Tabulka 1 – Webové stránky nenabízející nastavení souhlasu

Spolu s množstvím třetích stran koreluje ve většině případů i poměr typu těchto třetích stran. U webů, které komunikují s větším počtem třetích stran je také větší poměr trackerů oproti obsahovým třetím stranám.



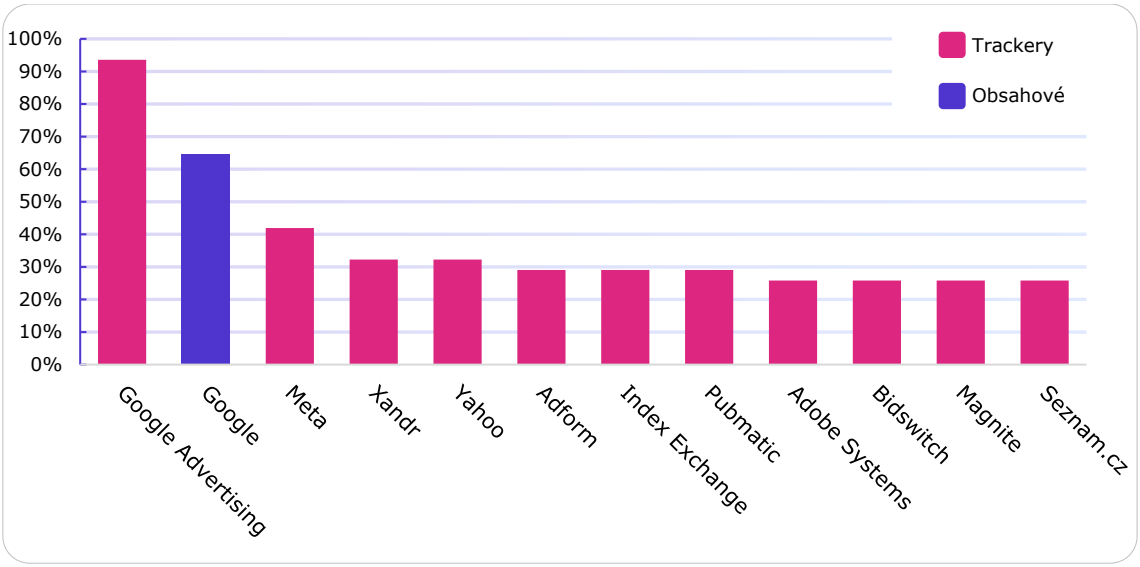
Graf 2 – Poměr obsahových třetích stran a trackerů

Neboť trackery sbírají osobní údaje uživatelů, v zájmu transparentnosti by uživatelé měli být seznámeni se všemi třetími stranami se kterými jsou jejich údaje sdíleny. Skutečné údaje z analýzy ukazují, že v průměru je uživatel seznámen jen s 53 % třetích stran pracujícími s jeho údaji. Osm webových stránek uživatele neinformuje o žádných z nalezených trackerů, z toho pět jich je českých.



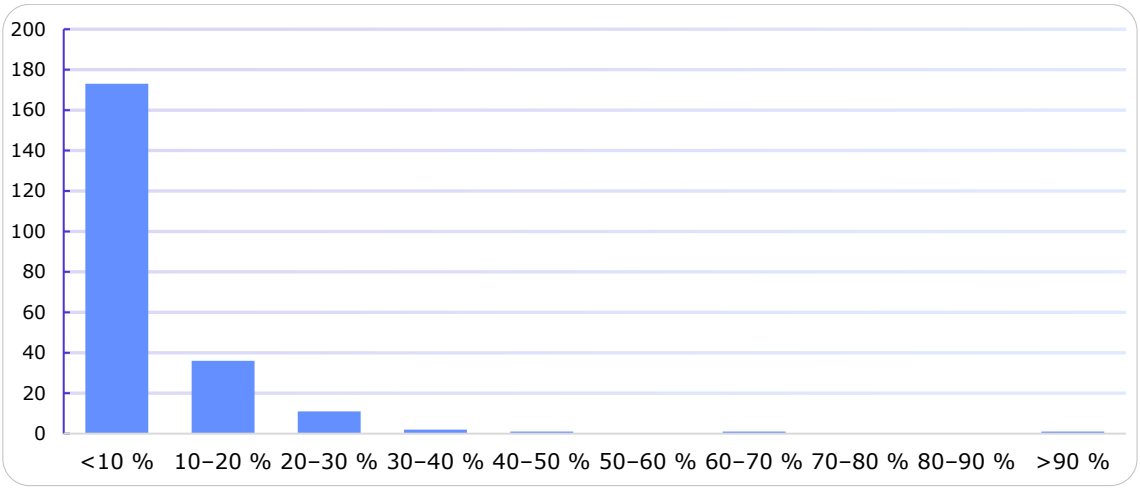
Graf 3 – Průměr (ne)oznámených trackerů v jednotlivých kategoriích webů

Společným prvkem pro téměř každou s testovaných stránek byla přítomnost Googlu, především v podobě reklam či analytik. Jedinými stránkami nevyužívající sledovacích služeb Googlu byly Hacker News a Facebook. Ostatní trackery se vyskytovali zhruba na třetině. Google zároveň poskytuje služby ve formě webových fontů, vložení map, CDN obrázků nebo YouTube videa. Obsah od Googlu byl pouze na 65 % stránek.



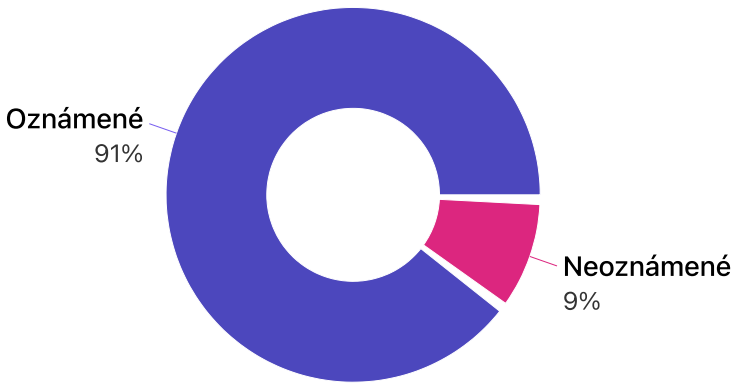
Graf 4 – Nejčetnější třetí strany

Celkem bylo na 31 stránkách 225 unikátních třetích stran. Majoritní podíl těchto třetích stran byl pouze na třech a méně stránkách. S rostoucí prominencí třetí strany zároveň logaritmicky klesá jejich počet. Složení třetích stran tedy tvoří „U“ křivku, kdy na jedné straně velké množství třetích stran na malém počtu webů, zatímco na druhé straně je malé množství třetích stran na velkém počtu webů.



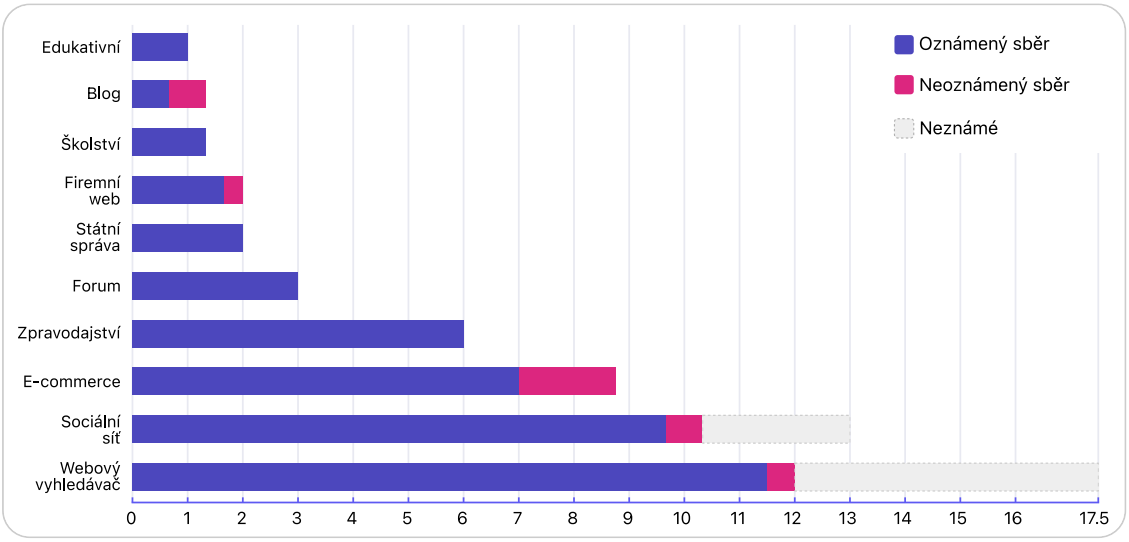
Graf 5 – Prominence třetích stran na webových stránkách

Osobní údaje sbírané přímo od prvních a druhých stran na rozdíl od třetích stran jsou více komunikovány s uživatelem. Ze zjištěných sbíraných údajů je poměr údajů, o kterých je uživatel informován 91 %. Sbírané údaje byly zjišťovány pouze přímo ze stránek, kdy údaje explicitně odesílali na server, nicméně některé společnosti sbírají údaje i ze stránek třetích stran například Google Analytics. Dá se tedy předpokládat, že procento a množství informovaného sběru je nižší v případě analýzy sběrů i mimo stránky společnosti.



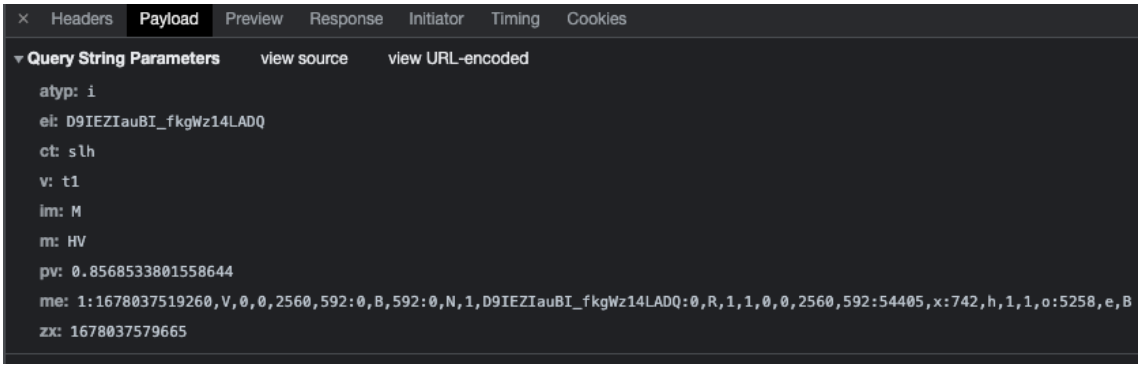
Graf 6 – Poměr oznámených a neoznámených údajů

V poměru se sbíranými údaji nejméně své uživatele informovali blogové stránky a internetové obchody. Podíváme-li se na sbírané osobní údaje první a druhých stran podle kategorií jednotlivých webů, zatímco zpravodajské weby sdílí údaje s největším počtem třetích stran, samy nesbírají tolik údajů jako webové vyhledávače. Vyhledávače prakticky třetí strany nevyužívaly, ale samy sbírají mnoho informací o uživatelích. Obdobný případ je tomu u sociálních sítí, a e-commerce. Odpovídá to předpokladu, že webové stránky, které samy těží z informací o uživateli sbírají jejich větší množství.



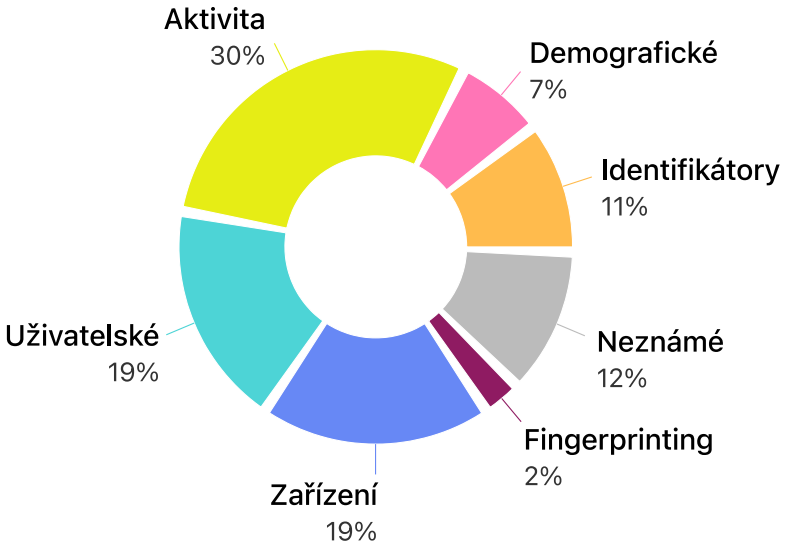
Graf 7 – Průměrný počet sbíraných údajů prvních a druhých stran dle kategorií webů

Osobní údaje prvních stran a druhých byly identifikovány a rozřazeny do kategorií. Některé webové stránky nicméně tvorbu odesílané datové části natolik znečitelní, že není možné rozklíčovat jednotlivé odesílané hodnoty (například vychází z dat vygenerovaných serverem). U takovýchto hodnot nelze posoudit, zdali je o jejich odesílání uživatel informován.



Obr. 13 – Nečitelná datová část odesílaných údajů z google.cz

Nečitelné hodnoty byly přítomné pouze u webů Facebook a Google. Naopak osm webových stránek sbíralo pouze údaje poskytnuté přímo od uživatele, jediný web přitom vůbec žádné vstupy od uživatele. Nadpoloviční většina webů vytvářela pro nepřihlášeného uživatele unikátní identifikátor. Ačkoliv všechny weby vytvářející identifikátory uživatele o této skutečnosti informovali, žádný s detekovaným fingerprintingem uživatelům tuto skutečnost nesdělil. Mezi vysokým počtem přizvaných třetích stran a počtem sbíraných údajů není spojitost.



Graf 8 – Rozdělení údajů dle jejich kategorie

Společnosti se nejvíce zaměřují na monitorování aktivity uživatelů na webové stránce. Nejčastěji zaznamenávají jednotlivé stránky, které uživatel navštívil, ale také doménu, ze které uživatel přišel. Mimo aktivitu, weby zaznamenávali informace o zařízení uživatele, typicky rozlišení, platformu a nastavený jazyk zařízení. Údaje toho typu slouží ve většině případů k analytické kategorizaci přístupových zařízení uživatelské základny. Malé množství stránek přitom zaznamenává každé kliknutí uživatele na stránce, včetně časů mezi načtením prvků a akcemi uživatele. Na přímo jen málo webů sbíralo demografické údaje o uživateli, a když ano, tak přibližnou polohu uživatele. Takovéto údaje jsou často získávány až následným zpracováním na serveru například z navštívených stránek a typu obsahu na nich. Do zahrnovaných sbíraných údajů byly započítávány pouze explicitně odeslané údaje, ale webové stránky mohou mnoho informací zjistit i nepřímým způsobem z automaticky odesílaných údajů.

7 Závěry a doporučení

V této práci byl představen vývoj webových technologií a jeho dopad na množství a druhy sbíraných osobních údajů uživatelů. V reakci na to byly v Evropě a po celém světě postupně zaváděny regulace rozšiřující práva občanů a ukládající nové povinnosti společnostem, to vše v zájmu zajištění řádné ochrany osobních údajů. Právě pokusy o ochranu osobních údajů čelí kritice z obou stran, kdy uživatelé vytýkají nedostatečnou ochranu a společnosti prosazují rozvolnění regulací, citující likvidační požadavky stran regulátorů [68][73][74][100].

Provedená analýza 31 webových stránek pomocí dostupných nástrojů cílí poskytnout obrázek o skutečném stavu ochrany osobních údajů uživatelů skrze zkoumání typů sbíraných údajů a třetí strany pracující s webovými stránkami. Prozkoumaná data ukazují, že i 5 let po zavedení GDPR v platnost některé weby dostatečně neinformují o sběru osobní údajů. Průměrná úroveň informování činí 91 % pro osobní údaje sbírané prvními a druhými strana ale u informování o všech třetích stranách na webu informuje už jen 53 %. Pochybení v seznámení se všemi sbíranými údaji bylo nalezeno u osmi webů (tří českých) a osm (čtyři z ČR) jich také neposkytlo žádné informace o třetích stranách. Devět webů (opět 4 z ČR) sbírající údaje neumožňuje uživateli odmítnout sběr kromě nezbytného pro technický provoz. Většina z těchto webů odkazuje uživatele na nastavení preferencí skrze nastavení v prohlížeči – odesílání požadavku Nesledovat (DoNotTrack), zakázání ukládání cookies. Nicméně tyto nastavení nic neřeší, cookies lze obejít skrze fingerprinting. Signál DNT nikdy nebyl podporován mnohými weby, zcela ignorován, či dokonce použit pro sledování uživatele [101].

Z analyzovaných stránek vyplynul trend, kdy weby zpeněžující uživatelské údaje, v podobě cílené reklamy, sdílí údaje s nejvíce třetími stranami. Velké nadnárodní společnosti vydělávající na uživatelských údajích naopak třetí strany příliš nevyužívají, ale sbírají větší množství údajů samy. V souvislosti lze také vnímat prominenci několik mála třetích stran na více než 90 % webů, zatímco většina zaznamenaných různých třetích stran se objevila jen na méně než 10 % webů. Toto pozorování je v souladu s výzkumem Englehardta a Narayanana na vzorku 1 milionu webových stránek [85].

Analyzované sbírané osobní údaje vycházeli pouze z rozpoznané jednoznačné komunikace mezi klientem a serverem prvních a druhých stran, aby bylo omezeno zahrnutí implicitních hodnot. Pouze společnosti Facebook a Google generování odesílaných údajů natolik znečitelnili, že není možné dopracovat se k počátečnímu iniciátoru. Webové stránky tedy pravděpodobně sbírají více údajů, přesné množství není možné zjistit bez přístupu k interním systémům, neboť mnoho se toho odehrává mimo prostor, který lze ověřit ze strany klienta. Z běžného požadavku na server, lze zjistit IP adresa (a z ní přibližná poloha), navštívené stránky, typ a verze prohlížeče a časy všech požadavků, to vyplývá ze samotné technologie http a nelze to změnit. Pokud tyto údaje společnosti ukládají se mohou uživatele dozvědět pouze z prohlášení těchto společností či provedeného auditu kontrolními úřady. Z toho vyplývá, že bez možnosti nahlédnutí „pod oponu“ jsou uživatelé odkázání na jakýsi systém důvěry v pravdivost prohlášení společností.

Uživatelé se nemusí spoléhat pouze na ochranou ruku regulačních úředníků, z výsledků analýzy doplněk *uBlock Origin* (či obdobný) ochránil soukromí uživatele před sledováním více než jiné možnosti. Jako další úroveň ochrany lze považovat vypovědění souhlasu s behaviorálním cílením reklamy na stránkách www.youronlinechoices.com často uváděných v *informacích o zpracování osobních údajů* na webových stránkách. Není však zcela jasné, zdali něco brání subjektům sbírat údaje a tento příkaz brát jen jako signál nezobrazovat cílenou reklamu [103].

Je zřejmé, že GDPR a obdobné zákony umožnily větší zapojení uživatelů do ochrany svých osobních údajů skrze zvýšenou transparentnost. Prohlášení společností byly také důležitým zdrojem pro zpracování této práce. Nicméně jak dokazuje efektivita programatických řešení jako jsou blokátory reklamy a sledování. Dalším logickým krokem posunující hranice kontroly uživatelů nad svými údaji je navržení a implementace jednotného systému oprávnění přímo v zařízeních a prohlížeči, který by s každým webem komunikoval uživatelovi preference. Dle páté schůzky Evropského parlamentu o názoru občanů jsou podobná řešení již diskutována a není nepravděpodobné, že to bude cesta, jakou se budou další regulace vyvíjet [74]. Objevují se i náznaky v podobě oživení DNT v nové podobě s legislativní podobou [102].

8 Seznam použité literatury

[1] BERNERS-LEE, Tim. *The original proposal of the WWW, HTMLized* [online]. CERN: Tim Berners-Lee, 1990 [cit. 2021-11-16]. Dostupné z: https://www.w3.org/History/1989/proposal.html

[2] Tags used in HTML. *World Wide Web Consortium* [online]. CERN: W3C, 1992 [cit. 2022-01-26]. Dostupné z: https://www.w3.org/History/19921103-hypertext/hypertext/WWW/MarkUp/Tags.html

[3] CAILLIAU, Robert; ASHMAN, Helen. Hypertext in the Web — a history. *ACM Computing Surveys*. 1999, **31**(4es). https://doi.org/10.1145/345966.346036

[4] LASAR, Matthew. Before Netscape: The forgotten Web browsers of the early 1990s. In: *Ars Technica* [online]. 2019 [cit. 2022-01-26]. Dostupné z: https://arstechnica.com/information-technology/2019/05/before-netscape-forgotten-web-browsers-of-the-early-1990s/

[5] KHALL, Wendy; TIROPANIS, Thanassis. Web evolution and Web Science. *Computer Networks*. 2012, **56**(18), 3859-3865. https://doi.org/10.1016/j.comnet.2012.10.004

[6] CLARK, David D.; WROCLAWSKI, John; SOLLINS, Karen R. et al. Tussle in cyberspace. *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications - SIGCOMM '02*. New York, New York, USA: ACM Press, 2002, 2002, , 347-. https://doi.org/10.1145/633025.633059

[7] BRIN, Sergey; PAGE, Lawrence. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*. 2012, **56**(18), 3825-3833. https://doi.org/10.1016/j.comnet.2012.10.007

[8] MCFERRIN, Joe, The History of eCommerce: How Did it All Begin. In: *IWD Agency* [online]. 2021 [cit. 2023-03-11]. Dostupné z: https://www.iwdagency.com/blogs/news/the-history-of-ecommerce-how-did-it-all-begin

[9] CHU, Sung-Chi; LEUNG, Lawrence C.; HUI, Yer Van et al. *Evolution of e-commerce Web sites: A conceptual framework and a longitudinal study*. 2007, **44**(2), 154-164. https://doi.org/10.1016/j.im.2006.11.003

[10] GULER, Bekir; TUFAN, Kadir. Unsuccessful e-commerce stories Dotcom boom. *2013 7th International Conference on Application of Information and Communication Technologies*. IEEE, 2013, 2013, , 1-4. https://doi.org/10.1109/ICAICT.2013.6722640

[11] The History of the Web [online]. New York: Jay Hoffmann, 2021 [cit. 2021-11-16]. Dostupné z: https://thehistoryoftheweb.com/

[12] GREER, Jennifer D. Advertising on traditional media sites: Can the traditional business model be translated to the Web? *The Social Science Journal*. 2004, **41**(1), 107-113. https://doi.org/10.1016/j.soscij.2003.10.009

[13] LEI, Richard M. An assessment of the World Wide Web as an advertising medium. *The Social Science Journal*. 2000, **37**(3), 465-471. https://doi.org/10.1016/S0362-3319(00)00081-1

[14] FUCHS, Jay. How Facebook Ads Have Evolved [+What This Means for Marketers]. In: *HubSpot* [online]. 2020, 2021 [cit. 2023-03-11]. Dostupné z: https://blog.hubspot.com/marketing/history-facebook-adtips-slideshare

[15] TAPPENDEN, Andrew F.; MILLER, James. A survey of cookie technology adoption amongst nations. *Journal of Web Engineering*. 2009, **8**(3), 211-244. Dostupné z: https://journals.riverpublishers.com/index.php/JWE/article/download/4051/2825

[16] HODGES, Jeff; CORRY Bil. *'HTTP State Management Mechanism' to Proposed Standard* [online]. The Security Practice, 2011 [cit. 2021-11-16]. Dostupné z: https://www.thesecuritypractice.com/the_security_practice/2011/03/http-state-management-mechanism-to-proposed-standard.html

[17] HESTERBERG, Karla, A Brief History of Online Advertising In: *HubSpot* [online]. 2021 [cit. 2023-03-13]. Dostupné z: https://blog.hubspot.com/marketing/history-of-online-advertising

[18] OKO Digital, The History of Online Advertising - 1994 to the present. In: *OKO Digital* [online]. 2019 [cit. 2023-03-13]. Dostupné z: https://oko.uk/blog/the-history-of-online-advertising

[19] DONALDSON, Dean, *Online Advertising History*. Flash by name, Cookies by nature [online]. Londýn: Bournemouth University, 2008 [cit. 2023-03-13]. Dostupné z: https://nothingtohide.us/wp-content/uploads/2008/01/dd_unit-1_online_advertsing_history.pdf

[20] HTTP State Management Mechanism. *IETF Datatracker* [online]. Mountain View (California): Network Working Group, 1997 [cit. 2021-11-16]. Dostupné z: https://datatracker.ietf.org/doc/html/rfc2109#section-8.3

[21] SIEGEL, David. The Web Is Ruined and I ruined it. *XML.com* [online]. Sebastopol (California): O'Reilly Media, 1997 [cit. 2021-11-16]. Dostupné z: https://www.xml.com/pub/a/w3j/s1.people.html

[22] Web Design History Timeline 1990-2021 / *Web Design Museum* [online]. Prague: Web Design Museum, 2021 [cit. 2021-11-16]. Dostupné z: https://www.webdesignmuseum.org/web-design-history

[23] SHANE, Tommy. The psychology of misinformation: Why we're vulnerable In: *First Draft* [online]. 2020 [cit. 2023-03-14]. Dostupné z: https://firstdraftnews.org/articles/the-psychology-of-misinformation-why-were-vulnerable/

[24] ROSEN, Deborah E.; PURINTON, Elizabeth. Website design. *Journal of Business Research*. 2004, **57**(7), 787-794. https://doi.org/10.1016/S0148-2963(02)00353-3

[25] WIRFS-BROCK, Allen; EICH, Brendan. JavaScript: the first 20 years. *Proceedings of the ACM on Programming Languages*. 2020, **4**(HOPL), 1-189. https://doi.org/10.1145/3386327

[26] KALUŽA, Marin; VUKELIĆ, Bernard. Comparison of front-end frameworks for web applications development. *Zbornik Veleučilišta u Rijeci*. 2018, **6**(1), 261-282. https://doi.org/10.31784/zvr.6.1.19

[27] *CGI: Common Gateway Interface* [online]. MIT: W3C [cit. 2022-01-29]. Dostupné z: https://www.w3.org/CGI/

[28] PRECHELT, Lutz. *Are Scripting Languages Any Good? A Validation of Perl, Python, Rexx, and Tcl against C, C , and Java*. Elsevier, 2003, 2003, , 205-270. Advances in Computers. https://doi.org/10.1016/S0065-2458(03)57005-X

[29] Usage Statistics and Market Share of PHP for Websites, November 2021.c In: *W3Techs* [online]. W3Techs, 2021 [cit. 2021-11-16]. Dostupné z: https://w3techs.com/technologies/details/pl-php

[30] History of PHP. *PHP* [online]. PHP, 2021 [cit. 2021-11-16]. Dostupné z: https://www.php.net/manual/en/history.php.php

[31] BĄK, Tomasz. Companies Using PHP by Domain. In: *SoftKraft* [online]. Bielsko-Biała: SoftKraft [cit. 2022-01-29]. Dostupné z: https://www.softkraft.co/companies-using-php/

[32] YADAV, Neha; RAJPOOT, Dharmveer Singh; DHAKAD, Shri Krishna. LARAVEL: A PHP Framework for E-Commerce Website. *2019 Fifth International Conference on Image Information Processing (ICIIP)*. IEEE, 2019, 2019, , 503-508. https://doi.org/10.1109/ICIIP47207.2019.8985771

[33] LEI, Kai; MA, Yining; Zhi TAN. Performance Comparison and Evaluation of Web Development Technologies in PHP, Python, and Node.js. *2014 IEEE 17th International Conference on Computational Science and Engineering*. IEEE, 2014, 2014, , 661-668. https://doi.org/10.1109/CSE.2014.142

[34] JOVICIC, Suzana. Scrolling and the In-Between Spaces of Boredom: Marginalized Youths on the Periphery of Vienna. *Ethos*. 2020, **48**(4), 498-516. ISSN 0091-2131. Dostupné z: https://doi.org/10.1111/etho.12294

[35] LAMP Stack. *IBM* [online]. New York: IBM Cloud Education, 2021 [cit. 2021-11-16]. Dostupné z: https://www.ibm.com/cloud/learn/lamp-stack-explained

[36] Database. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2023 [cit. 2023-03-15]. Dostupné z: https://en.wikipedia.org/wiki/Database#Classification

[37] KUMAR, Ajitesh. NoSQL Databases List & Examples - Data Analytics In: *Vitalflux.com* [online] 2022 [cit. 2023-03-15]. Dostupné z: https://vitalflux.com/nosql-databases-list-examples/

[38] PAVLO, Andrew, ASLETT, Matthew. What's Really New with NewSQL?. *ACM SIGMOD Record* [online]. 2016, **45**(2), 45-55 [cit. 2023-03-15]. ISSN 0163-5808. Dostupné z: https://doi.org/10.1145/3003665.3003674

[39] CARBONNELLE, Pierre. TOPDB Top Database index. *TOPDB index* [online]. Leuven: Pierre Carbonnelle, 2021 [cit. 2021-11-16]. Dostupné z: https://pypl.github.io/DB.html

[40] KHASAWNEH, Tariq N.; AL-SAHLEE, Mahmoud H.; SAFIA, Ali A.. SQL, NewSQL, and NOSQL Databases: A Comparative Survey. *2020 11th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2020, 2020, , 013-021. https://doi.org/10.1109/ICICS49469.2020.239513

[41] SAHATQIJA, Kosovare; AJDARI, Jaumin; ZENUNI, Xhemal et al. Comparison between relational and NOSQL databases. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, 2018, , 0216-0221. https://doi.org/10.23919/MIPRO.2018.8400041

[42] MATA, Francisco J.; QUESADA, Ariella. Web 2.0, Social Networks and E-commerce as Marketing Tools. *Journal of theoretical and applied electronic commerce research*. 2014, **9**(1), 11-12. https://doi.org/10.4067/S0718-18762014000100006

[43] VORA, Pawan. Creativity in Web Design. *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Cham: Springer International Publishing, 2019, 2019-08-11, , 381-393. Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-319-96071-5_41

[44] KIM, Woo Gon; PILLAI, Souji Gopalakrishna; HALDORAI, Kavitha et al. Dark patterns used by online travel agency websites. *Annals of Tourism Research*. 2021, **88**. https://doi.org/10.1016/j.annals.2020.103055

[45] RAJABI, Enayat; BEHESHTI, Seyed-Mehdi-Reza. Interlinking Big Data to Web of Data. *Big Data Optimization: Recent Developments and Challenges*. Cham: Springer International Publishing, 2016, 2016-05-27, , 133-145. Studies in Big Data. https://doi.org/10.1007/978-3-319-30265-2_6

[46] RAUFI, Bujar; ISMAILI, Florije; AJDARI, Jaumin et al. *Web personalization issues in big data and Semantic Web: challenges andopportunities*. 2019, **27**(4), 2379-2394. https://doi.org/10.3906/elk-1812-25

[47] MERSCH, Max; MUIRHEAD, Richard. What Is Web 3.0 & Why It Matters. In: *Medium* [online]. Luxembourg City: Fabric Ventures, 2021 [cit. 2021-11-16]. Dostupné z: https://medium.com/fabric-ventures/what-is-web-3-0-why-it-matters-934eb07f3d2b

[48] ALABDULWAHHAB, Faten Adel. Web 3.0: The Decentralized Web *Blockchain networks and Protocol Innovation*. IEEE, 2018, 2018, 1-4. https://doi.org/10.1109/CAIS.2018.8441990

[49] HUH, Seungyeon; MURALIDHARAN, Shapna; KO, Heedong et al. XR Collaboration Architecture based on Decentralized Web. *The 24th International Conference on 3D Web Technology*. New York, NY, USA: ACM, 2019, 2019-07-26, , 1-9. https://doi.org/10.1145/3329714.3338137

[50] Introducing Meta: A Social Technology Company. *Newsroom | Meta* [online]. Meta, 2021 [cit. 2021-11-16]. Dostupné z: https://about.fb.com/news/2021/10/facebook-company-is-now-meta/

[51] DOOSTI, Bardia; CRANDALL, David J.; SU, Norman Makoto. A Deep Study into the History of Web Design. *Proceedings of the 2017 ACM on Web Science Conference*. New York, NY, USA: ACM, 2017, 2017-06-25, , 329-338. https://doi.org/10.1145/3091478.3091503

[52] AMANATIDIS, Theodoros; CHATZIGEORGIOU, Alexander. Studying the evolution of PHP web applications. *Information and Software Technology*. 2016, **72**, 48-67. https://doi.org/10.1016/j.infsof.2015.11.009

[53] NASSER, Hussein. Demystifying the Browser Networking Tab in Developer Tools With Examples. In: *Youtube.com* [online]. 2020-11-10 [cit. 2023-02-10]. Dostupné z: https://www.youtube.com/watch?v=LBgfSwX4GDI

[54] LANE, Kin. Intro to APIS: History of APIs. In: *Postman* [online]. 2019 [cit. 2023-03-16]. Dostupné z: https://blog.postman.com/intro-to-apis-history-of-apis/

[55] BURNS, Kevin. WebSockets vs Long Polling. In: *DEV Community* [online]. 2021. 2022 [cit. 2023-03-16]. Dostupné z: https://dev.to/kevburnsjr/websockets-vs-long-polling-3a0o

[56] Evolution of HTTP. In: *Mozilla Web Docs*. [online]. Mozilla Corporation, 2023 [cit. 2023-03-15]. Dostupné z: https://developer.mozilla.org/en-US/docs/Web/HTTP/Basics_of_HTTP/Evolution_of_HTTP

[57] FETTE, Ian; MELNIKOV, Alexey. The WebSocket Protocol. *Request for Comments* [online]. 2011, **2011**(6455), 71 [cit. 2023-03-15]. Dostupné z: https://doi.org/10.17487/RFC6455

[58] W3Techs. Usage statistics of HTTP/2 for websites. *W3Techs - World Wide Web Technology Surveys* [online]. 2023-03-16 [cit. 2023-03-16]. Dostupné z: https://w3techs.com/technologies/details/ce-http2

[59] DIVINELEMON. Why web servers still use http 1.1 instead of http 2? In: *StackOverflow*[online]. 2021-01-27 [cit. 2023-03-16]. Dostupné z: https://stackoverflow.com/a/65924929

[60] LANDAU, Susan. FTC lawsuit spotlights a major privacy risk In: *The Conversation* [online]. 2022-08-30 [cit. 2023-03-16]. Dostupné z: https://theconversation.com/ftc-lawsuit-spotlights-a-major-privacy-risk-from-call-records-to-sensors-your-phone-reveals-more-about-you-than-you-think-189618

[61] MELUMAD, Shiri a Robert MEYER. Full Disclosure: How Smartphones Enhance Consumer Self-Disclosure. *Journal of Marketing* [online]. 2020, **84**(3), 28-45 [cit. 2023-03-16]. ISSN 0022-2429. Dostupné z: https://doi.org/10.1177/0022242920912732

[62] ORTIZ-OSPINA, Esteban. The rise of social media In*: Our World in Data* [online]. 2019-09-18 [cit. 2023-03-16]. Dostupné z: https://ourworldindata.org/rise-of-social-media

[63] DEREMUK, Iryna. Web Application Architecture: A Guide Through the Intricate Process of Building an App In: *LITSLINK* [online]. 2021-04-23 [cit. 2023-03-16]. Dostupné z: https://litslink.com/blog/web-application-architecture

[64] CHIPOLINA, Scott. Kickstarter Revises Crypto Ambitions Following Customer Backlash In: *Decrypt* [online]. 2022-02-18 [cit. 2023-03-17]. Dostupné z: https://decrypt.co/93241/kickstarter-revises-crypto-ambitions-following-customer-backlash

[65] HUANG, Kalley; ISAAC, Mike. Scott. Instagram rolls back some product changes after user backlash. In: *The New York Times* [online]. 2022-07-28 [cit. 2023-03-17]. Dostupné z: https://www.nytimes.com/2022/07/28/technology/instagram-reverses-changes.html

[66] WELLS, Georgie; HORWITZ, Jeff; SEETHARAMAN, Deepa. Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents In: *The Wall Street Journal* [online]. 2021-09-14

[cit. 2023-03-17]. Dostupné z: https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline

[67] Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů). In: *Úřední věstník*. L 119, 4.5.2016, p. 1–88. Dostupné také z: https://eur-lex.europa.eu/legal-content/cs/TXT/?uri=CELEX%3A32016R0679

[68] FORBES TECHNOLOGY COUNCIL. Unexpected Consequences Of GDPR, In: *Forbes* [online]. 2018-08-15 [cit. 2023-03-18]. Dostupné z: https://www.forbes.com/sites/forbestechcouncil/2018/08/15/15-unexpected-consequences-of-gdpr

[69] Směrnice Evropského parlamentu a Rady 2002/58/ES ze dne 12. července 2002 o zpracování osobních údajů a ochraně soukromí v odvětví elektronických komunikací (Směrnice o soukromí a elektronických komunikacích). In: *Úřední věstník*. L 201, 31.7.2002, s. 37—47. Dostupné také z: https://eur-lex.europa.eu/legal-content/CS/ALL/?uri=CELEX%3A32002L0058

[70] Zákon č. 127/2005 Sb., o elektronických komunikacích a o změně některých souvisejících zákonů (zákon o elektronických komunikacích). In: *Sbírka zákonů*. 2023. Dostupné z: https://www.zakonyprolidi.cz/cs/2005-127/zneni-20220701

[71] Směrnice Evropského parlamentu a Rady 2009/136/ES ze dne 25. listopadu 2009 , kterou se mění směrnice 2002/22/ES o univerzální službě a právech uživatelů týkajících se sítí a služeb elektronických komunikací, směrnice 2002/58/ES o zpracování osobních údajů a ochraně soukromí v odvětví elektronických komunikací a nařízení (ES) č. 2006/2004 o spolupráci mezi vnitrostátními orgány příslušnými pro vymáhání dodržování zákonů na ochranu zájmů spotřebitele. In: *Úřední věstník*. L 337, 18.12.2009, s. 11—36. Dostupné také z: https://eur-lex.europa.eu/legal-content/CS/TXT/?uri=celex%3A32009L0136

[72] LEENES, Ronald; KOSTA, Eleni. Taming the cookie monster with Dutch law – A tale of regulatory failure. *Computer Law & Security Review* [online]. 2015, **31**(3), 317-335 [cit. 2023-03-19]. ISSN 02673649. Dostupné z: https://doi.org/10.1016/j.clsr.2015.01.004

[73] SANTOS, Cristiana; BIELOVA, Nataliia; MATTE, Célestin). Are cookie banners indeed compliant with the law? *Technology and Regulation* [online]. 2020, 2020, 91–135. [cit. 2023-03-19]. ISSN 2666139X. Dostupné z: https://doi.org/10.26116/techreg.2020.009

[74] MILDEBRATH, Hendrik. The future of data protection and privacy. In: *European Parliament* [online]. 2022 [cit. 2023-03-20]. Dostupné z: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/729396/EPRS_BRI(2022)729396_EN.pdf

[75] DATAGUIDANCE. Comparing privacy laws: GDPR v. CCPA In: *FPF* [online]. 2018. [cit. 2023-03-20]. Dostupné z: https://fpf.org/wp-content/uploads/2018/11/GDPR_CCPA_Comparison-Guide.pdf

[76] California Privacy Rights Act (CPRA) | CCPA vs CPRA In: *Cookiebot* [online]. 2023. [cit. 2023-03-20]. Dostupné z: https://www.cookiebot.com/en/cpra/

[77] Bitcoin Energy Consumption Index In: *Digiconomist* [online]. 2023-03-10. [cit. 2023-03-20]. Dostupné z: https://digiconomist.net/bitcoin-energy-consumption

[78] DIEHL, Stephen. Web3 is bullshit [online]. 2021-12-04. [cit. 2023-03-20]. Dostupné z: https://www.stephendiehl.com/blog/web3-bullshit.html

[79] SCHESTOWITZ, Roy. IPFS: The Good, the Bad, and the Exceptionally Ugly In: *Techrights* [online]. 2021-09-28. [cit. 2023-03-20]. Dostupné z: http://techrights.org/2021/09/28/ipfs-issues/

[80] GRAHAM, Shauna. A List of Web3 Alternatives to Today's Popular Web 2.0 Apps In: *Medium* [online]. 2022-07-20. [cit. 2023-03-20]. Dostupné z: https://medium.com/@theshaunagraham/web3-alternatives-to-todays-popular-web-2-0-apps-97568dd6f322

[81] AHMED, Arooj. Chat GPT Achieved One Million Users in Record Time - Revolutionizing Time-Saving in Various Fields. In: *Digital Information World* [online]. 2023-01-27. [cit. 2023-03-21]. Dostupné z: https://www.digitalinformationworld.com/2023/01/chat-gpt-achieved-one-million-users-in.html

[82] What is generative AI? In: *McKinsey & Company* [online]. 2023-01-19. [cit. 2023-03-21]. Dostupné z: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

[83] VINCENT, James. 7 problems facing Bing, Bard, and the future of AI search In: *The Verge* [online]. 2023-02-09. [cit. 2023-03-21]. Dostupné z: https://www.theverge.com/2023/2/9/23592647/ai-search-bing-bard-chatgpt-microsoft-google-problems-challenges

[84] PAPADOGIANNAKIS, Emmanouil; PAPADOPOULOS, Panagiotis; KOURTELLIS, Nicolas; MARKATOS, Evangelos P. User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users. *Proceedings of the Web Conference 2021* [online]. New York, NY, USA: ACM, 2021, 2021-04-19, 2130-2141 [cit. 2023-03-21]. ISBN 9781450383127. Dostupné z: doi:10.1145/3442381.3450056 a

[85] ENGLEHARDT, Steven a Arvind NARAYANAN. Online Tracking. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* [online]. New York, NY, USA: ACM, 2016, 2016-10-24, 1388-1401 [cit. 2023-03-21]. ISBN 9781450341394. Dostupné z: doi:10.1145/2976749.2978313

[86] KRUMNOW, Benjamin; JONKER, Hugo; KARSCH, Stefan. How gullible are web measurement tools?. *Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies* [online]. New York, NY, USA: ACM, 2022, 2022-11-30, 171-186 [cit. 2023-03-22]. ISBN 9781450395083. Dostupné z: doi:10.1145/3555050.3569131

[87] LAOR, Tomer; MEHANNA, Naif; DUREY, Antonin; et al. DRAWN APART: A Device Identification Technique based on Remote GPU Fingerprinting. *Proceedings 2022 Network and Distributed System Security Symposium* [online]. Reston, VA: Internet Society, 2022, 2022, [cit. 2023-03-22]. ISBN 1-891562-74-6. Dostupné z: doi:10.14722/ndss.2022.24093

[88] Using HTTP cookies. In: *Mozilla Web Docs* [online]. 2023-03-03. [cit. 2023-03-22]. Dostupné z: https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies

[89] Chrome DevTools - Overview. In: *Chrome Developers* [online]. 2016-03-28. [cit. 2023-03-22]. Dostupné z: https://developer.chrome.com/docs/devtools/overview/

[90] HTTP response status codes. In: *Mozilla Web Docs* [online]. 2023-03-03. [cit. 2023-03-22]. Dostupné z: https://developer.mozilla.org/en-US/docs/Web/HTTP/Status

[91] DFPM is a browser extension for detecting browser fingerprinting. *Github* [online]. Github, 2019-06-11 [cit. 2023-03-23]. Dostupné z: https://github.com/freethenation/DFPM

[92] *Browserleaks - Check your browser for privacy leaks*. [online]. BrowserLeaks, 2011. 2021 [cit. 2023-03-23]. Dostupné z: https://browserleaks.com/

[93] STATISTA RESEARCH DEPARTMENT, Level of consent to the usage of cookies in selected countries worldwide as of June 2021. In: *Statista* [online]. 2023-01-10. [cit. 2023-03-23]. Dostupné z: https://www.statista.com/statistics/1273012/consent-cookies-worldwide/

[94] *Trackers | Better*. [online]. Better, 2021. [cit. 2023-03-23]. Dostupné z: https://better.fyi/trackers/

[95] *Netify Resources*. [online]. Netify, 2023. [cit. 2023-03-23]. Dostupné z: https://www.netify.ai/resources

[96] *Tracker Radar Wiki*. [online]. DuckDuckGo, 2023. [cit. 2023-03-23]. Dostupné z: https://slayterdev.github.io/tracker-radar-wiki/

[97] *WhoTracks.me – Bringing Transparency to Online Tracking*. [online]. Ghostery, 2023. [cit. 2023-03-23]. Dostupné z: https://whotracks.me/

[98] *Browser Market Share Worldwide.* [online]. Statcounter Global Stats, 2023. [cit. 2023-03-23]. Dostupné z: https://gs.statcounter.com/browser-market-share

[99] BRETOUS, Martina. A Plain English Guide to Real Time BiddingIn: *HubSpot* [online]. 2022-02-21 [cit. 2023-03-24]. Dostupné z: https://blog.hubspot.com/marketing/real-time-bidding

[100] WONG, Rebecca. Data Protection in the Online Age. *SSRN Electronic Journal* [online]. 2013, 141-143 [cit. 2023-03-26]. ISSN 1556-5068. Dostupné z: doi:10.2139/ssrn.2220754

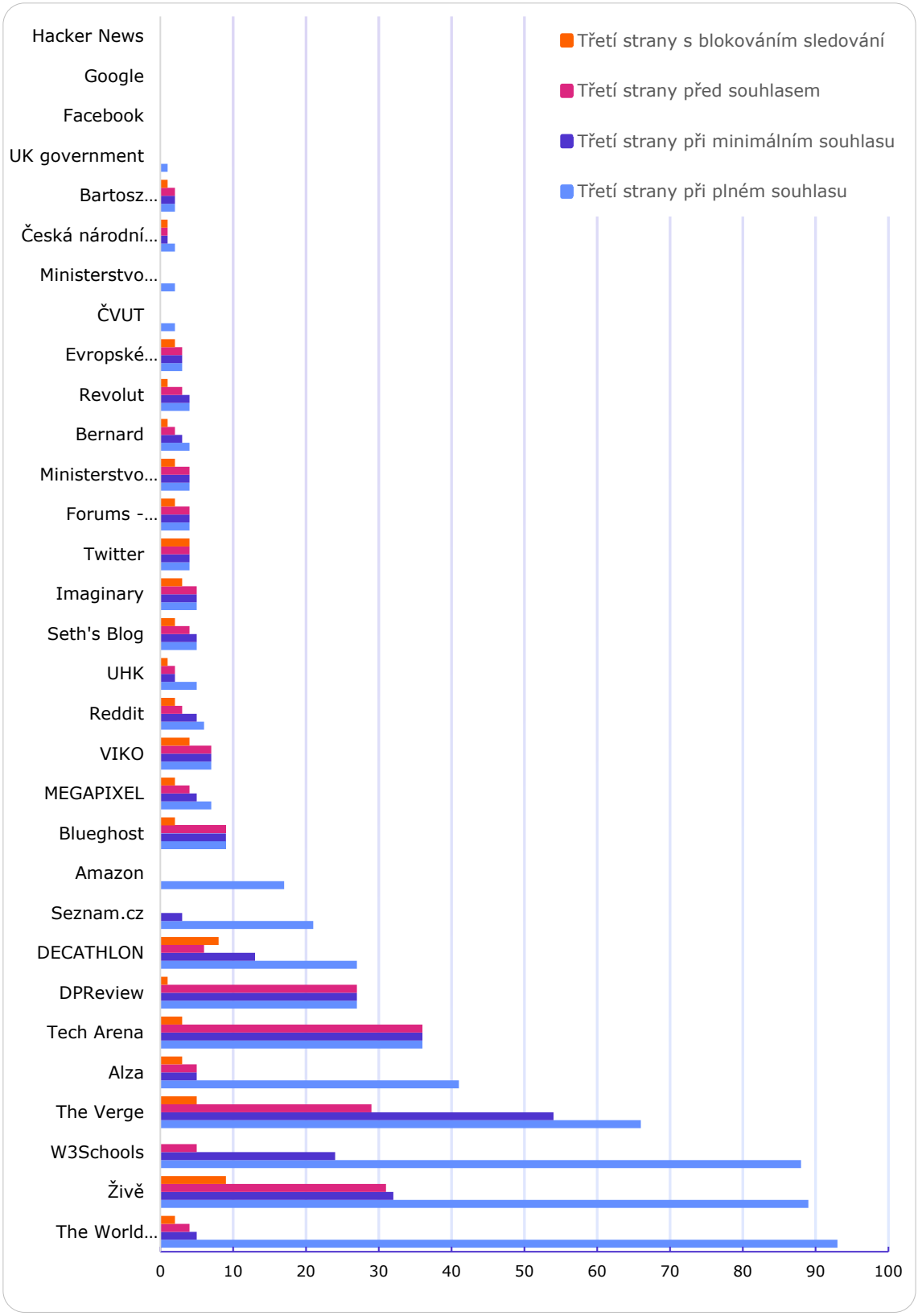
[101] SIMON, Michael. Apple is removing the Do Not Track toggle from Safari, but for a good reason. In: *Macworld* online]. 2019-02-06, [cit. 2023-03-26]. Dostupné z: https://www.macworld.com/article/232426/apple-safari-removing-do-not-track.html

[102] RUBASH, Julie. What is Global Privacy Control? Frequently Asked Questions. In: *Sourcepoint* online]. 2022-09-03, [cit. 2023-03-26]. Dostupné z: https://sourcepoint.com/blog/what-is-global-privacy-control-frequently-asked-questions/

[103] Často kladené dotazy. *Your Online Choices* [online] Brusel: EDAA., 2023 [cit. 2023-03-26]. Dostupné z: https://www.youronlinechoices.com/cz/casto-kladene-dotazy

9 Přílohy

Příloha č. 1 – Počet třetích stran při různých úrovních ochrany osobních údajů



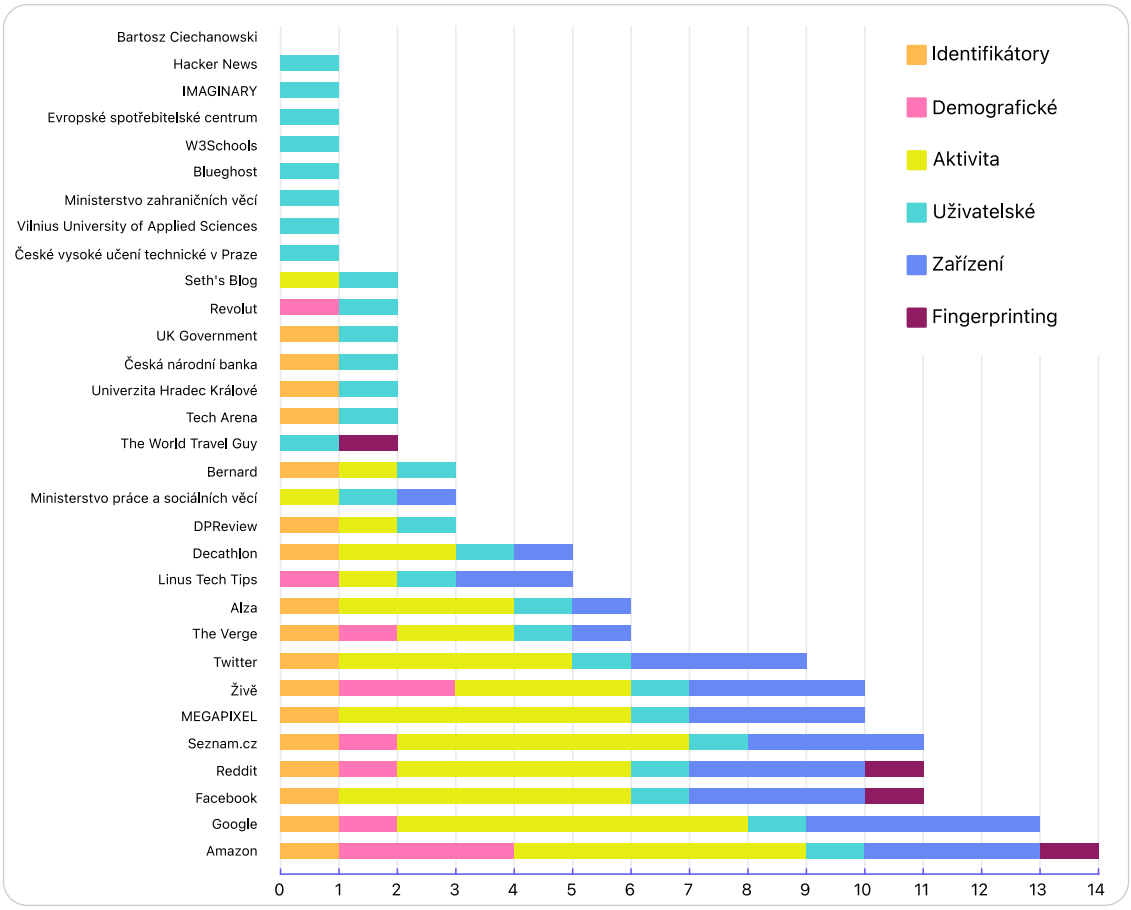
Příloha č. 2 – Počty webových stran v kategoriích

Zpravodajství	3
Blog	3
Edukativní	3
E-commerce	4
Forum	3
Webový vyhledávač	2
Firemní web	3
Školství	3
Sociální síť	3
Státní správa	4

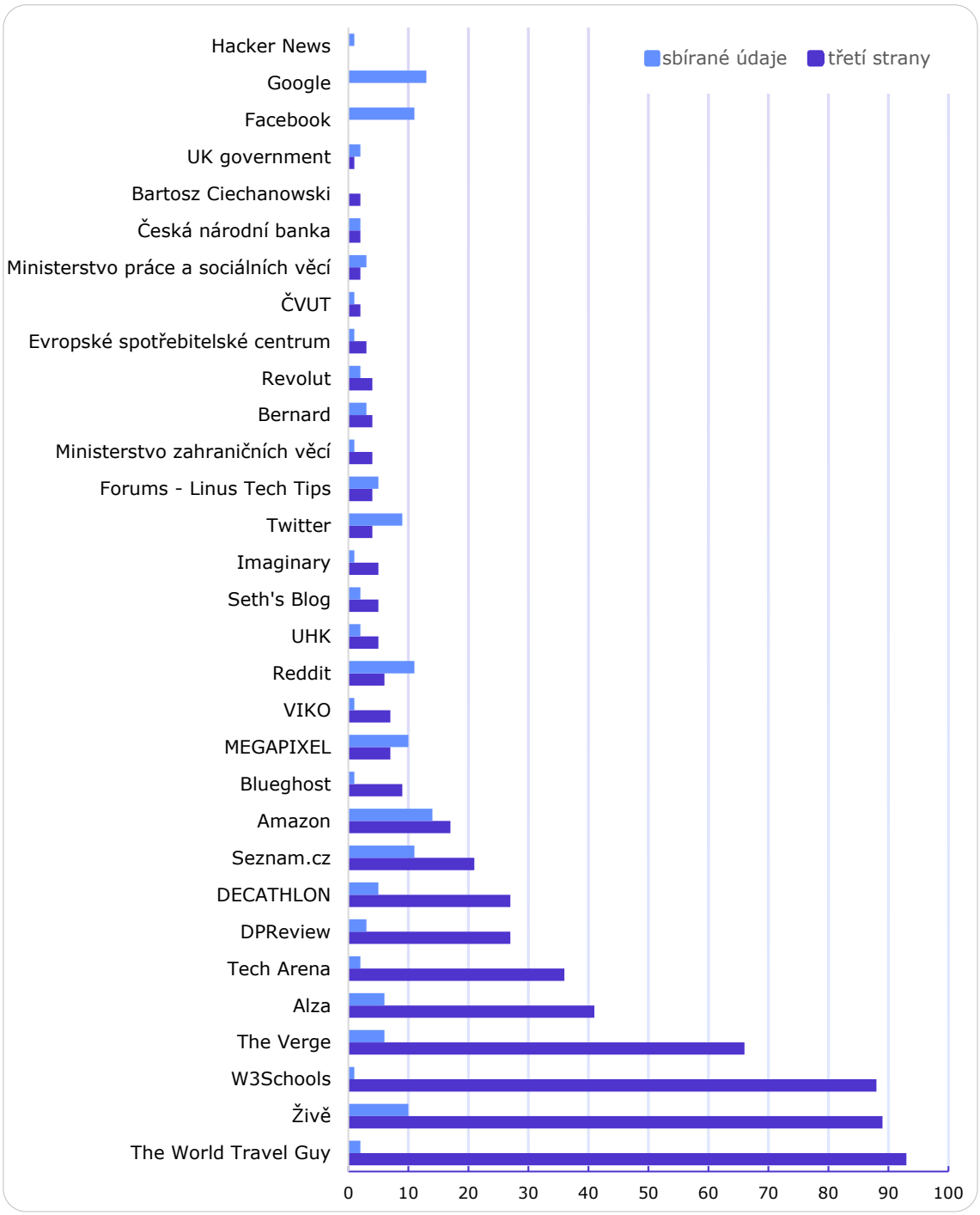
Příloha č. 3 – Webové stránky neoznamující žádné třetí strany

1. Reddit – reddit.com
2. DPReview – dpreview.com
3. České vysoké učení technické – cvut.cz
4. Ministerstvo zahraničních věcí ČR – mzv.cz
5. Blueghost – blueghost.cz
6. Evropské spotřebitelské centrum – evropskyspotrebitel.cz
7. IMAGINARY – imaginary.org
8. Bartosz Ciechanowski – ciechanow.ski

Příloha č. 4 – Počet sbíraných údajů prvních a druhých stran dle kategorií webů



Příloha č. 5 – Sbírané údaje a počet třetích stran na jednotlivých webových stránkách



Univerzita Hradec Králové
Fakulta informatiky a managementu

Zadání bakalářské práce

Autor:	**Martin Soukup**
Studium:	I1900641
Studijní program:	B0688A140001 Informační management
Studijní obor:	Informační management
Název bakalářské práce:	**Analýza sběru osobních údajů na internetu**
Název bakalářské práce AJ:	Analysis of Personal Data Collection on the Internet

Cíl, metody, literatura, předpoklady:

Cílem práce je prozkoumat neustále se vyvíjející se oblast jakou je internet a osobní údaje uživatelů. Práce se primárně zaměřuje na webové stránky, jelikož umožňují lépe kontrolovat sbírané údaje nežli například mobilní aplikace. Naprostá většina mobilních aplikací existuje také ve webové variantě.

Teoretická část práce obsahuje úvod do problematiky, stručným shrnutím vývoje webových stránek až do současné podoby. Dále představuje aktuální vládní opatření v boji proti nadměrnému sběru údajů bez kontroly uživatele. Pohled na opatření z obou stran a také ukázka reálných hrozeb neregulovaného sběru a užívání údajů.

Praktická část porovnává teoretické poznatky s praxí formou analýzy sběru údajů na vzorku webových stránek. Hlavním cílem analýzy je zjistit, zdali vládní regulace mají skutečný dopad na běžného uživatele. Druhým cílem je ověření pravdivosti prohlášení subjektů provozujících webové stránky o sbíraných uživatelských údajích.

Osnova práce:

- Úvod
- Teoretická část
  - Historie webových stránek
  - Technologický vývoj
  - Současný web a budoucnost
- Praktická část
  - Metodologie analýzy
  - Výsledky analýzy
- Zhodnocení
- Závěr

- NENADIĆ, Iva. Unpacking the "European approach" to tackling challenges of disinformation and political manipulation. Internet Policy Review. 2019, 8(4). ISSN 2197-6775. Dostupné z: doi:10.14763/2019.4.1436
- PAPADOGIANNAKIS, Emmanouil, Panagiotis PAPADOPOULOS, Nicolas KOURTELLIS a Evangelos P. MARKATOS. User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users. Proceedings of the Web Conference 2021. New York, NY, USA: ACM, 2021, 2021-04-19, , 2130-2141. ISBN 9781450383127. Dostupné z: doi:10.1145/3442381.3450056
- CAILLIAU, Robert a Helen ASHMAN. Hypertext in the Web — a history. ACM Computing Surveys. 1999, 31(4es). ISSN 0360-0300. Dostupné z: doi:10.1145/3091478.3091503
- MATA, Francisco J a Ariella QUESADA. Web 2.0, Social Networks and E-commerce as Marketing Tools. Journal of theoretical and applied electronic commerce research. 2014, 9(1), 11-12. ISSN 0718-1876. Dostupné z: doi:10.4067/S0718-18762014000100006
- DOOSTI, Bardia, David J. CRANDALL a Norman Makoto SU. A Deep Study into the History of Web Design. Proceedings of the 2017 ACM on Web Science Conference. New York, NY, USA: ACM, 2017, 2017-06-25, , 329-338. ISBN 9781450348966. Dostupné z: doi:10.1145/3091478.3091503

Zadávající pracoviště:	Katedra managementu, Fakulta informatiky a managementu
Vedoucí práce:	prof. PhDr. Marek Franěk, CSc., Ph.D.
Datum zadání závěrečné práce:	15.10.2021