UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH
PRÍRODOVEDECKÁ FAKULTA

COOKIES

2022 **Diana FORTUNOVÁ**

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH
PRÍRODOVEDECKÁ FAKULTA

COOKIES

BAKALÁRSKA PRÁCA

Študijný program:	informatika
Pracovisko (katedra/ústav):	Ústav informatiky
Vedúci diplomovej práce:	RNDr. JUDr. Pavol Sokol, PhD.
Konzultant diplomovej práce: (nepovinný)	JUDr. Laura Bachňáková Rózenfeldová, PhD.

Košice 2022

Diana FORTUNOVÁ

Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Diana Fortunová
Študijný program: informatika (jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: Informatika
Typ záverečnej práce: Bakalárska práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Cookies

Názov EN: Cookies

Cieľ: (1) Analyzovať používanie cookies z pohľadu technickej realizácie právnych požiadaviek.
(2) Porovnať aktuálne prístupy k implementácii právnych požiadaviek kladených na používanie cookies.
(3) Navrhnúť, implementovať a vyhodnotiť nástroj na testovanie používania cookies.

Literatúra:
(1) Sharma, Sanjay. Data Privacy and GDPR Handbook. John Wiley & Sons, 2019.
(2) Degeling, Martin, et al. "We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy." arXiv preprint arXiv:1808.05096 (2018).
(3) Hu, Xuehui, and Nishanth Sastry. "Characterising Third Party Cookie Usage in the EU after GDPR." Proceedings of the 10th ACM Conference on Web Science. 2019.
(4) Sanchez-Rola, Iskander, et al. "Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control." Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. 2019.
(5) Baesens, Bart. "Practical Web Scraping for Data Science Best Practices and Examples with Python." Apress (2018).

Vedúci: RNDr. JUDr. Pavol Sokol, PhD.
Konzultant: JUDr. Laura Bachňáková Rózenfeldová, PhD.
Ústav : ÚINF - Ústav informatiky
Riaditeľ ústavu: doc. RNDr. Ondrej Krídlo, PhD.

Dátum schválenia: 12.05.2021

Poďakovanie

Rada by som poďakovala svojmu vedúcemu práce RNDr. JUDr. Pavlovi Sokolovi, PhD. a svojej konzultantke JUDr. Laure Bachňákovej Rózenfeldovej za ich odborné vedenie, cenné rady a hlavne veľkú pomoc počas tvorby mojej záverečnej práce.

Abstrakt v štátnom jazyku

V tejto práci skúmame uplatňovanie právnych požiadaviek upravených vo všeobecne záväzných právnych aktoch Európskej únie a Slovenskej republiky, ktorých splnenie zaručuje zákonné použitie a používanie súborov cookies používateľmi navštevujúcim rôzne webové sídla na internete. Implementáciu týchto požiadaviek v praxi manuálne analyzujeme na príklade webových sídiel prevádzkovaných rôznymi internetovými predajcami na území Slovenskej republiky a identifikujeme bežnú prax online obchodníkov, pokiaľ ide o používanie súborov cookies na ich webových sídlach. Súčasne identifikujeme hlavné problémy, ktoré bránia úspešnej implementácii právnych požiadaviek v praxi. Následne porovnávame aktuálne dostupné prístupy k implementácii týchto požiadaviek. Výstupom tejto práce je návrh , implementácia a vyhodnocovanie testovacieho nástroja, ktorý by celý proces zautomatizoval.

Kľúčové slová: súbory cookies, osobné údaje, súkromie, GDPR, informačná bezpečnosť

Abstrakt v cudzom jazyku

In this paper, we examine the application of legal requirements regulated in generally binding legal acts of the European Union and the Slovak Republic, the fulfillment of which guarantees the legal use and use of cookies by users visiting various websites on the Internet. We manually analyze the implementation of these requirements in practice on the example of websites operated by various online retailers in the Slovak Republic and identify the common practice of online merchants regarding the use of cookies on their websites. At the same time, we identify the main problems that hinder the successful implementation of legal requirements in practice. We then compare the currently available approaches to implementing these requirements. The output of this work is to design, implement and evaluate a testing tool that would automate the whole process.

Keywords: cookies, personal data, privacy, GDPR, information security

Obsah

Zoznam ilustrácií

Zoznam tabuliek

Zoznam skratiek a značiek

kB **kilobajt,** je jednotkou kapacity pamäťových médií, 1kB = 10^3 bajtov

Slovník termínov

Ambientná autorita počítačový program používa ambientnú (okolitú) autoritu, keď určuje oprávnenia, ktoré chce vykonávať z globálneho menného priestoru. Autorita je "okolitá" v tom zmysle, že existuje v široko viditeľnom prostredí, kde si ju ostatní môžu vyžiadať podľa mena. Počítačový bezpečnostný model má ambientnú autoritu, ak týmto spôsobom poskytuje prístup k chráneným zdrojom [1].

Identifikátor relácie je jedinečné číslo, ktoré server webovej stránky prideľuje konkrétnemu používateľovi na dobu trvania session (relácie) tohto používateľa. ID relácie môže byť uložené ako súbor cookie, pole formulára alebo adresa URL (Uniform Resource Locator) [2].

DNS Systém názvov domén (DNS) je hierarchický systém názvov postavený na distribuovanej databáze. Tento systém transformuje názvy domén na adresy IP a umožňuje priraďovať názvy domén skupinám internetových zdrojov a používateľov bez ohľadu na fyzické umiestnenie entít [3].

Úvod

Existenciu ohrození ochrany základných práv a slobôd fyzických osôb na internete, najmä právo na súkromie a ochranu osobných údajov, nemožno poprieť. Tieto práva sú často ohrozené, ak nie priamo porušené pri bežných činnostiach jednotlivcov na internete. Tieto porušenia žiaľ ostávajú často nepovšimnuté. Zhromažďované údaje ako napríklad meno, IP adresa, emailová adresa, telefón a iné sa môžu použiť na profiláciu jednotlivca a tým umožniť personalizáciu reklám a marketingových stratégií na základe preferencií a aktivít identifikovaných používateľmi.

Zber údajov je možné vykonať mnohými spôsobmi. Jedna z najbežnejších metód je založená na použití súborov cookies. Tieto informácie môžu byť zhromažďované na rôzne účely, napríklad zapamätať si predvoľby používateľov (zvolený jazyk, veľkosť písma alebo iné predvoľby zobrazenia), zjednodušiť proces prihlásenia (zapamätaním si prihlasovacích údajov), analyzovať efektívnosť webových stránok (napr. prostredníctvom počtu návštevníkov a ich činnosti na danom webovom sídle) alebo efektívnosť reklám na webovej stránke. V tejto súvislosti je potrebné vyriešiť skutočnosť, že zhromaždené informácie môžu obsahovať osobné údaje používateľov webových stránok, a teda zasahovať do ich práva na ochranu osobných údajov, alebo iným spôsobom zasahovať do ich osobného života. Z dôvodu dôležitosti ochrany súkromia a osobných údajov v digitálnom prostredí je použitie súborov cookies a iných podobných technológií obmedzené platnou právnou úpravou. V tejto práci sa predovšetkým venujeme právnej úprave Európskej únie.

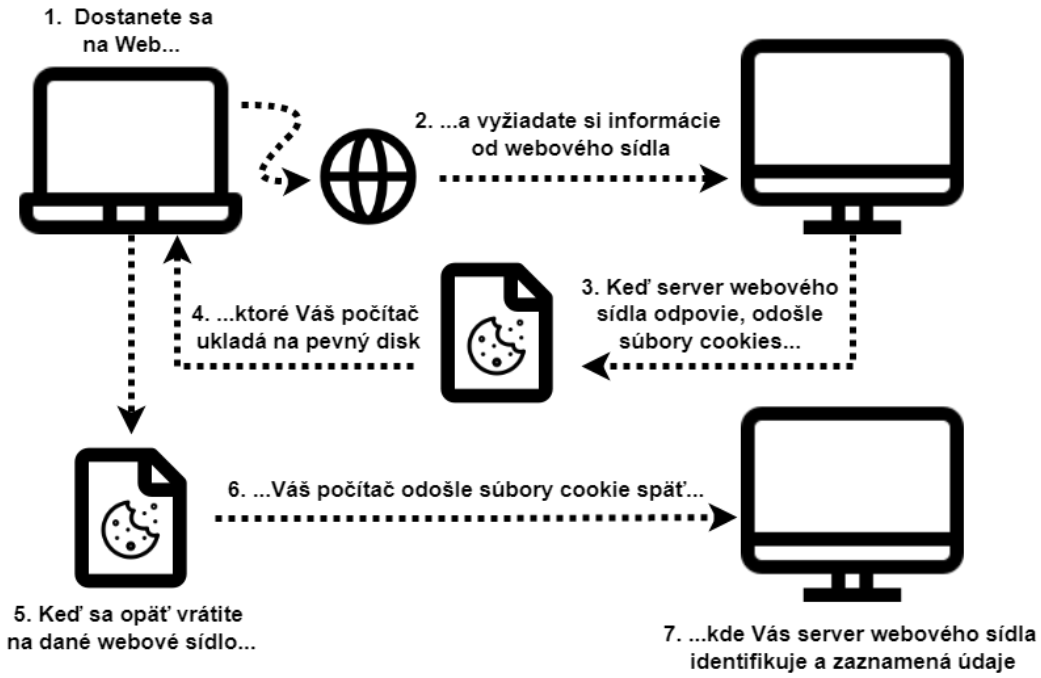
V tejto bakalárskej práci sa zameriavame hlavne na analýzu použitia a technického spracovania súborov cookies danými webovými sídlami z hľadiska právnych požiadaviek. Súčasne porovnávame aktuálne prístupy k implementácii týchto požiadaviek ako aj dostupných prostriedkov na analýzu súborov cookies používaných webovými sídlami.

V prvej kapitole popisujeme súbory cookies, ich funkcionalitu, zraniteľnosti a možnosti zneužitia týchto súborov pre škodlivé účely. Následne porovnávame už existujúce metódy a prostriedky pre analýzu webových sídiel. V tretej kapitole pomocou vhodne zvolenej sady atribútov, napríklad korektné spracovanie súhlasu s používaním súborov cookies z právneho hľadiska, použitie third-party cookies pred udelením súhlasu alebo existenciu cookies policy, hľadáme metódy ako tieto atribúty efektívne spracovať a analyzovať na vzorke webových sídiel slovenských internetových predajcov.

Na základe tejto analýzy navrhneme a implementujeme nástroj, ktorý tento manuálny proces zautomatizuje a vyhodnotí tieto údaje na rozsiahlejšej vzorke. ktorý tento manuálny proces zautomatizuje a vyhodnotí tieto údaje na rozsiahlejšej vzorke.

1 Cookies

Hypertext Transfer Protocol (HTTP) je bezstavový protokol, čo znamená, že všetky zdroje požiadaviek na server sú úplne rovnaké a server nemôže určiť, či požiadavka prichádza od klienta, ktorý už požiadavku urobil, alebo je táto požiadavka nová. HTTP taktiež nie je perzistentný. Ak vytvoríme požiadavku na server, server odošle požadované dáta a transakcia sa preň stáva ukončenou, teda neexistuje tu žiadna možnosť trvácnosti [4]. Pomocou **súborov cookies**, malých kúskov stavových dát, ktoré sú ukladané webovými sídlami vo forme textových dokumentov v zariadení používateľa [5], si však môžeme vymieňať informácie medzi serverom a prehliadačom. Je to spôsob, akým vieme poskytnúť možnosť prispôsobenia relácie používateľa a taktiež, aby servery rozpoznali používateľa medzi požiadavkami. Súbory cookies sa v podstate používajú na ukladanie ID relácie [4]. Príklad takejto relácie je znázornený na Obr. 1.



Obr. 1 Klient-server relácia s použitím súborov cookies

Po príchode na webové sídlo používateľ zadá požiadavku na webový server. Server odpovie spolu s odoslaním súborov cookies, ktoré sa uložia v zariadení používateľa. Po opätovnom príchode používateľa na dané sídlo jeho zariadenie odošle uložené súbory cookies serveru. Server tieto súbory identifikuje a zaznamená všetky dáta potrebné pre danú reláciu. Okrem toho súbory cookies slúžia aj na sledovanie (tracking cookies)**.** Tieto súbory cookies slúžia prevažne na sledovanie online aktivity používateľa medzi

všetkými stránkami a tieto informácie predáva tretím stranám za účelom online marketingu.

1.1 Obmedzenia súborov cookies

Súbory cookies so sebou neprinášajú len pozitívne stránky, ale majú aj určité obmedzenia, resp. nevýhody. Medzi tieto nevýhody zaraďujeme [6]:

1. súbory cookies nie sú schopné uniesť viac ako 4 kB údajov,
2. súbory cookies nie sú bezpečné, pretože sú ukladané v čistom texte. Môžu predstavovať možné bezpečnostné riziko, pretože ich môže otvoriť a manipulovať s nimi ktokoľvek,
3. používateľ má možnosť zakázať súbory cookies na svojom zariadení, čo predstavuje problém pre webové aplikácie, ktoré ich vyžadujú. Súbory cookies taktiež nebudú fungovať, ak je úroveň zabezpečenia v prehliadači nastavená na vysokú,
4. celkový počet súborov cookies je obmedzený (presný počet však závisí od konkrétnej implementácie prehliadača). Ak sa tento počet prekročí, nové súbory cookies nahradia staršie

1.2 Ako fungujú súbory cookies

Súbory cookies umožňujú službe založenej na HTTP, vytvárať stavové relácie, ktoré pretrvávajú vo viacerých HTTP transakciách. Keď server prijme HTTP požiadavku od klienta, môže vo svojej odpovedi obsiahnuť jeden alebo viac súborov so **Set-Cookie** hlavičkou. Klient následne interpretuje dané hlavičky odpovede **Set-Cookie** a akceptuje tie súbory cookies, ktoré neporušujú jeho pravidlá ochrany súkromia a bezpečnosti. Neskôr, keď klient odošle novú požiadavku na pôvodný server, použije hlavičku **Cookie** na prenos súborov cookies [7].

1.2.1 Set-Cookie hlavička

Set-Cookie hlavička HTTP odpovede sa používa na odosielanie súborov cookies zo servera na používateľského agenta. Mala by obsahovať názov hlavičky „Set-Cookie“, za ktorým nasleduje „**:**“ a súbor cookie. Každý súbor cookie začína s párom **meno-hodnota**, za ktorým nasleduje nula alebo viac párov **atribút-hodnota** [8].

```
Set-Cookie: <cookie-name>=<cookie-value>; Expires=<date>; Secure; HttpOnly
```

Set- Cookie hlavička obsahuje niekoľko dôležitých atribútov [9]:

1. `<cookie-name>=<cookie-value>` definuje názov súboru cookie a jeho hodnotu,
2. `Expires=<date>` označuje maximálnu životnosť súboru cookie, ako časovú pečiatku dátumu HTTP. Ak nie je zadaný, súbor cookie sa stane tzv. dočasný (session cookie), ktorých platnosť vyprší po zavretí prehliadača (ukončí sa session = relácia) [10]. Keď je nastavený dátum vypršania platnosti, daný súbor cookie sa stáva perzistentným, teda ostáva na pevnom disku, kým ho nevymaže používateľ alebo prehliadač, v závislosti od dátumu exspirácie [10]. Konečný termín sa vzťahuje na klienta, na ktorom sa súbor cookie nastavuje, nie na server,
3. `Max-Age=<number>` označuje počet sekúnd do uplynutia platnosti súboru cookie. Nulové alebo záporné číslo okamžite skončí platnosť súboru cookie. Ak sú nastavené hodnoty Expires aj Max-Age, Max-Age má prednosť. Max-age atribút nemusí byť podporovaný. Ak používateľský agent tento atribút nepodporuje, jednoducho ho ignoruje,
4. `Domain=<domain-value>` definuje hostiteľa, ktorému sa súbor cookie odošle. Ak je tento atribút nešpecifikovaný, súbor cookie je predvolene nastavený ako **HostOnly**, bez subdomén [8]. Viacnásobné hodnoty hostiteľa/domény nie sú povolené, ale ak je špecifikovaná doména, subdomény sú vždy zahrnuté,
5. `Path=<path-value>` označuje cestu, ktorá musí existovať v požadovanej adrese URL, aby prehliadač odoslal hlavičku súboru cookie. Ak server vynechá tento atribút, používateľský agent použije "adresár" komponentu

request-uri cesty ako predvolenú hodnotu [8]. Znak lomky (/) sa interpretuje ako oddeľovač adresára a priraďujú sa aj podadresáre,

6. `Secure` označuje, že súbor cookie je odoslaný na server iba vtedy, keď je zadaná požiadavka pomocou schémy https: (okrem localhostu), a preto je odolnejší voči útokom typu útočník v strede (man-in-the-middle),
7. `HttpOnly` myšlienkou je dať prehliadaču pokyn, že súbory cookies by nikdy nemali byť prístupné cez JavaScript prostredníctvom vlastnosti „document.cookie“. Táto funkcia bola navrhnutá ako bezpečnostné opatrenie, ktoré má pomôcť zabrániť cross-site scripting (XSS) útokom, páchaným krádežou súborov cookies prostredníctvom JavaScriptu [11]. Atribút HttpOnly je nezávislý od atribútu **Secure**. Súbor cookie môže mať HttpOnly aj Secure atribút [8].

1.2.2 Cookie hlavička

Hlavička Cookie HTTP požiadavky obsahuje uložené HTTP cookies spojené so serverom (t. j. predtým odoslané serverom s hlavičkou Set-Cookie alebo nastavené v JavaScripte pomocou document.cookie). Hlavička Cookie je voliteľná a môže byť vynechaná, ak napríklad nastavenia ochrany osobných údajov prehliadača blokujú súbory cookies. Mala by obsahovať názov hlavičky „Cookie“, za ktorým nasleduje „**:**“ a súbor cookie, obsahujúci zoznam párov **meno-hodnota**, v tvare `<cookie-name>=<cookie-value>`. Páry v zozname sú oddelené bodkočiarkou a medzerou „**;** “ [9]. Server nemôže zo samotnej **Cookie** hlavičky určiť, kedy platnosť súboru cookie vyprší, pre ktorých hostiteľov a pre ktoré cesty je súbor cookie platný alebo či bol súbor cookie nastavený s atribútmi Secure alebo HttpOnly [8].

```
Cookie: <cookie-list>
Cookie: name=value ; name2=value2; name3=value3
```

Keď používateľský agent prijme hlavičku **Set-Cookie**, ukladá súbory cookies spolu s ich atribútmi. Následne, pri vytváraní požiadavky HTTP, zahŕňa použiteľné súbory cookies, ktorým neuplynula doba platnosti v hlavičke **Cookie**. Ak používateľský agent dostane nový súbor cookie s rovnakým názvom, hodnotou domény a hodnotou cesty ako súbor cookie, ktorý už uložil, existujúci súbor cookie je odstránený a nahradený novým.

Servery môžu mazať súbory cookies odoslaním nového súboru cookie s atribútom **Expires** s hodnotou v minulosti. Ak atribúty súboru cookie neuvádzajú inak, súbor cookie je vrátený iba na pôvodný server a jeho platnosť vyprší na konci aktuálnej relácie [8].

1.3 Zraniteľnosti a hrozby cookies

Problémovým aspektom informačnej bezpečnosti taktiež je, že metódy útokov sa vyvíjajú rýchlo a obranné mechanizmy sa vyvíjajú relatívne pomaly. Protokol cookie bol založený na návrhu, ktorý bol podpísaný pred viac ako dvoma desaťročiami a bolo by viac než vhodné, aby sa aktualizoval na novú úroveň zabezpečenia. Z tejto skutočnosti vyplýva, že súbory cookies so sebou prinášajú množstvo bezpečnostných problémov. Medzi zraniteľnosti súvisiace s používaním súborov cookies, môžeme zaradiť [8]:

- ambientná autorita,
- čistý text,
- identifikátory relácie,
- slabá dôveryhodnosť,
- slabá integrita,
- spoliehanie sa na DNS

1.3.1 Ambientná autorita

Hoci má tento bezpečnostný problém viacero mien (CSRF, confused deputy), problém pramení z faktu, že súbory cookies sú formou ambientnej autority. Súbory cookie podporujú server operátorov v tom, aby oddelili označenie (vo forme URL) od autorizácie (vo forme cookies). V dôsledku toho môže užívateľský agent poskytnúť autorizáciu pre zdroj určený útočníkom, čo môže spôsobiť, že server alebo jeho klienti budú vykonávať akcie určené útočníkom, ako keby boli autorizované používateľom [8].

Namiesto používania súborov cookies na autorizáciu môžu operátori serverov zvážiť „zamotanie“ označenia a autorizácie tak, že adresy URL budú považovať za funkcie. Namiesto ukladania citlivých dát do súborov cookies tento prístup ukladá citlivé dáta do adries URL, čo vyžaduje, aby vzdialená entita dodala dáta sama [8]. Hoci tento prístup nie je všeliekom, rozumné uplatňovanie týchto zásad môže viesť k silnejšej bezpečnosti.

1.3.2 Čistý text

Aj keď sa súbory cookies používajú v kombinácii s HTTPS, dôvernosť a integrita súborov cookies nie sú zabezpečené. Ak sa súbory cookies neodosielajú cez zabezpečený kanál (ako je napr. TLS), informácie obsiahnuté v hlavičke súboru **Cookie** a **Set-Cookie** sa prenesú ako plain text [12]. Táto skutočnosť predstavuje niekoľko bezpečnostných obáv:

- všetky citlivé informácie prenášané v týchto hlavičkách sú vystavené odpočúvaniu,
- zlý prostredník by mohol zmeniť hlavičky pri ich ceste oboma smermi s nepredvídateľnými výsledkami,
- klient so zlým úmyslom by mohol pred prenosom zmeniť hlavičku súboru cookie s nepredvídateľným výsledkom

Servery by mali zašifrovať a podpísať obsah súborov cookies (pomocou akéhokoľvek formátu, ktorý si server želá), počas ich odoslania používateľskému agentovi (aj počas odosielania súborov cookies cez zabezpečený kanál). Zašifrovanie a podpísanie obsahu súboru cookie však nezabráni útočníkovi v prenose súboru cookie z jedného používateľského agenta na iného alebo v prehratí súboru cookie neskôr [8].

Okrem šifrovania a podpisovania obsahu každého súboru cookie by servery, ktoré vyžadujú vyššiu úroveň zabezpečenia, mali používať hlavičky súboru Cookie a hlavičky Set-Cookie iba cez zabezpečený kanál. Pri používaní súborov cookies cez zabezpečený kanál by si servery taktiež mali nastaviť atribút **Secure** (pozri 1.2.1.) pre každý súbor cookie. Ak server nenastaví tento atribút, ochrana poskytovaná zabezpečeným kanálom bude z veľkej časti zbytočná [8].

Majme príklad webového e-mailového servera, ktorý si ukladá identifikátor relácie do súboru cookie a zvyčajne k nemu pristupuje cez HTTPS. Ak server na svojich súboroch cookies nenastaví atribút **Secure**, útočník môže zachytiť každú prichádzajúcu požiadavku HTTP od používateľského agenta a presmerovať túto požiadavku na webmailový server cez HTTP. Aj keď tento server nepočúva na pripojenia HTTP, používateľský agent stále zahrnie súbory cookies do požiadavky. Aktívny sieťový útočník môže zachytiť tieto súbory cookies, prezrieť ich na serveri a zistiť obsah e-mailu

používateľa. Ak by server namiesto toho nastavil na svoje súbory cookies atribút **Secure**, používateľský agent by tieto súbory cookies nezahrnul do požiadavky vo forme čistého textu.

1.3.3 Identifikátory relácie

Namiesto ukladania informácií o relácii priamo do súboru cookie (kde môžu byť vystavené alebo prezreté útočníkom), servery v súboroch cookies bežne ukladajú identifikátory relácie. Keď server prijme požiadavku HTTP s daným identifikátorom, môže ho použiť ako kľúč a vyhľadať informácie o stave súvisiace so súborom cookie [8].

Používanie súborov cookies s identifikátorom relácie obmedzuje škody, ktoré môže útočník spôsobiť, ak sa dozvie o obsahu súboru cookie, pretože tento identifikátor je užitočný iba pre interakciu so serverom (na rozdiel od obsahu súboru cookie, ktorý neobsahuje daný identifikátor a teda môže byť sám osebe citlivý). Okrem toho použitie jedného identifikátora relácie bráni útočníkovi „spojiť“ obsah súboru cookie z dvoch interakcií so serverom, čo by mohlo spôsobiť neočakávané správanie servera.

Používanie identifikátorov relácie nie je bez rizika. Server by mal dbať na to, aby sa vyhol zraniteľnostiam typu „fixácia relácie“ (session fixation). Tento útok na fixáciu relácie prebieha v troch krokoch [8]:

- útočník prenesie identifikátor relácie zo svojho používateľského agenta do používateľského agenta obete,
- obeť používa tento identifikátor relácie na interakciu so serverom, pričom môže do identifikátora relácie vložiť poverenia používateľa alebo dôverné informácie a
- útočník používa identifikátor relácie na priamu interakciu so serverom, pričom môže získať oprávnenie používateľa alebo dôverné informácie

1.3.4 Slabá dôveryhodnosť

Súbory cookies neposkytujú izoláciu ani podľa sieťového portu, ani schémy či dokonca cesty [8]. Ak je súbor cookie čitateľný službou spustenou na jednom sieťovom porte, je súbor cookie čitateľný aj službou spustenou na inom porte toho istého servera. Rovnaký problém nastáva aj pri zapisovaní do súborov cookies. Z tohto dôvodu by servery nemali prevádzkovať vzájomne nedôveryhodné služby na rôznych portoch toho

istého hostiteľa a používať súbory cookies na ukladanie informácií citlivých na bezpečnosť.

Hoci sa súbory cookies najčastejšie používajú so schémami HTTP a HTTPS, pre daného hostiteľa môžu byť dostupné aj pre iné schémy, ako napríklad ftp a gopher. Hoci tento nedostatok izolácie podľa schémy je najzreteľnejší v API iných ako HTTP, ktoré umožňujú prístup k súborom cookie (napr. HTML document.cookie API), nedostatok izolácie podľa schémy je v skutočnosti prítomný v požiadavkách na samotné spracovanie súborov cookies.

Aj keď protokol na úrovni siete neposiela súbory cookies uložené pre jednu cestu do druhej, niektorí používateľskí agenti sprístupňujú súbory cookies prostredníctvom rozhraní API, ktoré nie sú HTTP, ako je napríklad HTML document.cookie API. Pretože niektorí z týchto používateľských agentov (napr. webové prehliadače) neizolujú zdroje prijaté z rôznych ciest, zdroj získaný z jednej cesty môže mať prístup k súborom cookies uloženým pre inú cestu [8].

1.3.5 Slabá integrita

Súbory cookies neposkytujú takmer žiadne záruky integrity. Tento nedostatok súborov cookies je známy a veľmi závažný problém. Avšak, dôsledky v reálnom svete sú často veľmi podceňované. Útoky umožnené iba jednoduchým vložením škodlivých súborov cookies môžu byť nepolapiteľné a následky môžu byť vážne.

Poznáme dve závažné formy narušenia integrity súborov cookies [13]:

1. **Prepisovanie súborov cookies** ak súbor cookie zdieľa rozsah domény so súvisiacou doménou, môže byť touto doménou priamo prepísaný pomocou iného súboru cookie s presne rovnakým názvom/doménou/cestou. Na tomto mieste potrebné poznamenať, že hoci zabezpečený súbor cookie možno prečítať iba procesom HTTPS, môže byť jednoducho zapísaný alebo prepísaný požiadavkou HTTP.
2. **Zatienenie súborov cookies** alternatívne môže útočník, ktorý ovláda súvisiacu doménu, úmyselne zatieniť súbor cookie vložením iného súboru cookie s rovnakým názvom, ale iným rozsahom domény/cesty. Ak chceme napríklad zatieniť súbor cookie s hodnotou

value=sun; domain=www.eclipse.com; Path=/; Secure

súvisiaca doména lunar.eclipse.com môže zapísať súbor cookie s hodnotou

value=moon; domain=.eclipse.com; Path=/home.

Keď neskôr prehliadač vydá požiadavku na adresu https://www.eclipse.com/home, oba súbory cookie sa zhodujú s adresou URL a sú zahrnuté. Pre väčšinu prehliadačov bude hlavička súboru Cookie bude vyzerať nasledovne:

Cookie: value=moon; value=sun;

„Slnečný“ súbor cookie môže byť zatienený „mesačným“ súborom, ak webová stránka náhodou uprednostňuje hodnotu „mesiac“ pred „slnko“. Všimnime si, že aj keď má súbor cookie s hodnotou „slnko“ zabezpečený príznak a odosiela sa cez protokol HTTPS, stále ho možno zatieniť súborom cookie nastaveným z pripojenia HTTP.

1.3.6 Spoliehanie sa na DNS

Súbory cookies sa z hľadiska bezpečnosti spoliehajú na systém názvov domén (Domain Name System - DNS). Ak je DNS čiastočne alebo úplne ohrozený, protokol cookie nemusí poskytovať bezpečnostné vlastnosti požadované aplikáciami [8].

1.4 Útoky na súbory cookies

1.4.1 Cookie poisoning

Útoky, ktorých cieľom je manipulovať, zachytiť alebo falšovať obsah súborov cookies HTTP. Sú to rôzne typy útokov, ktoré môžu mať vplyv na aplikáciu na strane klienta, na dátový prenos alebo na webový server [14].

1.4.2 Client-side Cookie poisoning

Pojem poisoning v zabezpečení webových aplikácií a zabezpečení siete sa najčastejšie používa na označenie útokov, pri ktorých má útočník za cieľ upraviť uložené informácie, ktoré neskôr s nepriaznivými účinkami použije tretia strana [14]. Napríklad DNS cache poisoning popisuje útoky, pri ktorých je obsah medzipamäte DNS upravený tak, aby nasmeroval používateľov tejto medzipamäte na škodlivé webové stránky.

Skutočná forma Cookie Poisoningu nie je bežná, pretože aplikační programátori v dnešnej dobe zriedkavo robia také základné chyby. Útok iniciuje útočník, ktorý manipuluje s obsahom súborov cookies vo svoj prospech ešte predtým, ako je súbor cookie odoslaný na webový server. Útočník musí iba stlačiť kláves F12 a pomocou

používateľského rozhrania prehliadača upraviť súbory cookies. Pokročilý útočník môže samozrejme tiež vytvoriť od začiatku vhodnú požiadavku HTTP podľa svojich potrieb.

Napríklad zle napísaná webová aplikácia môže ukladať používateľské meno aktuálne prihláseného používateľa do súboru cookie. Aplikácia potom môže pomocou obsahu súboru cookie skontrolovať, ktorý používateľ vykonáva konkrétnu operáciu. Ak je to tak, útočník môže zmeniť obsah súboru cookie tak, aby sa vydával za niekoho iného.

1.4.3 Man-in-the-Middle Cookie hijacking

Hijacking (únos) súborov cookies, je forma útoku typu man-in-the-middle (MITM). V takom prípade útočník použije inú techniku útoku na odpočúvanie komunikácie medzi webovým prehliadačom a webovým serverom a získa prístup k prenášanému obsahu súborov cookie [12].

Pri typickom útoku MITM útočník komunikáciu nielen počúva, ale môže s ňou aj manipulovať. Môže buď ukradnúť citlivé informácie obsiahnuté v prenesenom súbore cookie, alebo ich upraviť pre svoj vlastný prospech.

1.4.4 Cookies a Sessions(relácie)

Väčšina útokov zameraných na súbory cookies súvisí s identifikátormi relácií, ktoré sa najčastejšie ukladajú a prenášajú pomocou súborov cookies. Session cookies môžu byť cieľom rôznych útokov, napríklad:

- **Session hijacking:** Cieľom tohto typu útoku je ukradnúť identifikátor relácie používateľa. Útočník potom použije odcudzený identifikátor relácie na odcudzenie identity používateľa [14].
- **Session fixation:** Pri tomto type útoku útočník oklame používateľa pomocou identifikátora relácie, ktorý je už útočníkovi známy [15]. Po prihlásení používateľa môže útočník vydávať používateľa za iného. Keď útočník získa prístup k funkčnej relácii používateľa, získa plný prístup k účtu obete. Ich ďalšie kroky závisia od typu aplikácie. Napríklad v online banke môžu mať prístup k niektorým finančným operáciám a rozsiahlym osobným informáciám, ktoré sa môžu použiť na ďalšiu eskaláciu útoku.
- **Session forgery:** Falšovanie relácií je možné, ak aplikácia generuje identifikátory relácií nezabezpečeným spôsobom a útočník môže vypočítať alebo odhadnúť, ako sa identifikátor vytvára.

Sfalšovanie relácie je možné v nasledujúcich prípadoch [14]:

a) Keď sa identifikátory relácie generujú pomocou predvídateľného algoritmu. Najjednoduchším prípadom by bol identifikátor relácie, ktorý je číslom, ktoré sa pri každej novej relácii zvyšuje o jeden. Nájsť identifikátor pracovnej relácie je v takom prípade triviálne.
b) Keď sa identifikátory relácie generujú na základe nezabezpečeného generátora náhodných čísel. Nájsť identifikátor pracovnej relácie je v takom prípade ťažšie, ale stále možné.
c) Keď aplikácia neobmedzuje neúspešné pokusy. V takom prípade môže útočník použiť jednoduchý útok hrubou silou na nájdenie funkčného identifikátora relácie.

- **Cross-site Scripting:** Cross-site Scripting (XSS) útoky sú vynikajúcim spôsobom prístupu k obsahu súborov cookies vrátane identifikátorov relácií. Ak je webová aplikácia zraniteľná dokonca aj voči jednoduchému XSS, útočníkovi stačí, aby dostal identifikátor relácie a oklamal tak obeť kliknutím na dodaný odkaz [12]. Akonáhle obeť klikne na odkaz, obsah súboru cookie relácie sa odošle v žiadosti na web útočníka. V najjednoduchšom prípade útočníkovi stačí analyzovať protokoly webového servera, aby sa zobrazil obsah súborov cookies relácie.
- **Cross-site Request Forgery:** keď používateľ navštívi legitímnu stránku, obdrží legitímny súbor cookie. Potom však navštívi škodlivú stránku, ktorá prikáže prehliadaču používateľa vykonať akciu zameranú na legitímnu stránku, ktorú predtým navštívil. Požiadavka je prijatá legitímnou stránkou spolu s legitímnym súborom cookie a vykoná sa rovnaká akcia, zdanlivo iniciovaná legitímnym používateľom, ale iniciovala ju škodlivá stránka [15].

1.5 Súbory cookies z právneho hľadiska

Definícia pojmu cookies nie je obsiahnutá v žiadnych právnych aktoch, či už na národnej alebo európskej úrovni. V technickej terminológii sa tento termín vzťahuje na informácie, ktoré prechádzajú medzi pôvodným serverom a užívateľským agentom a sú uložené užívateľským agentom [8].

Z právneho hľadiska je najdôležitejšou klasifikáciou pre súbory cookies ich rozdelenie podľa triedy, teda súbory cookies prvej strany a tretej strany. Súbory cookies prvej strany (first-party) sú nevyhnutné pre používanie webového sídla, pretože ich hlavnou funkciou je rozpoznať návštevníka ako jednotlivca. Sú ukladané priamo navštívenými webovými sídlami [16]. V praxi ich koncový používateľ zvyčajne

neblokuje. Nie je na nich cielené žiadne riešenie proti sledovaniu alebo nastavenia ochrany osobných údajov z dvoch hlavných dôvodov. Prvým dôvodom je skutočnosť, že bez nich je obsah mnohých webových sídel nedostupný. Druhým dôvodom je skutočnosť, že, sú prospešné pre koncového užívateľa, napr. umožňujú automatické prihlásenie alebo prispôsobenie obsahu webového sídla [17].

Naopak, používanie súborov cookies tretích strán (third-party) vyplýva zo skutočnosti, že webové sídla obsahujú nielen svoj vlastný obsah, ale aj obsah z iných webových sídel využívajúcich ich vlastné súbory cookies. Používajú sa prevažne na reklamné účely a na webové sídlo sa umiestňujú pomocou skriptu alebo značky [16]. Z tohto dôvodu sa mnohí koncoví používatelia rozhodnú odmietnuť tieto súbory cookies, napriek tomu, že informácie o nich zhromaždené a uložené sú anonymné. Toto odmietnutie zvyčajne nezabráni koncovému používateľovi v prehliadaní webového sídla.

Ďalším relevantným právnym aspektom súborov cookies je skutočnosť, že môžu byť použité ako online identifikátory [18]. V súlade s novou právnou úpravou ochrany osobných údajov obsiahnutou vo Všeobecnom nariadení o ochrane osobných údajov (**GDPR** - General Data Protection Regulation), všeobecne platnou právnou úpravou v oblasti ochrany osobných údajov. Táto úprava sa stala účinnou 25. mája 2018. Toto nariadenie má pomerne širokú pôsobnosť, keďže ukladá povinnosti kakémukoľvek subjektu kdekoľvek na svete, pokiaľ spracúva údaje týkajúce sa občanov EÚ [19]. Podľa GDPR je možné osobné údaje definovať ako akúkoľvek informáciu týkajúcu sa identifikovanej alebo identifikovateľnej fyzickej osoby (dotknutej osoby). Dôležitým v tomto smere je, že musí ísť o fyzickú osobu a súčasne postačuje, ak bude identifikovateľná (napr. pomocou fyziologických parametrov). Pre účely identifikácie fyzickej osoby rozlišujeme rôzne faktory, pričom jedným z nich sú rôzne typy online identifikátorov. GDPR v recitáli 30 špecifikuje, že „fyzické osoby môžu byť spojené s online identifikátormi poskytovanými ich zariadeniami, aplikáciami, nástrojmi a protokolmi“. Ako jeden z príkladov takýchto identifikátorov uvádza súbory cookies. Ďalej stanovuje základný právny rámecop, ktorý existuje medzi súbormi cookies a osobnými údajmi, pričom uvádza, že zariadenia, aplikácie, nástroje a protokoly „môžu zanechávať stopy, ktoré, najmä v kombinácii s jedinečnými identifikátormi a inými informáciami prijatými servermi, môžu byť použité na vytváranie profilov. fyzických osôb a identifikovať ich“ (recitál 30 GDPR). Vzhľadom na vyššie uvedené a keďže súbory cookies samotne alebo v kombinácii s inými údajmi možno použiť na jedinečnú

identifikáciu zariadenia, údaje o jednotlivcoch priradených k zariadeniu alebo používajúcich zariadenie by sa mohli považovať za osobné údaje. To platí aj napriek tomu, že súbory cookies používajú pseudonymné identifikátory, ktoré im poskytujú jedinečnosť. Podľa autora v článku [18], bude v dôsledku toho každý súbor cookie alebo iný online identifikátor (napríklad adresa internetového protokolu) jednoznačne priradený k zariadeniu, ktoré dokáže identifikovať osobu ako jednotlivca, podliehať príslušnému nariadeniu o ochrane osobných údajov [17].

1.6 Súhlas

GDPR definuje v článku, bod 11. súhlas dotknutej osoby ako „akýkoľvek slobodne daný, konkrétny, informovaný a jednoznačný prejav vôle dotknutej osoby, ktorým formou vyhlásenia alebo jednoznačného potvrdzujúceho úkonu vyjadruje súhlas so spracovaním osobných údajov týkajúcich sa danej dotknutej osoby“.

1.6.1 Kritériá pre súhlas

GDPR a Smernica Európskeho parlamentu a Rady 2002/58/ES z 12. júla 2002, týkajúca sa spracovávania osobných údajov a ochrany súkromia v sektore elektronických komunikácií (smernica o súkromí a elektronických komunikáciách, ďalej len „smernica 2002/58/ES“) definujú nasledujúce povinné kritériá na to, aby bol súhlas považovaný za platný [18, 21]:

- **Voľne daný** - v kontexte tejto práce môžeme túto požiadavku interpretovať ako slobodnú voľbu používateľa prijať alebo odmietnuť používanie súborov cookies.
- **Špecifický**- vo všeobecnosti môžeme špecifickosť súhlasu definovať ako jeho vnútornú konzistentnosť, kedy je možné prípadné vzniknuté problémy riešiť napr. prostredníctvom výkladu [17]. V súvislosti so súbormi cookies by to znamenalo, že používatelia by vyjadrili svoj súhlas priamo s používaním konkrétnych typov súborov cookies. Používanie rôznych typov súborov cookies s rôznymi účelmi spracovania by si vyžadovalo platné mechanizmy súhlasu pre každý účel. Konkrétny súhlas vyžaduje nastavenie príznaku **HostOnly**, aby sa zabránilo súhlasu so súbormi cookies subdomény
- **Informovaný** - informovaný súhlas si vyžaduje vopred dané, jasné a komplexné oznámenie používateľa o účeloch spracovania. Informácie „by mali byť poskytnuté spôsobom, ktorý umožní bežnému používateľovi pochopiť daný

problém a následne vykonať zámernú voľbu“ [22]. V súčasnosti je používateľ oboznámený s podmienkami služby v oznámení o súhlase so súbormi cookies, ktoré obsahuje len krátke vysvetlenie, prečo webové sídlo používa súbory cookies, ale zvyčajne sa neposkytujú žiadne konkrétne informácie [17].

- **Jednoznačný** - poskytnutie jednoznačného súhlasu možno označiť za nesporné a objektívne vyjadrenie súhlasu užívateľa s používaním súborov cookies, posilňujúce princíp **Opt-in**, zavedený v smernici 2002/58/ES. Tento princíp je používaný na vyjadrenie prijatia programu používateľom alebo jeho účasti na danom programe. Opačným princípom je tzv. **Opt- out** [20]. Napríklad používateľ, ktorý sa rozhodne prihlásiť sa (opt-in) na odber noviniek internetového obchodu, začne dostávať emailové správy tohto druhu, až kým sa nerozhodne odstúpiť (opt-out). V súčasnosti väčšina webových sídel považuje získanie implicitného súhlasu za dostatočné [21]. V praxi možno na väčšine webových sídel nájsť viditeľné upozornenie týkajúce sa zhromažďovania súborov cookies spolu s odkazom na dokument so zásadami používania súborov cookies (napr. „Naša webová stránka používa súbory cookies. Pokračovaním predpokladáme Vaše povolenie ukladať súbory cookies, ako je podrobne uvedené v časti o ochrane osobných údajov a súboroch cookies.”). Významnou zmenou, ktorú GDPR uzákonilo je, že takáto prax sa už nerešpektuje. Webové sídla sú povinné zabezpečiť, aby používateľ výslovne súhlasil s používaním cookies [18]. Príkladom, ktorý spĺňa požiadavky novej právnej úpravy, je stmavnutie obrazovky po vstupe na webovú stránku s upozornením, že po získaní súhlasu budú zbierané súbory cookies. Okrem toho bude toto upozornenie obsahovať aspoň tri tlačidlá, a to: tlačidlo **Nastavenia súborov cookies**, tlačidlo **Prijať súbory cookies** a tlačidlo **Odmietnuť** [17]. Tlačidlo Prijať súbory cookies akceptuje nastavenia, ktoré si používateľ zvolí. V prípade tlačidla Odmietnuť súbory cookies nebudú súbory cookies spracovávané, kým nebudú potrebné.

1.6.2 Ďalšie požiadavky nevyhnutné pre súhlas

GDPR špecifikuje ďalšie požiadavky týkajúce sa súhlasu [18, 21]. Po prvé, ak je spracovanie údajov založené na súhlase, „prevádzkovateľ musí byť schopný preukázať, že dotknutá osoba súhlasila so spracovaním svojich osobných údajov“ podľa článku 7 ods. 1 GDPR. Na preukázanie súhlasu používateľov s používaním súborov cookies sa od

správcu bude vyžadovať, aby preukázal, že daný používateľ súhlasí s používaním konkrétnych súborov cookies, a to buď prostredníctvom nastavenia svojho prehliadača, alebo prostredníctvom súhlasu konkrétne udeleného v čase návštevy webového sídla. Digitálne stopy v podobe nastavenia webového prehliadača nemusia zaručiť, že prevádzkovateľ bude schopný preukázať poskytnutie súhlasu. Problémom v tomto smere je, že komunikáciu so serverom webového sídla iniciuje webový prehliadač. Server v podstate nevie, ktorý webový prehliadač je priradený k určitému súboru cookie. Až po inicializácii pripojenia je zariadenie identifikované. Ďalší problém môže nastať, ak používateľ vymaže všetky súbory cookies. Bude prakticky nemožné identifikovať jeho zariadenie, keďže prehliadač neposiela na server žiadne súbory cookies [18]. V tejto súvislosti jediný spôsob, ako môže prevádzkovateľ splniť túto povinnosť, je preukázať postup, ktorým bol súhlas získaný a že tento spôsob bol aplikovaný v čase získania súhlasu [17].

Po druhé, používateľ je tiež oprávnený svoj súhlas kedykoľvek odvolať a „odvolať ho bude rovnako jednoduché ako udeliť súhlas“ podľa čl. 7 ods. 3 GDPR. Z tohto ustanovenia môžeme predpokladať, že odstúpenie by malo byť prístupné rovnakým spôsobom, akým bolo vyjadrenie súhlasu, napr. v rovnakej forme, na rovnakom mieste webovej stránky alebo rovnakým postupom, akým bol predtým udelený súhlas. Technicky by zrušenie malo byť vyriešené požiadavkou používateľa, aby obnovil súhlas s tým, aby stále zahŕňal súbory cookies používateľa. Následne webové sídlo odstráni súbory cookies a všetky súvisiace údaje. Problém môže nastať, ak používateľ zablokuje odosielanie súborov cookies a vymaže ich pre danú webovú stránku na zariadení (napr. prehliadač). Koncový používateľ už nevie, ako vymazať údaje súborov cookies na webovom sídle, pretože vedel iba odoslať tento pseudoanonymizovaný identifikátor [17].

Keďže existuje možnosť poskytnutia výslovného súhlasu (napr. tlačidlo súhlasu so súbormi cookies), musí existovať možnosť jeho výslovného odvolania. Zabezpečiť jednoduchosť odvolania súhlasu nie je ľahká úloha. Požiadavky na odvolanie súhlasu nebudú splnené, ak sa používateľ bude musieť pre odvolanie preklikávať cez niekoľko úrovní nastavení [21]. Riešenie môže byť vo forme špecifickej ikony alebo tlačidla na stiahnutie, ktoré by malo byť umiestnené na prvej stránke webového sídla alebo vhodnejšie v hlavnom menu webového sídla [18]. Nakoniec musí byť dotknutá osoba pred udelením súhlasu informovaná o svojom práve na odstúpenie od zmluvy. Odvolanie

súhlasu však nebude mať retroaktívne účinky, a teda neovplyvní zákonnosť spracúvania založeného na súhlase poskytnutom pred jeho odvolaním [17].

2 Porovnanie nástrojov a metód

Táto kapitola poskytuje prehľad prác, ktoré sa sústreďujú nielen na zmeny v spracovaní údajov, súhlasov a dokumentov o ochrane osobných údajov po právnej úprave a ich analýzu. Taktiež nazerajú či už do problematiky web trackingu alebo analýzy grafického prevedenia oznámení o používaní súborov cookies a ich vplyv na koncového používateľa. Súčasťou tejto sekcie je taktiež porovnanie už dostupných riešení pre analýzu súborov cookies na vybranom webovom sídle.

2.1.1 Podobné práce

V prvej sledovanej práci [19] autori analyzovali zmeny foriem spracovania dát, úpravy súhlasov a dokumentov o ochrane osobných údajov na 500 najpopulárnejších webových sídlach pre každý štát EÚ pred a po máji 2018.

Použili systém pre automatické skenovanie webových sídel, ktorý vyhľadával zmienky o **Privacy Policy**, (zásady ochrany osobných údajov) prehláseniach na webových sídlach, ktoré popisujú, ako prevádzkovatelia sídel zhromažďujú, uchovávajú, chránia a využívajú osobné údaje poskytované používateľmi [25]. Webové sídla, bez jednoznačnej cesty, skenovali manuálne. Použili Bolerpipe – knižnicu pre extrakciu textu z HTML kódu. Na porovnanie politiky používali Jaccard similarity index a s využitím Polyglot knižnice na rozdelenie textu do viet uložili politiky ako list MD-5 hešovaných viet, aby tak urýchlili proces porovnávania. Prišli k záveru, že aj keď prevažná väčšina týchto stránok už mala politiku súkromia, veľa z nich spravilo po máji výrazné zmeny v ich znení.

Druhá časť článku sa venuje udeleniu súhlasov so spracovaním a zbieraním údajov a použitiu súborov cookies. Súhlasy hľadali a manuálne analyzovali ich podľa druhu implementácie a kategorizovali ich do skupín podľa možnosti interakcie (analýza knižníc pre kľúčové atribúty potrebné pre aktívny súhlas).

Zistili, že veľa sídel používa knižnice tretích strán pre svoje cookies súhlasy. Našli len 37 webových sídel, ktoré žiadali o explicitný súhlas s používaním súborov cookies pred tým ako ich použili. Po máji malo o 16% viac stránok cookies policy oproti januáru 2018. Veľa webových sídel v máji zakomponovalo súhlasy s používaním súborov cookies. Skoro žiadne zmeny v počte a druhu používaných súborov cookies a sledovacích služieb. Väčšina súhlasov aj tak nespĺňa GDPR požiadavky.

V druhej práci [26] sa autori zamerali na to, v akom rozsahu vie používateľ, po nadobudnutí účinnosti GDPR kontrolovať postup zvaný **web tracking,** ktorým prevádzkovatelia webových sídel zhromažďujú, ukladajú a zdieľajú informácie o aktivitách ich návštevníkov [27]. Sústreďujú sa prevažne na to, koľko trackingu je vykonávaného na používateľovi pred udelením súhlasu, čo reálne znamená odmietnuť súhlas, a ako ťažké je na stránke nájsť a zrušiť tento súhlas.

Zistili, že sledovanie je prevládajúce, prebieha väčšinou bez súhlasu používateľa je ťažké ho deaktivovať. Väčšina webových stránok vykonáva sledovanie a 92% z nich to robí skôr, ako používateľa upozorní. Manuálne analyzovali 2000 populárnych webových sídiel a prišli so sériou otázok, ktoré následne implementovali do pluginu Chrome prehliadača. Dospeli k záverom, že väčšina stránok sleduje používateľov aj bez udelenia ich súhlasu, len málo webov poskytuje jednoduchý spôsob ako zrušiť sledovanie, sledovanie je často neúčinné, väčšina webových sídel vytvára pretrvávajúce cookies identifikátory a deaktivácia prostredníctvom externých služieb nezabráni všetkému trackingu. Odmietnutie súborov cookies nemá vplyv na sledovanie. Po odmietnutí sledovania sa vytvorí viac súborov cookies.

Autori tretej práce [28] sa sústredili na základné vlastnosti grafického používateľského rozhrania oznámení o súhlase s používaním súborov cookies. Vykonali tri experimenty s viac ako 80 000 a analyzovali ako poloha oznámenia a typ súhlasu vplýva na jeho udelenie používateľom. Systematicky študovali dizajnové vlastnosti existujúcich oznámení a ich vplyvu na konanie používateľa. Zistili, že používatelia prevažne interagujú s oznámením zobrazeným v dolnej (ľavej) časti obrazovky. Taktiež, že pri binárnej voľbe je viac používateľov ochotných akceptovať sledovanie v porovnaní s mechanizmami, ktoré od nich vyžadujú, aby umožnili použitie súborov cookies pre každú kategóriu zvlášť. Ukazujú tiež, že rozšírená prax upútania pozornosti na tlačidlo súhlasu (napríklad zelenou farbou) má veľký vplyv na konečnom rozhodnutí používateľov.

2.2 Porovnanie existujúcich nástrojov

2.2.1 Metodológia

Na základe literatúry a prevládajúceho trendu vo výbere analyzovaných atribútov u sledovaných nástrojov, sme zvolili nasledujúce porovnávacie kritéria:

- **Skenované webové sídla: tento atribút udáva počet analyzovaných webových stránok domény konkrétnym nástrojom,**
- **Nájdené cookies:** udáva počet zistených súborov cookies, ktoré daná doména ukladá na zariadenie používateľa, ktorý ju navštívil,
- **Kategorizácia:** atribút sleduje, či nástroj poskytol informáciu o tom, aké typy súborov cookies sú zistiteľné z danej domény (napr. funkčné, analytické, reklamné, a. i.),
- **Hľadanie súhlasu:** zisťuje, či daný nástroj vyhľadáva banner s oznámením o používaní súborov cookies a udelením súhlasu,
- **Party klasifikácia:** atribút sleduje, či daný nástroj poskytuje informáciu o tom, koľko first a third-party cookies sa nachádza na danej webovej stránke,
- **Destinácia dát:** udáva, či sledovaný nástroj vyhľadáva a oboznamuje používateľa o tom, do ktorej krajiny sa posielajú jeho údaje,
- **Exspirácia:** sleduje, či daný nástroj poskytol informáciu o životnosti jednotlivých súborov cookies,
- **Popis:** atribút zisťuje, či daný nástroj poskytol používateľovi taktiež informáciu o tom, za akým účelom sú jednotlivé súbory cookies ukladané

Tab. 1 Predstavuje analýzu dostupných riešení voľne prístupných z prostredia Internetu. Analýza bola vyhotovená na sídle www.upjs.sk**,** v režime Inkognito a pred udelením akéhokoľvek súhlasu. Zameriava sa na atribúty, medzi ktoré zaraďujeme: koľko podstránok bolo skenovaných, koľko súborov cookies skenoval daný prehľad, či skener kategorizoval nájdené súbory cookies podľa funkcií a funkcionality, či zahrnul destináciu a i.

Tab. 1 Porovnanie voľne dostupných riešení

	1.	2.	3.	4.	5.	6.	7.	8.
Cookie script [24]	10	9	A	N	A	N	A	A
Cookie Metrix [25]	1	7	A	A	A	N	A	A
cookie- serve [26]	?	12	A	N	N	N	A	A
Piwik PRO [27]	?	6	A	N	A	N	A	N
Cookie-bot [28]	5	15	A	N	A	A	A	A
Juksta [29]	1	5	A	N	A	A	A	A
BIT SENTINEL [30]	?	7	A	N	A	N	A	A
EditThisCookie [31]	1	2	N	N	N	N	A	N
Manuálna analýza	1	13	A	A	A	A	A	N
Poskytnuté webom	?	24	A	A	A	N	A	A

Prvý stĺpec tabuľky obsahuje názov sledovaného nástroja/prístupu pre analýzy cookies na doménovom sídle ku dňu 23.3.2022.http://www.upjs.sk/

Pri prvom sledovanom atribúte (Skenovanie webového sídla) vidíme, že len veľmi málo voľne dostupných prostriedkov pre analýzu súborov cookies na webových sídlach skenuje viac ako len úvodnú stránku webového sídla. Niektoré riešenia ani neinformujú používateľa o počte skenovaných webových stránok, na ktorých by sa prípadné súbory cookies mohli nachádzať. Konkrétne sme v tejto vzorke našli len dva skenery, ktoré nahliadli aj na stránky iné, ako je tá úvodná.

Pri počte súborov cookies už dáta vyzerajú o niečo lepšie. Všetky dostupné riešenia poskytli informáciu o danej skutočnosti, z toho sa však len dve priblížili počtom nájdených cookies manuálnej analýze. Čo sa týka informácií poskytnutých webovým sídlom obsiahnutých v ich privacy policies, predpokladáme, že v danom zozname sú zahrnuté aj cookies tretích strán, ktoré budú zbierané až po ich odsúhlasení a do zariadenia používateľa sa pred udelením súhlasu neukladajú.

V treťom atribúte (Kategorizácia nájdených súborov cookies) obstál najhoršie nástroj EditThisCookie, ktorý používateľovi neposkytol žiadne informácie o typoch zbieraných cookies na analyzovanom webovom sídle.

Hľadanie oznámenia o súhlase s používaním súborov cookies umožňuje len Cookie Metrix, ktorý okrem informácie, že banner súhlasu je na stránke prítomný, dokáže aj vyhotoviť snímku obrazovky hlavnej webovej stránky.

Delenie súborov cookies podľa klasifikácie strany (party) je poskytované väčšinou skenerov. Cookie-serve a EditThisCookie však touto funkcionalitou nedisponovali.

O tom, kde všade konkrétne súbory cookies zasielajú získané dáta informuje len malé množstvo z dostupných riešení. Sú nimi nástroje Cookie-bot a Juksta.

Informáciou o životnosti sledovaných súborov cookies úspešne disponujú všetky analyzované riešenia.

Posledný atribút (Popis nájdených súborov cookies) hovorí o tom, či daný nástroj poskytol používateľovi akékoľvek dodatočné informácie o súboroch cookies, ktoré nie sú obsiahnuté v predchádzajúcich atribútoch. Prevažná väčšina týchto nástrojov poskytuje informácie ako rozsah zbieraných údajov alebo účel, za ktorým sa tieto údaje zbierajú.

3 Návrh a implementácia

3.1 Návrh systému

Obr. 2 Štruktúra nástroja na analýzu súborov cookies

Ako prvé je potrebné získať všetky dáta potrebné pre analýzu súborov cookies na sledovaných webových sídlach. Tieto dáta následne extrahujeme do vhodných typov súborov, ktoré nám zabezpečia ľahkú prácu. Následne nad týmito súbormi vyhľadáme zvolené atribúty, a zapíšeme k nim ich prislúchajúce hodnoty. Tieto hodnoty následne analyzujeme.

3.2 Extrakcia údajov

Návrh systému pre analýzu súborov cookies na nami zvolenej vzorke pozostáva z dvoch primárnych častí. Prvou z nich je získanie a extrakcia potrebných údajov, nad ktorými budeme následne vykonávať analýzu nami zvolených atribútov. V tejto kapitole si bližšie popíšeme metodiku práce pri vytváraní nástroja pre extrakciu informácií z daného webového sídla.

3.2.1 Automatizácia prehliadania webových sídel

Otvorenie a prehliadanie jedného alebo desiatich webových sídiel by sa na prvý pohľad mohlo zdať ako časovo relatívne nenáročná úloha. Problém však nastáva, ak by sme týchto webových sídiel chceli naraz prezerať viacero. Automatizáciu tohto úkonu však vieme veľmi rýchlo vyriešiť použitím WebDriver-a.

WebDriver riadi prehliadač natívne tak, ako by to robil používateľ, či už lokálne alebo na vzdialenom počítači pomocou servera Selenium. Tento systém znamená veľký skok vpred z hľadiska automatizácie prehliadača [37].

V kóde je jednoducho použiteľný po importovaní základných funkcionalít, potrebných pre jeho plynulý chod:

```
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.by import By
import selenium.webdriver as webdriver
```

Následne volaním metódy

```
driver = webdriver.Firefox(service=Service
('geckodriver-v0.30.0-win64/geckodriver.exe'))
```

WebDriver uvedieme do chodu.

3.2.2 Automatizácia ukladania nájdených údajov

Na vstupe driver dostane zoznam všetkých url adries webových sídiel, nad ktorými budeme realizovať analýzu súborov cookies. Nasleduje for cyklus, ktorý pre každú url adresu webového sídla vykoná požadovanú akciu:

- **Uloží HTML kód:** WebDriver pomocou metódy, ktorá je obsiahnutá priamo v jeho implementácii, vyhľadá a načíta obsah stránky, ktorú práve prehliada. Následne môžeme tieto dáta uložiť do vopred zvoleného priečinku, kde sa budú ukladať HTML súbory všetkých sledovaných webových sídiel:

```
#loads source code
content = driver.page_source

# saves cleaned source code
with open(folder_html + '/cleaned_source_' + hostname       +
'.html', 'w', encoding='utf-8') as outfile:
outfile.write(content)
```

- **Uloží viditeľný text hlavnej stránky:** Ďalším dôležitým postupom je uložiť viditeľný text úvodnej stránky. Na začiatku sa z raw HTML dát obsiahnutých v premennej `content` vytvorí súvislý string. Následne sa v danom dokumente postupne odstraňujú všetky tagy a časti kódu, ktoré nie sú čistým textom. Ako posledný krok sa výsledný vyčistený text zapíše do vytvoreného súboru. V spojení s už extrahovaným HTML kódom nám tieto dáta ponúkajú dobrý základ pre extrakciu informácií o dostupnosti akejkoľvek formy súhlasu so spracovaním a používaním súborov cookies jednotlivými webovými sídlami:

```
# saves visible content
doc = LH.fromstring(content)
visible_content = ""
# creates new file result_hostname.txt
with open(folder_siteText + '/result_' + hostname +
    '.txt', 'w', encoding='utf-8') as f:
#cleaning out the code parts
forelt in doc.iterdescendants():
if elt.tag in ignore_tags:
continue
text = elt.text or ''
tail = elt.tail or ''
words = ' '.join((text, tail)).strip()
    #writing into the file
if words:
f.write(words)
f.write('\n')
visible_content = visible_content +
words + '\n'
```

- **Uložiť viditeľný text policies:** Prevažná väčšina atribútov, ktoré sledujeme, nadväzujú vo veľkej miere na prítomnosť a korektnosť dokumentov o ochrane osobných údajov a dôležitých informácií spojených s dostupnými právnymi predpismi upravujúcimi súbory cookies. Cieľom tejto funkcionality je teda zistiť prítomnosť daných dokumentov a ich uloženie pre ďalšiu analýzu. Postup ukladania čistého textu je obdobný ako tomu bolo pri ukladaní čistého textu hlavnej stránky webového sídla. Poďme sa ale pozrieť na spôsob hľadania odkazov, ktoré sa na tieto dokumenty odvolávajú:

```
    elems = driver.find_elements(By.TAG_NAME, 'a')
privacy_url = ""
cookiesText_url = ""
```

Keďže prevažná väčšina odkazov na stránky v rámci webového sídla sa v HTML kóde nachádza pod tagom 'a' využijeme túto skutočnosť pre zjednodušenie hľadanie jednotlivých stránok, na ktorých majú jednotlivé sídla uložené informácie o ochrane osobných údajov svojich používateľov:

```
for elem in elems:
    elem_text = elem.get_property("text")
    if re.search(r'\bosob.+daj.+',
     elem_text,re.IGNORECASE) != None:
            privacy_url = elem.get_attribute("href")

    if re.search(r'\bcookie+',
    elem_text, re.IGNORECASE) != None:
            cookiesText_url = elem.get_attribute("href")
```

Následne pomocou regulárnych výrazov prehľadávame v poli odkazov na všetky podstránky daného webového sídla, pokiaľ nenájdeme odkaz, ktorý odkazuje na jeden alebo oba dokumenty. Tieto url adresy následne načítame obdobne, ako tomu bolo v časti o uložení HTML kódu a uložíme obsah ako čistý text to prislúchajúcich priečinkov.

- **Uložiť súbory cookies:** Ako posledné potrebujeme uložiť súbory cookies. To spravíme jednoducho a to pomocou metódy, ktorú nám ponúka WebDriver. Cookies uložíme do súboru .json pre jednoduché a prehľadné prehliadanie:

```
cookies = driver.get_cookies()
```

3.2.3 Testované atribúty

Vybrané atribúty, ktoré sme sledovali na danej vzorke (Príloha A) sa dajú rozdeliť do štyroch skupín, a to:

1. Spôsob, akým jednotlivé webové sídla umožňujú používateľom vyjadriť súhlas s používaním súborov cookies
2. Existencia a správnosť cookie policy
3. Druhy používaných súborov cookies
4. Opt-out

3.2.4 Spôsob, akým jednotlivé sídla umožňujú používateľom vyjadriť súhlas s používaním súborov cookies

d) Možnosť udelenia súhlasu s používaním súborov cookies: Má používateľ možnosť udeliť súhlas s používaním súborov cookies? Nachádza sa vôbec na sídle oznámenie o ich používaní?

Nadobúdané hodnoty: **ÁNO/NIE**

e) Aktívne udelenie súhlasu: Ak sa na webovom sídle oznámenie nachádza, je súhlas s používaním súborov cookies implicitný (Prehliadaním tejto stránky súhlasíte...)? Je aktívny (tlačidlo Súhlasím)?

Nadobúdané hodnoty: **ÁNO/NIE**

f) Možnosť výberu alternatívy súhlasu: Obsahuje súhlas binárny výber (Súhlasím/Nesúhlasím), výber preferovaných súborov cookies (checkbox, slider...)?

Nadobúdané hodnoty: **ÁNO/NIE**

g) Použitie súborov cookies pred udelením súhlasu, Použitie len first-party: Ukladá webové sídlo súbory cookies pred udelením súhlasu ? Ak áno, sú to len first-party cookies potrebné na chod stránky alebo sú to aj súbory cookies tretích strán ?

Nadobúdané hodnoty: **ÁNO/NIE**

h) Spojenie s iným právnym úkonom: Je udelenie súhlasu s používaním súborov cookies spojené s iným právnym úkonom? (udelením súhlasu na inú skutočnosť/iné spracovanie údajov)?

Nadobúdané hodnoty: **ÁNO/NIE**

i) Prístup k webovému sídlu bez udelenia súhlasu: Je sídlo prístupné aj bez udelenia súhlasu ? Zakrýva banner obsah sídla?

Nadobúdané hodnoty: **ÁNO/NIE**

3.2.5 Existencia a správnosť cookies policy

a) Existencia cookies policy: Je v zásadách ochrany osobných údajov zmienka o použití súborov cookies ?

Nadobúdané hodnoty: **ÁNO/NIE**

b) Sú v cookie policy obsiahnuté informácie ako: čo sú súbory cookies, aké typy súborov cookies webové sídlo používa, za akým účelom, aký rozsah údajov zbiera a komu tieto údaje odosiela?

Nadobúdané hodnoty: **ÁNO/NIE**

c) Poskytlo webové sídlo kontakt na prevádzkovateľa alebo zodpovednú osobu za GDPR ?

Nadobúdané hodnoty: **ÁNO/NIE**

3.2.6 Druhy používaných súborov cookies

a) Používa webové sídlo first aj third party cookies? Ukladá sídlo aj trvalé súbory cookies alebo sú všetky dočasné? Aké typy súborov cookies používa (analytické, reklamné, nevyhnutné,...)?

Nadobúdané hodnoty: **ÁNO/NIE**

b) Koľko súborov cookies webové sídlo uložilo?

Nadobúdané hodnoty: **Počet**

3.2.7 Opt-out

a) Informácia o možnosti odvolania súhlasu: Obsahuje webové sídlo informácie o možnosti odvolania poskytnutého súhlasu s používaním súborov cookies?

Nadobúdané hodnoty: **ÁNO/NIE**

b) Možnosť odvolania súhlasu v rámci sídla: je možnosť zrušenia súhlasu realizovateľná zaškrtnutím nejakého políčka, emailom, ...?

Nadobúdané hodnoty: **ÁNO/NIE**

c) Spôsob odvolania súhlasu:

Nadobúdané hodnoty: **banner, email, zaškrtávajúce políčko, žiadne ...**

3.3 Návrh hľadania atribútov

V tejto kapitole nazrieme hlbšie do toho, ako jednotlivé vybrané atribúty v našom kóde hľadať a vyhodnocovať. Opierame sa prevažne o množinu kľúčových slov získaných z manuálnej analýzy malej časti našej vzorky. Následne používame regulárne výrazy, ktoré nám pri hľadaní potrebných atribútov výrazne pomôžu:

3.3.1 Spôsob, akým jednotlivé sídla umožňujú používateľom vyjadriť súhlas s používaním súborov cookies:

1) Pri atribúte, ktorý sleduje prítomnosť možnosti udeliť súhlas s používaním súborov cookies vykonávame naše hľadanie nad súbormi obsahujúcimi HTML kód. Sústreďujeme sa prevažne na tlačidlá, ktoré by mali byť identifikátorom možnosti udeliť súhlas s používaním súborov cookies.

   Hľadané výrazy: Súhlasím, Súhlasím a pokračovať, Rozumiem, Povoliť všetko, Prijať všetko, Prijať všetky, Povoliť všetky cookies, OK, OK v poriadku

```
#Test "Súhlasím"
    suhlasim_bool = False
    print ("Možnosť udelenia súhlasu:")
    for line in lines:
        if   re.search(r"Súhlasím.*|Rozumiem.*|P.+všetk.+",
line, re.IGNORECASE):
            #print(line, end="")
            suhlasim_bool = True
```

```
    if (suhlasim_bool == True):
        print("Je uvedené Súhlasím: ÁNO")
        df_cookies.loc[index,"Suhlas"] = True
    else:
        print ("Je uvedené Súhlasím: NIE")
        df_cookies.loc[index,"Suhlas"] = False
```

2) Atribút pre aktívne udelenie súhlasu je z právneho hľadiska veľmi dôležitý a hovorí o tom aký reálny vplyv má používateľ daného webového sídla na používanie cookies. Sústreďujeme sa pri ňom na formu oznámenia ako aj formu súhlasu.

Tento web používa súbory cookie. **Ďalším prechádzaním tohto webu vyjadrujete súhlas s ich používaním.**

Tieto stránky používajú na poskytovanie služieb, personalizáciu reklám a analýzu návštevnosti súbory cookie. **Používaním týchto stránok s tým súhlasíte.**

Táto stránka používa súbory cookies, aby Vám zaistila najlepší zážitok z prehliadania. **Pokračovaním v prehliadaní tejto stránky súhlasíte s používaním súborov cookies.**

Používame cookies. **Ak budete naďalej prehliadať túto stránku, predpokladáme, že ste s ňou spokojní.**

Tento eshop používa pre skvalitňovanie služieb, správne fungovanie a analýzu návštevnosti súbory cookies. **Prehliadaním týchto stránok s tým súhlasíte.**

Implicitné súhlasy tohto typu, ktoré majú v sebe obsiahnuté aj súhlasy s inými právnymi úkonmi a skutočnosťami a preto môžeme tieto dve hľadania v našom kóde prepojiť. Následný výpis a vloženie hodnót potom prebieha rovnakým spôsobom ako pri predchádzajúcom atribúte.

```
#Test "Pokračovaním v prehliadaní"
    suhlasim_bool = False
    print ("Aktívne udelenie súhlasu:")
    for line in lines:
        if re.search(r"Pokračovaním v
prehliadaní|Pokračovaním v
prezeraní|Prehliadaním.+súhlasíte|Prezeraním.+súhlasíte",
line, re.IGNORECASE):
            #print(line, end="")
            suhlasim_bool = True
    if (suhlasim_bool == True):
```

```
        print("Je uvedené Pokračovaním v prehliadaní.
Súhlas je aktívny: NIE")
           else:
        print("Nie je uvedené Pokračovaním v prehliadaní.
Súhlas je aktívny: ÁNO")
```

3) Pri hľadaní alternatívy poskytnutia súhlasu postupujeme podobne ako v odseku "a)" kde v súbore HTML hľadáme kľúčové výrazy, ktoré naznačujú prítomnosť tohto atribútu na webovom sídle.
Hľadané výrazy: Odmietnuť všetko, Zakázať všetko, Disagree, Nechcem cookies

Ďalej hľadáme zmienku o nastavení ochrany úkromia alebo nastavenie súborov cookies, ktoré idukuje, že svoj súhlas si používateľ môže ľubovoľne upraviť podľa svojich preferencií:

Hľadané výrazy: Nastavenia, Nastavenie, Nastavenie súkromia, Nastavenie cookies, Nastaviť, Upraviť nastavenia cookies, Podrobné nastavenia, Prispôsobiť

Regulárny výraz použitý v kóde:

```
if re.search(r"Odmietnuť.*|Zakázať.*|Disagree|
Nechcem cookies", line, re.IGNORECASE):
```

3.3.2 Existencia a správnosť cookies policy

Pri hľadaní v tejto kategórii budeme pracovať so súbormi privacy policy a cookie policy. V týchto súboroch budeme následne analyzovať výskyty kľúčových výrazov:

1) Hľadanie informácie o tom, či sa v dokumente o ochrane osobných údajov nachádza zmienka o cookies je relatívne jednoduché, Stačí sa len pozrieť na to, či sa v niektorých z textov nachádza reťazec cookie.

   Regulárny výraz použitý v kóde:

   ```
   if re.search(r"cookie.+", line, re.IGNORECASE):
   ```

2) V atribúte o dostupnosti informácií o nájdených súboroch cookies máme spojenie viacerých dôležitých požiadaviek, ktorých splnenie zaručuje správnosť policies. Pozrime sa, ako tieto požiadavky vyhodnocovať:

   a) Čo sú súbory cookies:

   Túto požiadavku budeme vyhľadávať pomocou skupiny výrazov, akými sú napríklad:

Cookies sú malé textové súbory, Súbory cookie sú malé súbory, krátké textové súbory, Cookies sú textové súbory, Cookies jsou malé textové soubory, Cookies sú malé informácie, Súbory cookies sú malé štandardizované textové súbory, Cookies je malý súbor písmen a čísel, Súbor cookie je malý textový súbor,...

Alebo môžeme hľadať výrazy ako: Čo sú cookies, Čo sú súbory cookies, Čo sú súbory cookie, niečo ako nadpis, ktorým jednotlivé sídla uvedú danú kapitolu.

Regulárny výraz použitý v kóde:

```
if re.search(r"cookie.+mal.+text.+súbor.*|čo sú.+cookie.+",
line, re.IGNORECASE):
```

b) Aký typ súborov cookies:

Hľadáme výrazy ako analytické cookies, funkčné, remarketingové, marketingové, nevyhnutné, nutné, ostatné, esenciálne, preferenčné, výkonové, technické, konverzné, trackingové, prevádzkové, základné,...

Možno hľadať aj výrazy ako delenie podľa funkcie, typy súborov cookies, druhy používaných...

Regulárny výraz použitý v kóde:

```
if re.search(r"analytic.*cookie|.*marketingové.*cookie|
.*nutné.*cookie|funkčné.*cookie|esenciálne.*cookie|
konverzné.*cookie|trackingové.*cookie|preferenčné.*cookie"
, line, re.IGNORECASE):
```

c) Rozsah údajov:

Pri tomto atribúte možno hľadať výraz, ktorý bude obsahovať slová ako údaj, zhromažďovať, zhromažďované informácie, zbierané údaje,...

Regulárny výraz použitý v kóde:

```
if  re.search(r"zhromažďov.+údaj.*|zhromažďov.+informác.+",
line, re.IGNORECASE):
```

d) Komu odosiela:

Hľadať spoločnosť, Google, Facebook, Twitter,... tretie krajiny, prenášané do tretích krajín, odovzdanie, odosielanie reklamným a sociálnym sieťam, spracované ďalšími spracovateľmi, sprostredkovatelia, externí poskytovatelia,...

Regulárny výraz použitý v kóde:

```
if re.search(r"Google|Facebook|Twitter|tret.+kraj.+|
odosielan.+", line, re.IGNORECASE):
```

3) Pri existencii kontaktu na osobu zodpovednú za ochranu osobných údajov môžeme vyhľadávať výrazy ako: kto je správca, kontaktné údaje správcu/prevádzkovateľa, Spracovanie vykonáva správca, kontaktovať našu poverenú osobu, spracovávame ako správca, zodpovednú osobu, správce, osobních údaju, Prevádzkovateľom/Správcom je, Správca osobných údajov, spracovateľom, Poverenec,...

Regulárny výraz použitý v kóde:

```
if re.search(r"Správc|prevádzkovateľ.*|poveren.+osob.+|
zodpoved.+osob.+", line, re.IGNORECASE):
```

3.3.3 Druhy používaných súborov cookies

Súbory cookies na sledovaných webových sídlach budeme hľadať v súboroch s príponou .json, ktoré obsahujú zoznam všetkých súborov cookies, ktoré sa do prehliadača našou návštevou uložili. Vychádzame zo skutočnosti, že všetky cookies, ktorých doménový atribút je viazaný na doménu daného sídla, môžeme považovať za first-party súbor cookie. Teda v tomto súbore budeme hľadať pomocou premennej hostname, názvu domény, ktorý sme si v kóde uložili pre zjednodušenie práce. Taktiež prechádzaním cez všetky doménové atribúty v súbore vieme spočítať celkový počet súborov cookies, ktoré sa v súbore nachádzajú.

```
#Test "Len first-party cookies"
    #firstParty_bool = False
    counter_first = 0
    c = 0
    result = 0
    print ("Len first-party cookies:")
    for line in lines:
        if re.search(r"\bdomain.+,$", line, re.IGNORECASE):
            #print(line, end="")
            #firstParty_bool = True
            c = c +1
```

```
        if   re.search(r"\bdomain.+:.+%s.+,$"   %   name,   line,
 re.IGNORECASE):
            #print(line, end="")
            #firstParty_bool = True
            counter_first = counter_first +1
    result = c - counter_first
    if (result  != 0):
        print("NIE")
    else:
        print("ÁNO")

    print("-------------")
    print("Celkový počet cookies je: " + str(c))
```

Keďže časová pečiatka v atribúte udávajúcom expiráciu súboru cookie je v type UNIX time, môžeme si jednoducho zistiť čas uloženia súboru .json, ktorý obsahuje súbory cookies, v rovnakom type a následne vykonať porovnanie daných dvoch časových pečiatok. Ak je čas uvedený v atribúte "expiry" súboru cookie väčší ako dátum uloženia súboru o nejakú vopred zvolenú časovú jednotku, súbor cookie je perzistentný, inak ho môžeme považovať za dočasný.

Hľadanie a klasifikáciu cookies podľa funkcionality vykonáme s pomocou predpripraveného zoznamu často používaných cookies, ktoré sme spolu s ich prislúchajúcou kategóriou extrahovali z voľne dostupnej databázy [38].

```
#Test "Delenie podľa funkcionality"
    for cookies in cookies_list:
            #print(cookies)
            #print(r"\Functional,%s" %cookies)
            if  re.search(r"Functional,%s"  %cookies,  database,
    re.IGNORECASE):
                functional = functional +1
            if  re.search(r"Analytics,%s"  %cookies,  database,
    re.IGNORECASE):
                analytics = analytics +1
            if  re.search(r"Marketing,%s"  %cookies,  database,
    re.IGNORECASE):
                marketing = marketing +1
            if     re.search(r",%s"     %cookies,     database,
    re.IGNORECASE) == None:
                unknown = unknown +1
```

3.3.4 Opt-out

Pri našom poslednom atribúte nazrieme opäť do súborov obsahujúcich čistý text oboch policies. Nad týmito textami následne spustímehľadanie výrazov ako Odmietnutie/Odmietnuť (používanie) súborov cookies, zakázať cookies,...

Regulárny výraz použitý v kóde:

```
If re.search(r"Odmietnu.+cookie.*|
zakázať cookie.*|Odvola.+súhlas.+cookie", line, re.IGNORECASE):
```

4 Testovanie používania súborov cookies v rámci e-shopov

4.1 Manuálna analýza

Keďže sme si správnosť údajov získaných pomocou nášho crawlera potrebovali otestovať a taktiež zistiť, ktoré výrazy sa naprieč webovými sídlami vyskytujú najčastejšie, rozhodli sme sa vyhotoviť manuálnu analýzu na zlomku sledovanej vzorky. Tieto výstupy nám vo veľkej miere pomohli zefektívniť náš nástroj, ako aj poskytnúť akúsi testovaciu sadu, na ktorej okamžite uvidíme správnosť našej implementácie.

Ako sme však očakávali, technické spracovanie cookies z právneho hľadiska a poskytnuté informácie o ochrane osobných údajov a používaní súborov cookies jednotlivými prevádzkovateľmi malo veľa medzier:

1. V kategórii súhlas poskytlo používateľom možnosť udeliť súhlas s používaním súborov cookies 85% webových sídiel, možnosť aktívneho súhlasu poskytlo 45%, možnosť výberu alternatívy mali všetky sledované eshopy a v spojení s iným právnym úkonom sa súhlas vyskytol v 15%
2. Pri manuálnej analýze tejto vzorky sme taktiež prišli na to, že prevažná väčšina webových sídiel ukladá súbory cookies už pred udelením súhlasu používateľom. Tieto súbory cookies sú ale prevažne first-party cookies. Väčšina webových sídiel ukladala aj dočasné, aj trvalé súbory cookies. Čo sa týka súborov cookies z hľadiska funkcie, väčšina predajcov ukladala funkčné aj analytické cookies potrebné pre správny chod ich stránok. 67,5% ukladalo aj marketingové cookies.
3. O správnosti policies rozhodovala prítomnosť atribútov. Informáciu o tom, čo sú súbory cookies obsiahlo vo svojich dokumentoch 55% sledovaných webových sídiel. Informácie o rôznych typoch používaných cookies obsahovalo 15% zo všetkých policies, rozsah zbieraných údajov len 22,5% a informácie o tom, kam údaje o používateľoch smerujú nám dalo len 5%.
4. Len 15,5% nemalo na svojich stránkach zverejnený kontakt na zodpovednú osobu za spracovanie osobných údajov.
5. Informáciu o možnosti odvolať súhlas s používaním súborov cookies poskytlo 37,5% webových sídiel, z toho len 17% umožnilo tento súhlas odvolať v rámci niektorej zo stránok. V spôsobe odvolania súhlasu prevládala možnosť kontaktovať správcu.

4.2 Analýza pomocou nástroja

Náš nástroj sme následne spustili nad celou vzorkou, ktorá obsahuje 100 najlepších online predajcov v rámci Slovenskej republiky. Pre jednotlivé atribúty nástroj vyhodnotil nasledujúce hodnoty:

1. Kategória, v ktorej sledujeme technické a právne spracovanie súhlasu s používaním súborov cookies obstála v rámci eshopov najlepšie. Len 4% všetkých webových sídiel nemalo na svojich stránkach možnosť udelenia súhlasu ani možnosť alternatívy súhlasu, či už tlačidlom, alebo možnosťou nastaviť si preferencie ukladaných súborov cookies. Počet webových sídiel, ktoré obsahovali oznámenie o používaní súborov cookies v spojení s iným právnym úkonom bol 4%.
2. Po analýze súborov obsahujúcich informácie o uložených súboroch cookies sme prišli k nasledujúcim záverom. Len 2 webové sídla z celej vzorky neukladajú žiadne súbory cookies. Reklamné cookies pritom na zariadenie používateľa ukladá 70% predajcov.
3. Pri analýze korektnosti policies sme prišli na to, že 10% webových sídiel neobsahuje v daných dokumentoch žiadku zmienku o súboroch cookies. 47% všetkých eshopov poskytlo používateľom informáciu o tom, čo sú súbory cookies a na čo sa používajú. Len 13% privacy policies obsahovalo delenie súborov cookies podľa funkcionality a 27% malo v sebe obsiahnutú aj informáciu o rozsahu zbieraných údajov. O tom, komu tieto citlivé údaje ďalej predávajú informovali len 3% všetkých predajcov.
4. Len 10% nemalo na svojich stránkach zverejnený kontakt na zodpovednú osobu za spracovanie osobných údajov.
5. Informáciu o možnosti odvolania súhlasu s používaním súborov cookies poskytlo len 23% eshopov.

Záver

Keďže súbory cookies sa vo veľmi krátkom čase stali dôležitou súčasťou pri prehliadaní Internetu. Súčasne len veľmi málo prevádzkovateľov webových sídel si dá záležať na korektnom technickom spracovaní súborov cookies a dokumentov o ochrane osobných údajov. Z tohto dôvodu je potrené sa zaoberať problémami súvisiacimi s používaním súborov cookies Používaním súborov cookies webovými sídlami nás ako používateľov vystavujú riziku ohrozenia našich osobných údajov. Súbory cookies sa taktiež využívajú na rôznu škálu škodlivých aktivít. Z tohto vyplýva nutnosť vyvíjať stále efektívnejšie riešenia na ich detekciu, správnu manipuláciu a ich použitie, ako aj informovanie koncového používateľa o skutočnostiach, ktoré tieto malé kúsky dát so sebou prinášajú.

Prvým cieľom práce bolo analyzovať súbory cookies z pohľadu technickej realizácie právnych požiadaviek. V práci sme popísali fungovanie súborov cookies. To bolo dôležité pre pochopenie toho, ako sa tieto súbory správajú vo webovom rozhraní a v zariadení používateľa. Popísali sme zraniteľnosti, ktoré si so sebou tieto malé kúsky údajov nesú a taktiež ich využitie pre škodlivé aktivity, ak sa dostanú do nesprávnych rúk. Zistili sme, aké údaje tieto súbory obsahujú, čo nám pomohlo navrhnúť nástroj na ich detekciu. Porovnali sme rôzne typy riešení, ktoré sa využívajú na detekciu a analýzu cookies, ako aj dokumentov o ochrane osobných údajov. Rozhodli sme sa pre využitie detekcie na základe regulárnych výrazov.

Druhý cieľ práce bol zameraný na porovnanie aktuálnych prístupov k implementácii právnych požiadaviek súborov cookies a dokumentov o ochrane osobných údajov používaných webovými sídlami. Popísali sme už existujúce prístupy na detekciu týchto údajov. Prostredníctvom analýzy týchto prístupov sme a identifikovali niekoľko atribútov, ktoré sú pre nás najdôležitejšie a ktorým je nutné sa venovať. Na základe tohto porovnania sme sa taktiež rozhodli využiť viacero prístupov, ako tieto atribúty z webového sídla získať a spracovať, napríklad vyhľadávaním kľúčových slov pomocou regulárnych výrazov

Tretím cieľom práce bol návrh a implementácia nástroja pre detekciu súborov cookies, dokumentov o ochrane osobných údajov a spracovaní súborov cookies. Dôležitosť daného nástroja sme načrtli v úvodnej kapitole. V rámci návrhu nástroja popisujeme získavanie a následné spracovanie údajov potrebných pre našu analýzu. Popísali sme softvér Selenium WebDriver, ktorý je pre nás najvhodnejším pomocníkom

pri automatickom zbere všetkých potrebných údajov. Tento systém poskytuje veľmi rýchle a jednoduché prehľadávanie webového rozhrania. Následne popisujeme metódy, ktoré sme si zvolili pre využitie v systéme. Prvým krokom je identifikovanie všetkých atribútov, ktoré sme si pre návrh nástroja zvolili a následne ich úspešná realizácia. Rozhodli sme sa pre využitie zoznamu najlepších slovenských eshopov poskytnutých webovým sídlom najeshopy.sk, ktorý obsahuje veľké množstvo domén najobľúbenejších eshopov zoradených v abecednom poradí.. Časť tejto vzorky sme podrobili manuálnej analýze kde sme si odskúšali jednoduchosť nájdenia zvolených atribútov. Predstavili sme filtráciu pomocou kľúčových slov, ktorými vieme jednoznačne určiť, či sú všetky požiadavky pre technickú a právnu správnosť spracovania súborov cookies na danom sídle korektné. Následne sme navrhli analýzu s využitím regulárnych výrazov, ktorá nám pomôže pri hľadaní širšieho spektra kľúčových slov, potrebných pre vyslovenie záveru. Určili sme hodnoty, ktoré indikujú, či je technické spracovanie cookies a policies na danej doméne z právneho hľadiska korektné.

Napokon sme pomocou navrhnutého nástroja analyzovali celú vzorku 100 najlepších online predajcov. Prišli sme na to, že prevažná väčšina slovenských online predajcov svojich používateľov informuje o používaní súborov cookies počas prvej návštevy webového sídla. Čo sa týka dokumentov o ochrane osobných údajov a informáciách približujúcich súbory cookies, správne spracovanie poskytlo len veľmi malé percento. Možnosť odvolania súhlasu s používaním súborov cookies neposkytli takmer žiadne eshopy.

Nástroj, ktorý sme navrhli a implementovali, dokáže upozorniť používateľa na to, koľko súborov cookies sa na webovom sídle nachádza, aké typy a pre aké účely sú tieto údaje ukladané, sú tieto údaje v bezpečí, alebo sú posielané tretím stranám či na aký časový úsek sú tieto údaje uchovávané. Taktiež upozorňuje na medzery v dokumentoch a o spracovávaní osobných údajov daným online predajcom, ako aj informáciu o tom, či svoje rozhodnutie o poskytnutí súhlasu vie používateľ kedykoľvek zmeniť. To vo veľkej miere prispeje k zvýšeniu bezpečnosti používateľov na webových sídlach a zabráni tak klikaniu na tlačidlo "Súhlasím" bez hlbšieho uváženia. V budúcnosti je možné model nástroja rozširovať a zefektívniť pre detekciu ďalších atribútov, ktoré môžu priniesť nové právne úpravy.

Zoznam použitej literatúry

1. Academic Dictionaries and Encyclopedias, Ambient authority [online] Dostupné z: https://en-academic.com/dic.nsf/enwiki/4760084
2. TechTarget, Session ID [online] Dostupné z: https://www.techtarget.com/searchsoftwarequality/definition/session-ID
3. Techopedia, Domain Name System [online] Dostupné z: https://www.techopedia.com/definition/24201/domain-name-system-dns
4. Learn How HTTP Cookies Work [online] Dostupné z: https://flaviocopes.com/cookies/
5. Cookiepedia, What are Cookies? [online] Dostupné z: **Chyba! Neplatné hypertextové prepojenie.**
6. SOFFAR, H. 2017. Cookies uses , features , advantages and disadvantages [online] Dostupné z: https://www.online-sciences.com/computer/cookies-uses-features-advantages-and-disadvantages/
7. Chuan Yue; Mengjun Xie; Haining Wang (2010). An automatic HTTP cookie management system. , 54(13), 2182–2198. doi:10.1016/j.comnet.2010.03.006
8. BARTH, A. 2011. HTTP State Management Mechanism. ISSN: 2070-1721 [online] Dostupné z: https://datatracker.ietf.org/doc/html/rfc6265
9. mdn, Set-Cookie [online] https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/Set-Cookie
10. Cookies, the GDPR, and the ePrivacy Directive [online] Dostupné z: https://gdpr.eu/cookies/
11. Human Who Codes, http cookies explained [online] Dostupné z: https://humanwhocodes.com/blog/2009/05/05/http-cookies-explained/
12. JUSSILA, J. 2018. HTTP Cookie Weaknesses, Attack Methods And Defense Mechanisms: A Systematic Literature Review
13. XIAOFENG. Z. et al. Cookies Lack Integrity: Real-World Implications 24th USENIX Security Symposium August 12–14, 2015 • Washington, D.C. ISBN 978-1-939133-11-3
14. Acutenix, What is Cookie Poisoning [online] Dostupné z: https://www.acunetix.com/blog/web-security-zone/what-is-cookie-poisoning/

15. HTML.COM, Are You Using Cookies? Then This Ultimate Guide Is For You [online] Dostupné z: https://html.com/resources/cookies-ultimate-guide/#The_Risks_of_Cookies_and_What_You_Need_to_Watch_out_For
16. URBAN, T. et al. The Unwanted Sharing Economy: An Analysis of Cookie Syncing and User Transparency under GDPR. arXiv:1811.08660v1 [cs.CR] 21 Nov 2018 [online] Dostupné z: https://arxiv.org/pdf/1811.08660.pdf
17. Rózenfeldová, L., Sokol, P. New Initiatives and Approaches in the Law of Cookies in the EU. In: IDIMT-2018 Strategic Modeling in Management, Economy and Society 26th Interdisciplinary Information Management Talks. Linz: Gerhard Chroust, 2018. pp. 303-310.
18. Rózenfeldová, L. Protection of Privacy and Personal Data as regards the Use of Cookies. In: STUDIA IURIDICA Cassoviensia, vol. 8, no. 1, 2020. pp. 60-72.
19. DEGELING. M. et al. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. arXiv:1808.05096 [cs.CY] 25 Jun 2019 [online] Dostupné z: https://arxiv.org/pdf/1808.05096.pdf
20. Computer Hope, Opt [online] Dostupné z: https://www.computerhope.com/jargon/o/opt.htm
21. Rózenfeldová, L., Sokol, P. and Fortunová D. Privacy and Personal Data Protection Issues as regards the Use of Cookies on the Example of Healthcare Providers in Slovakia
22. Abbamonte, G. B. (2014) The Protection of Computer Privacy under EU Law. The Columbia Journal of European Law, Vol. 21, No. 1, pp 71-87, ISSN 1076-6715.
23. European Parliament. (2002) Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). OJ L 201, 31.7.2002, p. 37–47.
24. European Parliament. (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) Official Journal of the European Union, Vol. L119 (4 May 2016), pp. 1-88

25. WebsitePolicies, What is a Privacy Policy: The Definitive Guide [online] Dostupné z: https://www.websitepolicies.com/blog/what-is-privacy-policy
26. SANCHEZ, R. et al. (2019, July). Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (pp. 340-351).
27. CookiePro, Website Tracking: WHz and How Do Websites Track You? [online] Dostupné z: https://www.cookiepro.com/blog/website-tracking/
28. DEGELING, M. et al. (Un)informed Consent: Studying GDPR Consent Notices in the Field arXiv:1909.02638. 2019 Oct 22.
29. Cookie Script, Cookie Scanner [online] Dostupné z: https://cookie-script.com/cookie-scanner 23.03.2022
30. CookieMetrix, Determines if your website is not comply with EU Cookie Law [online] Dostupné z: https://www.cookiemetrix.com/
31. Cookieserve, Free Cookie Checker Tool for Websites [online] Dostupné z: https://www.cookieserve.com/
32. Piwik PRO, Free online Cookie Scanner [online] Dostupné z: https://piwik.pro/cookie-scanner/
33. Cookiebot, Is my website compliant? [online] Dostupné z: https://www.cookiebot.com/en/
34. Juksta, Which cookies does your site use? [online] Dostupné z: https://juksta.eu/consent-management
35. BIT SENTINEL, Free Cookies Scanner [online] Dostupné z: https://bit-sentinel.com/free-cookies-scanner/
36. EditThisCookie, Chrome rozšírenie [online] Dostupné z: https://chrome.google.com/webstore/detail/editthiscookie/fngmhnnpilhplaeedifhccceomclgfbg?hl=sk
37. Selenium, WebDriver Documentation [online] Dostupné z: chttps://www.selenium.dev/documentation/webdriver/
38. jkwakman, Open-Cookie-Database [online] Dostupné z: https://github.com/jkwakman/Open-Cookie-Database/blob/master/open-cookie-database.csv

Prílohy

Príloha A: Extrakcia údajov

Príloha B: Hľadanie atribútov

Príloha C: Výsledky

Príloha A: Extrakcia údajov

```
    folder_html = 'reports_html'
    folder_siteText = 'reports_siteText'
    folder_cookies = 'reports_cookies'
    folder_privacyText = 'reports_privacyText'
    folder_cookiesText = 'reports_cookiesText'
    urls = ["https://bittersweetparis.sk""http://budchlap.sk"]

    dictionary = {}
    # creates folder for cleaned page source code and all visible text from
page
    Path(folder_html).mkdir(parents=True, exist_ok=True)
    Path(folder_siteText).mkdir(parents=True, exist_ok=True)
    Path(folder_cookies).mkdir(parents=True, exist_ok=True)
    Path(folder_privacyText).mkdir(parents=True, exist_ok=True)
    Path(folder_cookiesText).mkdir(parents=True, exist_ok=True)

    driver      =      webdriver.Firefox(service=Service('geckodriver-v0.30.0-
win64/geckodriver.exe'))

    for url in urls:
        parsed_url = urllib.parse.urlparse(url)
        hostname = ''.join(parsed_url.netloc.split('.')[1])

        ignore_tags = ('script', 'noscript', 'style')

        driver.get(url)  # Load page

        content = driver.page_source
        #cleaner = clean.Cleaner()
        #content = cleaner.clean_html(content)

        # saves cleaned source code
        with open(folder_html + '/cleaned_source_' + hostname + '.html', 'w',
encoding='utf-8') as outfile:
            outfile.write(content)

        # saves visible content
        doc = LH.fromstring(content)
        visible_content = ""
        with  open(folder_siteText  +  '/result_'  +  hostname  +  '.txt',  'w',
encoding='utf-8') as f:
            for elt in doc.iterdescendants():
                if elt.tag in ignore_tags:
                    continue
                text = elt.text or ''
                tail = elt.tail or ''
```

```
                words = ' '.join((text, tail)).strip()

                if words:
                    f.write(words)
                    f.write('\n')
                    visible_content = visible_content + words + '\n'
    ###moje_privacy
        elems = driver.find_elements(By.TAG_NAME, 'a')
        privacy_url = ""
        cookiesText_url = ""

        for elem in elems:
                #elem.get_property("text")  ==  "Ochrana  osobných  údajov":
#pole textov, ako class hladat
            elem_text = elem.get_property("text")
            if re.search(r'\bosob.+daj.+', elem_text, re.IGNORECASE) != None:
                privacy_url = elem.get_attribute("href")

            if  re.search(r'\bcookie+', elem_text, re.IGNORECASE) != None:
                cookiesText_url = elem.get_attribute("href")

        if privacy_url != "":
            driver.get(privacy_url)  # Load privacy page
            hostname
            privacy_content = driver.page_source

            doc = LH.fromstring(privacy_content)
            visible_privacy_content = ""
            with open(folder_privacyText + '/privacy_' + hostname + '.txt',
'w', encoding='utf-8') as f:
                for elt in doc.iterdescendants():
                    if elt.tag in ignore_tags:
                        continue
                    text = elt.text or ''
                    tail = elt.tail or ''
                    words = ' '.join((text, tail)).strip()

                    if words:
                        f.write(words)
                        f.write('\n')
                        visible_privacy_content  =  visible_privacy_content  +
words + '\n'
            privacy_url = ""
        else:
            print("Couldn't find privacy policy url of " + url)

        if  cookiesText_url != "":
            driver.get(cookiesText_url)  # Load privacy page
```

```
            hostname
            cookiesText_content = driver.page_source

            doc = LH.fromstring(cookiesText_content)
            visible_cookiesText_content = ""
            with open(folder_cookiesText + '/cookiesText_' + hostname + '.txt',
'w', encoding='utf-8') as f:
                for elt in doc.iterdescendants():
                    if elt.tag in ignore_tags:
                        continue
                    text = elt.text or ''
                    tail = elt.tail or ''
                    words = ' '.join((text, tail)).strip()

                    if words:
                        f.write(words)
                        f.write('\n')
                        visible_cookiesText_content                          =
visible_cookiesText_content + words + '\n'
            privacy_url = ""
        else:
            print("Couldn't find cookie policy url of " + url)
    ###moje_privacy

    ###moje

        cookies = driver.get_cookies()
        #print(cookies)

        obj_page = {}
        obj_page['hostname'] = hostname
        obj_page['visibleText'] = visible_content
        obj_page['cookies'] = cookies

        with open(folder_cookies + '/cookies_' + hostname + '.json', 'w',
encoding='utf-8') as outfile:
            json.dump(obj_page,     outfile,     indent=4,     sort_keys=True,
ensure_ascii=False)

driver.close()
```

Príloha B: Hľadanie atribútov

```
    # Vytvorenie dataframe df_cookies
    #df_cookies = pd.DataFrame(lst)
    df_cookies                                                                    =
pd.DataFrame(columns=['Web_name','Suhlas','Suhlas_alter','Spojenie_PU','Cooki
es_fun','Cookies_anal','Cookies_rekl','Cookies_unkn','Cookies_session','Cooki
es_persis','Cookies_exists','Co_su_cookies','Typy_cookies','Rozsah_cookies','
Prijemca','Kontakt_spravca','Odovolanie_suhlasu'])
    print(df_cookies)



    ###################################
    #
    # HTML file analyser
    #
    ###################################



    # Prejde vsetky subory a prida ich do pola files
    files_HTML = []
    path_HTML= "testing/reports_nonCleanedHtml/"
    print("Zoznam HTML suborov: ")
    for (dirpath, dirnames, filenames) in walk(path_HTML):
        files_HTML.extend(filenames)
        print(files_HTML)

    index = 0 # index sa nastavi na 0
    #  Kontrola vsetkych suborov a postupne vykonanie testov
    for file in files_HTML:
        print("==============")
        print(file)
        print("==============")
        lines = open(path_HTML+file, encoding="utf8").readlines()



     #Test "Súhlasím"
        suhlasim_bool = False
        print ("Možnosť udelenia súhlasu:")
        for line in lines:
            if      re.search(r"Súhlasím.*|Rozumiem.*|P.+všetk.+",      line,
re.IGNORECASE):
                #print(line, end="")
                suhlasim_bool = True
        if (suhlasim_bool == True):
            print("Je uvedené Súhlasím: ÁNO")
            df_cookies.loc[index,"Suhlas"] = True
        else:
```

```
            print ("Je uvedené Súhlasím: NIE")
            df_cookies.loc[index,"Suhlas"] = False

        print("-------------")

    #Test "Nechcem cookies"
        suhlasim_bool = False
        print ("Možnosť výberu alternatívy súhlasu: ")
        for line in lines:
            if  re.search(r"Odmietnuť.*|Zakázať.*|Disagree|Nechcem  cookies",
line, re.IGNORECASE):
                #print(line, end="")
                suhlasim_bool = True
            elif re.search(r"Nastavenie", line, re.IGNORECASE):
                #print(line, end="")
                nastavenie_bool = True
        if (suhlasim_bool == True) or (nastavenie_bool == True):
            print("Možnosť výberu alternatívy: ÁNO")
            df_cookies.loc[index,"Suhlas_alter"] = True
        else:
            print ("Možnosť výberu alternatívy: NIE")
            df_cookies.loc[index,"Suhlas_alter"] = False

        print("-------------")

    #Test "Pokračovaním v prehliadaní"
        suhlasim_bool = False
        print ("Aktívne udelenie súhlasu:")
        for line in lines:
            if   re.search(r"Pokračovaním   v   prehliadaní|Pokračovaním   v
prezeraní|Prehliadaním.+súhlasíte|Prezeraním.+súhlasíte",               line,
re.IGNORECASE):
                #print(line, end="")
                suhlasim_bool = True
        if (suhlasim_bool == True):
            print("Je uvedené Pokračovaním v prehliadaní. Súhlas je aktívny:
NIE")
            print("-------------")
            print("Spojenie s iným právnym úkonom:")
            print("Súhlas je spojený s iným právnym úkonom: ÁNO")
            df_cookies.loc[index,"Spojenie_PU"] = True
        else:
            print("Nie  je  uvedené  Pokračovaním  v  prehliadaní.  Súhlas  je
aktívny: ÁNO")
            print("-------------")
            print("Spojenie s iným právnym úkonom:")
            print("Súhlas je spojený s iným právnym úkonom: NIE")
            df_cookies.loc[index,"Spojenie_PU"] = False
```

```
    index=index+1

print("*********************************")


##################################
#
# Cookies file analyser
#
##################################


files_cookies = []
path_cookies= "testing/reports_cookies/"
for (dirpath, dirnames, filenames) in walk(path_cookies):
    files_cookies.extend(filenames)
    #print(files_cookies)

index = 0
for file in files_cookies:
    print("==============")
    print(file)
    print("==============")
    filepath = path_cookies+file
    lines = open(filepath, encoding="utf8").readlines()
    saveTime = os.path.getmtime(filepath)
    print(str(saveTime))
    name = (Path(filepath).stem).replace('cookies_', '')
    #print(name)

    #Nazov weboveho sidla
    for line in lines:
        if re.search(r"\bhostname.+,$", line, re.IGNORECASE):
           hostName = line.replace('\"hostname\": "', '')
           hostName = hostName[:-3]
           hostName = hostName.replace('      ', '')
    df_cookies.loc[index,"Web_name"] = hostName

 #Test "Len first-party cookies"
    #firstParty_bool = False
    counter_first = 0
    c = 0
    result = 0
    print ("Len first-party cookies:")
    for line in lines:
        if re.search(r"\bdomain.+,$", line, re.IGNORECASE):
            #print(line, end="")
```

```
                #firstParty_bool = True
                c = c +1
            if re.search(r"\bdomain.+:.+%s.+,$" % name, line, re.IGNORECASE):
                #print(line, end="")
                #firstParty_bool = True
                counter_first = counter_first +1
        result = c - counter_first
        if (result  != 0):
            print("NIE")
        else:
            print("ÁNO")

        print("-------------")
        print("Celkový počet cookies je: " + str(c))

    #Test "Delenie podľa funkcionality"
        print('Delenie podľa funkcionality')
        cookies_list=[]
        for line in lines:
            if re.search(r"\bname.+,$", line, re.IGNORECASE):
               cookieName = line.replace('\"name\": "', '')
               cookieName = cookieName[:-3]
               cookieName = cookieName.replace('      ', '')
               cookies_list.append(cookieName)

        #print(cookies_list)

        with open('testing/cookies_by_functionality.txt', 'r') as file:
            #database = file.read().replace('\n', '')
            database = file.read()

        functional = 0
        analytics = 0
        marketing = 0
        unknown = 0
        #print(database)
        for cookies in cookies_list:
            #print(cookies)
            #print(r"\Functional,%s" %cookies)
            if re.search(r"Functional,%s" %cookies, database, re.IGNORECASE):
                functional = functional +1
            if re.search(r"Analytics,%s" %cookies, database, re.IGNORECASE):
                analytics = analytics +1
            if re.search(r"Marketing,%s" %cookies, database, re.IGNORECASE):
                marketing = marketing +1
            if re.search(r",%s" %cookies, database, re.IGNORECASE) == None:
                unknown = unknown +1
```

```
        df_cookies.loc[index,"Cookies_fun"] = functional
        df_cookies.loc[index,"Cookies_anal"] = analytics
        df_cookies.loc[index,"Cookies_rekl"] = marketing
        df_cookies.loc[index,"Cookies_unkn"] = unknown
        print("Funkčné cookies: " + str(functional))
        print("Analytické cookies: " + str(analytics))
        print("Reklamné cookies: " + str(marketing))
        print("Neznáme cookies: " + str(unknown))

        print("-------------")

     #Test "Delenie podľa doby uschovania" #nejde
        #firstParty_bool = False

        print ("Delenie podľa doby uschovania:")
        session = 0
        persistent = 0
        cookieTimeStamp_list =[]
        for line in lines:
            if re.search(r"\bexpiry.+,$", line, re.IGNORECASE):
                cookieTimeStamp = line.replace('\"expiry\": ', '')
                cookieTimeStamp = cookieTimeStamp[:-2]
                cookieTimeStamp = cookieTimeStamp.replace('      ', '')
                convert_time                                                    =
datetime.utcfromtimestamp(int(cookieTimeStamp)).strftime('%Y-%m-%d %H:%M:%S')
                cookieTimeStamp_list.append(convert_time)
        print(cookieTimeStamp_list)
        print ("Session cookies: ",c-len(cookieTimeStamp_list))
        print ("Persistent cookies: ",len(cookieTimeStamp_list))
        df_cookies.loc[index,"Cookies_session"] = c-len(cookieTimeStamp_list)
        df_cookies.loc[index,"Cookies_persis"] = len(cookieTimeStamp_list)

        index = index + 1

    print("**********************************")


    ###################################
    #
    # Privacy policy file analyser
    #
    ###################################

    files_privacyPolicy = []
    path_privacyPolicy= "testing/reports_privacyText/"
    for (dirpath, dirnames, filenames) in walk(path_privacyPolicy):
        files_privacyPolicy.extend(filenames)
```

```
        #print(files_privacyPolicy)

    index = 0
    for file in files_privacyPolicy:
        print("==============")
        print(file)
        print("==============")
        filepath = path_privacyPolicy+file
        lines = open(filepath, encoding="utf8").readlines()


    #Test "cookies"
        cookies_bool = False
        print ("Existuje zmienka o cookies:")
        for line in lines:
            if re.search(r"cookie.+", line, re.IGNORECASE):
                #print(line, end="")
                cookies_bool = True
        if (cookies_bool == True):
            print("Zmienka o cookies: ÁNO")
            df_cookies.loc[index,"Cookies_exists"] = True
        else:
            print ("Zmienka o cookies: NIE")
            df_cookies.loc[index,"Cookies_exists"] = False

        print("-------------")


     #Test "Čo sú cookies"
        coSuCookies_bool = False
        print ("Čo sú cookies:")
        for line in lines:
            if re.search(r"cookie.+mal.+text.+súbor.*|čo sú.+cookie.+", line,
re.IGNORECASE):
                #print(line, end="")
                coSuCookies_bool = True
        if (coSuCookies_bool == True):
            print("Je uvedené Čo sú cookies: ÁNO")
            df_cookies.loc[index,"Co_su_cookies"] = True
        else:
            print ("Je uvedené Čo sú cookies: NIE")
            df_cookies.loc[index,"Co_su_cookies"] = False

        print("-------------")


     #Test "Typy cookies"
        typy_bool = False
```

```
        print ("Typy cookies:")
        for line in lines:
            if
re.search(r"analytic.*cookie|.*marketingové.*cookie|.*nutné.*cookie|funkčné.*
cookie|esenciálne.*cookie|konverzné.*cookie|trackingové.*cookie|preferenčné.*
cookie", line, re.IGNORECASE):
                #print(line, end="")
                typy_bool = True
        if (typy_bool == True):
            print("Zmienka o typoch cookies: ÁNO")
            df_cookies.loc[index,"Typy_cookies"] = True
        else:
            print ("Zmienka o typoch cookies: NIE")
            df_cookies.loc[index,"Typy_cookies"] = False

        print("-------------")



    #Test "Rozsah údajov"
        rozsah_bool = False
        print ("Rozsah údajov:")
        for line in lines:
            if  re.search(r"zhromažďov.+údaj.*|zhromažďov.+informác.+",  line,
re.IGNORECASE):
                #print(line, end="")
                rozsah_bool = True
        if (rozsah_bool == True):
            print("Zmienka o rozsahu zbieraných údajov: ÁNO")
            df_cookies.loc[index,"Rozsah_cookies"] = True
        else:
            print ("Zmienka o rozsahu zbieraných údajov: NIE")
            df_cookies.loc[index,"Rozsah_cookies"] = False
        print("-------------")


    #Test "Príjemca"
        prijemca_bool = False
        print ("Príjemca:")
        for line in lines:
            if  re.search(r"Google|Facebook|Twitter|tret.+kraj.+|odosielan.+",
line, re.IGNORECASE):
                #print(line, end="")
                prijemca_bool = True
        if (prijemca_bool == True):
            print("Zmienka o príjemcoch: ÁNO")
            df_cookies.loc[index,"Prijemca"] = True
        else:
```

```
            print ("Zmienka o príjemcoch: NIE")
            df_cookies.loc[index,"Prijemca"] = False

        print("-------------")


     #Test "Správca"
        spravca_bool = False
        print ("Správca:")
        for line in lines:
            if
re.search(r"Správc|prevádzkovateľ.*|poveren.+osob.+|zodpoved.+osob.+",  line,
re.IGNORECASE):
                #print(line, end="")
                spravca_bool = True
        if (spravca_bool == True):
            print("Kontakt na správcu: ÁNO")
            df_cookies.loc[index,"Kontakt_spravca"] = True
        else:
            print ("Kontakt na správcu: NIE")
            df_cookies.loc[index,"Kontakt_spravca"] = False

        print("-------------")


     #Test "Odvolanie súhlasu"
        odvolanie_bool = False
        print ("Odvolanie súhlasu:")
        for line in lines:
            if                          re.search(r"Odmietnu.+cookie.*|zakázať
cookie.*|Odvola.+súhlas.+cookie", line, re.IGNORECASE):
                #print(line, end="")
                odvolanie_bool = True
        if (odvolanie_bool == True):
            print("Možnosť odvolania súhlasu: ÁNO")
            df_cookies.loc[index,"Odovolanie_suhlasu"] = True
        else:
            print ("Možnosť odvolania súhlasu: NIE")
            df_cookies.loc[index,"Odovolanie_suhlasu"] = False

        index = index + 1
        print("-------------")

    print("**********************************")

    # DF print and save to CSV file
    print(df_cookies)
    df_cookies.to_csv('cookies_ouput.csv')
```

Príloha C: Výsledky

Web_name	Suhlas	Suhlas_alter	Spojenie_PU	Cookies_fun	Cookies_anal	Cookies_rekl	Cookies_unkn	Cookies_session
123led	T	T	F	1	0	0	5	2
4home	T	T	F	1	2	2	7	0
adiel	T	T	F	1	0	0	7	1
afg	T	T	F	0	0	1	2	1
agatinsvet	T	T	F	0	0	1	2	1
ajprodukty	T	T	F	3	0	0	6	2
alensa	F	F	F	2	2	3	11	2
alza	T	T	F	0	0	0	18	1
amiatex	F	F	F	3	3	2	6	2
andreashop	F	F	F	1	0	0	1	2
anrdoezrs	T	T	F	1	0	0	2	1
anwell	T	T	F	1	0	0	2	1
ara-shoes	T	T	F	1	0	1	10	5
artforum	T	T	F	1	2	2	1	1
astoreo	T	T	F	2	2	3	15	4
audiolibrix	T	T	F	1	0	0	2	0
autoobuv	T	T	F	0	0	0	1	0
autovybava	T	T	F	5	0	0	13	3
babickinazahrada	T	T	F	7	2	4	8	3
bedario	T	T	F	4	3	3	14	1
belenka	T	T	F	1	0	0	1	1
benulekaren	T	T	F	0	0	1	6	3
bionatural	T	T	F	2	2	2	5	1
biotechusa	T	T	F	3	6	9	8	2
blancheporte	T	T	F	2	2	3	12	1
boel	T	T	F	6	2	2	10	3
bohatstvo-prirody	T	T	F	2	2	2	11	3
bomo	T	T	F	1	0	0	5	2
brainmarket	T	T	F	2	2	3	8	3
brloh	T	T	T	1	2	3	8	0
bronx	T	T	F	2	3	2	3	0
bubulakovo	T	T	F	9	4	3	12	2
cisteoblecenie	T	T	F	1	0	0	2	2
cz	T	T	F	1	0	0	2	1
daffer	T	T	F	1	2	3	4	0
danimani	T	T	F	1	2	2	3	1

darcekovy-raj	T	T	F	1	2	3	7	2
decathlon	T	T	T	6	2	4	23	11
dekortextil	T	T	F	3	3	2	2	1
delimano	T	T	T	3	3	1	21	7
dermacol	T	T	F	3	2	1	9	2
dobra-miska	T	T	F	2	3	2	6	1
dobrytextil	T	T	F	4	3	3	23	5
dormeo	T	T	T	3	3	1	21	7
drogeriashop	T	T	F	1	0	1	10	3
efarby	T	T	F	5	2	3	9	2
efitness	T	T	F	1	4	2	12	3
emobilshop	T	T	F	2	2	3	7	2
e-pneumatiky	T	T	F	2	2	2	6	1
esat	T	T	F	1	0	0	4	2
e-spotrebice	T	T	F	1	2	1	3	1
estilofina	T	T	F	8	2	3	11	1
etanikozmetika	T	T	F	1	0	0	5	1
evolutiongroup	T	T	F	2	0	0	1	2
eyerim	T	T	F	2	2	3	8	2
fashionformen	T	T	F	2	2	3	5	1
fexi	T	T	F	1	0	1	8	2
gamisport	T	T	F	3	3	1	7	2
gams-shop	T	T	F	1	0	0	1	1
gangstagroup	T	T	F	6	2	3	18	5
gatio	T	T	F	2	2	2	11	3
gigalash	T	T	F	1	1	2	3	0
givsport	T	T	F	2	2	2	10	4
glash	T	T	F	5	2	4	8	1
grizly	T	T	F	3	0	0	7	4
gsklub	T	T	F	0	0	0	0	0
herbatica	T	T	F	2	2	3	8	3
hillvital	T	T	F	3	2	2	8	4
homepoint	T	T	F	6	3	1	6	0
houseland	T	T	F	2	2	3	11	3
idea-nabytok	T	T	F	0	0	0	0	0
indickynabytok	T	T	F	3	3	2	5	1
insportline	T	T	F	2	2	2	12	3
invia	T	T	F	1	0	0	2	1

i-zahradnynabytok	T	T	F	1	0	0	3	1
jipro	F	F	F	2	2	1	1	0
johnc	T	T	F	2	2	3	10	2
johngarfield	T	T	F	2	2	1	8	3
kidtown	T	T	F	2	2	3	4	2
kinekus	T	T	F	1	0	0	2	1
kokiskashop	T	T	F	2	2	3	7	1
kompava	T	T	F	1	0	0	3	1
krasnevone	T	T	F	2	2	3	6	3
lampyasvetla	T	T	F	13	0	0	3	1
lucystyle	T	T	F	3	3	2	9	2
macaciamama	T	T	F	1	0	0	9	2
magnet-3pagen	T	T	F	5	2	3	13	2
manstyle	T	T	F	1	0	2	6	2
manutea	T	T	F	2	2	2	1	0
market24	T	T	F	0	0	0	2	0
nutraceutics	T	T	F	4	2	3	5	0
shop	T	T	F	1	0	1	2	1

Web_name	Cookies_persis	Cookies_exists	Co_su_cookies	Typy_cookies	Rozsah_cookies	Prijemca	Kontakt_spravca	Odovolanie_suhlasu
123led	4	T	T	F	F	T	T	T
4home	10	T	F	F	F	T	T	F
adiel	7	T	T	T	T	T	T	T
afg	2	T	F	F	F	T	T	F
agatinsvet	2	T	T	T	F	T	T	F
ajprodukty	7	T	T	T	F	T	F	F
alensa	14	T	F	T	F	T	T	F
alza	17	T	T	F	F	T	T	F
amiatex	9	T	F	F	F	F	T	F
andreashop	0	T	F	T	F	T	F	F
anrdoezrs	2	T	T	T	F	T	T	F
anwell	2	T	F	F	F	T	T	F
ara-shoes	7	T	T	T	F	T	T	F
artforum	2	T	T	F	T	T	T	F
astoreo	16	T	T	F	F	T	T	T
audiolibrix	3	T	F	T	F	T	T	F
autoobuv	1	F	F	F	F	F	F	F
autovybava	15	F	F	F	F	T	T	F

babickinazahrada	16	F	F	F	F	T	T	F
bedario	20	T	T	T	T	T	T	T
belenka	1	T	F	F	F	T	T	F
benulekaren	4	F	F	F	F	T	T	F
bionatural	8	T	F	T	F	T	T	F
biotechusa	19	T	T	T	T	T	T	T
blancheporte	16	T	T	T	T	T	T	F
boel	15	T	T	F	T	T	T	T
bohatstvo-prirody	12	F	F	F	F	T	T	F
bomo	4	T	F	F	F	T	T	F
brainmarket	10	F	F	F	F	T	T	F
brloh	12	F	F	F	F	T	T	F
bronx	7	T	F	F	F	T	T	F
bubulakovo	24	T	T	T	F	T	T	F
cisteoblecenie	1	T	F	T	F	T	F	F
cz	2	T	F	F	F	T	T	F
daffer	8	T	T	T	F	T	T	T
danimani	5	T	F	F	F	T	T	F
darcekovy-raj	9	T	F	T	F	T	T	F
decathlon	22	T	F	F	F	T	F	F
dekortextil	6	T	F	T	F	T	T	F
delimano	18	T	T	T	T	T	T	F
dermacol	11	T	T	T	F	T	T	F
dobra-miska	9	F	F	F	F	T	T	F
dobrytextil	25	T	T	T	F	T	T	F
dormeo	18	T	T	T	T	T	T	F
drogeriashop	9	T	F	F	F	T	T	F
efarby	15	T	F	F	F	T	T	F
efitness	16	T	T	T	F	T	T	F
emobilshop	10	T	F	F	F	T	T	T
e-pneumatiky	9	T	F	F	F	T	T	F
esat	3	T	F	F	T	T	T	F
e-spotrebice	4	T	F	T	F	T	T	F
estilofina	21	T	T	T	F	T	T	F
etanikozmetika	5	T	T	T	T	T	T	T

evolutiongroup	1	T	T	T	F	T	T	T
eyerim	11	T	F	T	T	T	T	T
fashionformen	9	T	T	T	F	T	T	F
fexi	8	T	T	T	F	T	T	F
gamisport	9	T	F	T	F	T	T	F
gams-shop	1	T	T	T	F	T	T	F
gangstagroup	22	T	T	T	F	T	T	T
gatio	12	T	F	F	F	T	T	F
gigalash	5	T	T	T	T	T	T	T
givsport	10	T	F	F	F	T	T	F
glash	16	T	F	T	F	T	T	F
grizly	6	T	F	T	F	T	F	F
gsklub	0	T	T	T	T	T	T	T
herbatica	10	T	F	F	F	T	T	F
hillvital	9	T	T	F	F	T	T	F
homepoint	13	T	T	T	T	T	T	F
houseland	13	T	F	T	F	T	T	F
idea-nabytok	0	T	F	T	T	T	T	F
indickynabytok	9	T	T	T	F	T	T	F
insportline	13	T	T	T	T	F	T	F
invia	2	T	T	T	F	T	T	F
i-zahradnynabytok	3	T	F	T	F	T	T	F
jipro	4	T	F	T	T	T	T	F
johnc	13	T	T	T	T	T	T	F
johngarfield	8	T	F	F	F	T	F	T
kidtown	7	F	F	F	F	T	T	F
kinekus	2	T	F	T	F	T	T	F
kokiskashop	11	T	T	T	F	T	T	T
kompava	3	T	F	T	T	T	T	T
krasnevone	8	T	F	F	F	T	T	F
lampyasvetla	15	T	T	T	F	T	T	F
lucystyle	12	T	F	F	F	T	F	F
macaciamama	8	T	T	T	F	T	T	F
magnet-3pagen	19	T	T	T	T	T	T	F
manstyle	7	T	T	T	F	T	T	T
manutea	5	T	F	T	T	T	T	F
market24	2	T	F	T	F	T	T	F

nutraceuti cs	12	T	T	T	F	T	T	F
shop	3	T	T	T	F	T	T	T