Univerzita Karlova

Pedagogická fakulta

Katedra informačních technologií a technické výchovy (41-KITTV)

BAKALÁŘSKÁ PRÁCE

Rozsah a dopady sběru osobních dat v digitálním prostředí s přesahem do vzdělávání

The scope and impact of the collection of personal data in the digital environment, extending to education

Jan Peterka

Vedoucí bakalářské práce:	PhDr. Josef Procházka, Ph.D.
Studijní program:	Specializace v pedagogice (B7507)
Studijní obor:	Informační technologie se zaměřením na vzdělávání

Praha 2021

Odevzdáním této bakalářské práce na téma Rozsah a dopady sběru osobních dat v digitálním prostředí s přesahem do vzdělávání potvrzuji, že jsem ji vypracoval pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

Praha, 17.4. 2021 Podpis autora

Děkuji svému vedoucímu práce doktoru Procházkovi za to, že jsem u něj mohl práci psát, za jeho pomoc s definováním její struktury, a jeho cennou zpětnou vazbu, která pomohla kvalitě práce.

Děkuji Emilovi Millerovi za poskytnutí TeX šablony a pomoc s formátováním práce.

Abstrakt

Cílem bakalářské práce je zmapovat téma vytváření digitální stopy a osobních dat v digitálním prostředí, a ukázat tuto problematiku zejména z pohledu jednotlivce, možného vlivu na něj a možností jeho vlivu na vytváření vlastní digitální stopy.

Dále se práce zabývá napojením tohoto tématu do stávajících i vznikajících rámcových vzdělávacích plánů a obecně do aktuálního školského vzdělávání.

Ze zjištěných poznatků pak vychází praktická část práce, která si dává za cíl vytvořit prototyp online prostředí pro seznamování lidí (a zejména žáků, studentů) s tímto tématem popularizační a interaktivní formou.

Klíčová slova

Sběr digitálních dat, Digitální stopa, Bezpečnost

Abstract

The aim of the bachelor thesis is to map the topic of creating a digital footprint and personal data in a digital environment, and to show this issue especially from the perspective of the individual, the possible influence on him and the possibility of his influence on creating his own digital footprint.

Furthermore, the work deals with the connection of this topic to existing and emerging educational plans and in general to the current school education.

The practical part of the work, based on these findings, aims to create a prototype online environment for acquainting people (and especially pupils, students) with this topic in a popularizing and interactive form.

Keywords

Collection of personal data, digital environment, security

Obsah

Úvod

Cambridge Analytica, Systém sociálního kreditu v Číně nebo předvolání ředitelů firem jako je Facebook a Google před americký Kongres. To je několik mediálně známých případů, které v posledních letech vynáší téma osobních dat, jejich směru a s tím spojených rizik, do povědomí širší veřejnosti. Jsou však uživatelé (tedy všichni, kdo používají digitální služby) o tématu dostatečně informováni? Jsou tyto informace takové, aby se lidé zvládali v tématu orientovat, a zároveň dělat zodpovědná, informovaná rozhodnutí o svém vlastním chování v digitálním světě?

Tato práce si dává za cíl téma osobních údajů a digitální stopy zmapovat, a to primárně z pohledu jednotlivce – uživatele. Pokouší se shrnout základní témata a pojmy, které pomáhají se v tématu zorientovat a mluvit o něm, prozkoumat typy dat a digitální stopy, která naším každodenním chováním vzniká, upozornit na rizika, a zároveň nabídnout konkrétní přístupy, kterými uživatel může tuto realitu ovlivňovat.

Tyto teoretické poznatky jsou pak kromě jejich samotného přínosu i podkladem pro praktickou část práce – vytvoření prostředí, ve kterém je uživatel tímto tématem provázen.

Kapitola 1

Problematika ochrany digitálních osobních dat uživatelů

Úvod

Rozvoj digitální části života s sebou mimo jiné přinesl čím dál větší objem digitálních (a tedy relativně snadno uchovávatelných, duplikovatelných a strojově či částečně strojově zpracovatelných) dat. V online světě tráví lidé čím dál větší množství času – v roce 2020 to bylo v průměru téměř 7 hodin denně (Kemp, 2019) – a spolu s tím jsou čím dál pokročilejší technologie pro sběr a vyhodnocování dat.

V této kapitole budou popsány důvody, které vedou k aktuálnímu stavu v oblasti sběru osobních dat, jak z pohledu ekonomické a jiné motivace, tak s pohledu technologického. Dále rozebereme rizika tohoto stavu.

Použité pojmy a koncepty

Networked privacy
Networked privacy, aj kho definují Marwick a Boyd, je model soukromí, který oproti klasickému pohledu individuálnímu přidává rozměr propojenosti – zohledňuje tedy fakt, že lidé sdílejí (ať vědomě, nebo nevědomě) i informace o ostatních, a tedy soukromí jednotlivce není oddělitelné od soukromí dalších osob (Marwick – danah boyd, 2014).
Do češtiny by šlo přeložit jako *propojené soukromí*, v práci je použit originální název.

1.1 Motivace pro sběr dat

V diskuzích o osobních datech a jejich sběru se nejčastěji objevují jména dvou digitálních firem – Google a Facebook. Důvodem je samozřejmě to, že jsou obě velkou součástí nabídky digitálních služeb,

a zejména to, že jejich hlavní oblastí podnikání je zprostředkování reklamy.

Facebook se blíží ke třem miliardám aktivních uživatelů (Tankovska, 2020), u Google je výpočet složitější (protože jde o různé služby), ale jen mezi vyhledávači (tedy esenciální služba pro většinu uživatelů internetu) má podíl přes 90 % (Statcounter, 2021). S celkovým odhadem počtu 4,6 miliardy digitálních uživatelů (Kemp, 2019) tedy jde o více než 4 miliardy uživatelů.

Obě tyto firmy mají svůj obchodní model založený na nabízení reklamy – v případě obou firem jde o primární zdroj příjmů. Jejich dominance na trhu (v USA má Google 30 % trhu reklam (eMarketer Editors, 2019) v digitálním prostředí, Facebook následuje s 20 %) je umožněna dobrým zacílením reklamy. Tato personalizace je umožněna právě sběrem dat napříč digitálním prostředím a jejich analýzou.

1.2 Technické možnosti sběru dat

Pro tvorbu profilů jednotlivých uživatelů webu a nabízení reklamy je potřeba sbírat co nejvíce informací o chování uživatele v online světě. Vytvoření profilu člověka nebylo technicky téměř možné před začátkem používání *cookies*.
Cookies byly poprvé implementovány v roce 1994 do prohlížeče Netscape jako řešení pro ukládání stavu nákupního košíku (Penland, 2020), a umožňují uchovávat informace o uživateli na dané stránce. S rozvojem webu a zejména vkládání skriptů a `iframů` na webové stránky to umožnilo sledovat uživatele skrze cookies napříč weby, včetně sledování historie prohlížení části webů. To dále umožnilo jednu z klíčových marketingových metod – remarketing, tedy nabízení produktů z obchodu, který uživatel na webu navštívil (Scott, 2020). Cookies se staly na dlouho dobu hlavním způsobeb profilování uživatelů pro reklamní účely. Postupně jsou možnosti cookies právně limitovány, a začínají se používat další techniky – *device* a *browser fingerprinting*, *tracking pixels* nebo *behavioral profiling*.

Jednotlivé technologie jsou více popsány v kapitole Hlavní zdroje dat digitální stopy uživatele.

1.3 Možnosti analýzy dat

Se zvyšujícím se množstvím veřejně dostupných osobních dat je snazší vytvářet statistické modely či samoučící systémy, které na základě dat dokáží identifikovat osobu, odhadovat její pohlaví, věk, náboženské a politické přesvědčení, charakterové rysy nebo sexualitu, jak ukázalo mnoho výzkumů (Azucar et al., 2018). Na to prakticky navázala například firma Cambridge Analytica, využívající tyto modely pro cílenou politickou reklamu. Je očividné, že to jsou často velmi citlivá data, která mohou být potenciálně zneužita mnoha způsoby a uživatel si nemusí být vůbec vědom, že si o něm takovouto představu někdo dělá a používá ji pro (například) nastavení zobrazovaného obsahu a reklam.

1.4 Povědomí o problematice

Důležitou otázkou je, jaké je aktuálně povědomí lidí o této realitě. To není snadná otázka, i z toho důvodu, že se oblast rychle mění a vyvíjí. Ovšem nějakou představu je možné si udělat.

Dle PEW Research je 59 % dotázaných občanů USA podle svého mínění nedostatečně informováno o tom, jak s jejich daty společnosti nakládají. Zároveň má 81 % z nich pocit nedostatečných možností kontroly nad tím, jak jsou data o nich sbírána, a shodné procento vnímá rizika spojená se sběrem osobních dat vyšší než benefity (Research, 2019).

Studie zkoumající povědomí a změny ve vnímání digitálního soukromí mezi studenty po medializované kauze firmy Cambridge Analytica, dochází k závěrům, že (Hinds et al., 2020):

- lidé málo vnímají dopady související s *networked privacy* – tedy uvědomění možností získávání a agregace dat o osobě z dat jiných osob
- lidé si myslí, že jsou imunní vůči cílené reklamě, a tedy nemají obavu z toho, že budou data o nich využita k ovlivnění jich samotných
- z toho plyne závěr, že lidé nemohou dělat opravdu uvědomělé závěry, pokud mají takto nepřesnou představu

Dalším důležitým faktorem ve vnímání problematiky je i *privacy fatigue (do češtiny lze zhruba přeložit jako "únava z tématu soukromí*), jak popisuje Choi (Choi et al., 2018). Ta plyne z pocitu ztráty kontroly, nepřehlednosti celého tématu a pocitu marnosti z dalších a dalších úniků dat a prolomení soukromí. Studie ukazuje, že tato únava vede k menšímu zájmu o téma (*behavioral disengagement*) a menší ochotě až úplné rezignaci na změnu vlastního chování.

Studie dále ukazuje, že uživatelé v oblasti soukromí a sběru dat nemají od používaných služeb definovaná očekávání (Pilton et al., 2021).

Shrnutí

Tato kapitola přinesla základní přehled o problematice osobních dat a pojmech s ní spojených.
Jedním z důležitých poznání je fakt, že pro mnoho lidí je tato problematika složitá (pro některé až tak, že rezignují na pokusy se v ní orientovat), nemají pocit informovanosti, a zároveň si nejsou vědomi všech aspektů spojených se soukromí a jejího narušení v souvislosti se sdílením osobních dat. To vede k přesvědčení, že má smysl vytvořit nástroje pro větší informovanost a pochopení problematiky.

Kapitola 2

Rizikové dopady sběru osobních dat

Úvod

Tato kapitola se pokouší zodpovědět na otázku, proč je téma osobních dat pro jednotlivce důležité, tedy „Proč by mě to mělo zajímat?“. Poukazuje na obecná i konkrétní rizika a problémy, které se sběrem osobních dat v dříve neuskutečnitelném měřítku souvisí.

2.1 Změna chování na základě vědomí vytváření stopy

Jedním z diskutovaných témat je riziko změny chování na základě vědomí toho, že naše chování je sledováno, ukládáno a potenciálně analyzováno (Bryce – Klang, 2009).

Tato oblast ještě není příliš prozkoumána, je tedy těžké říct, jak velké změny chování mohou nastávat. Tato sekce nabízí úvahy o těchto dopadech.

Jeden ze směru úvah o změnách chování vychází z modelu *Panopticonu* (Foucault, 1980). Panopticon je popsán jako vězení, kde může být každý vězeň kdykoli sledován strážným. Protože vězeň neví, jestli zrovna je nebo není sledován, chová se, jako by byl sledován stále. Je tedy vytvořeno prostředí, kde se člověk stále cítí pod dohledem, a reguluje na základě toho svoje chování.

Tento teoretický model se v současnosti stává realitou skrze Systém sociálního kreditu – systém, který každému čínskému občanovi určuje *sociální skóre* na základě chování. Toto je umožněno enormním propojením mnoha zdrojů dat soukromých společností i státu, a právě cílenou snahou o co největší možnosti monitorování občanů. Dokonce můžeme říct, že tento model je zesílenou verzí Foucaultova Panopticonu – na rozdíl od něj tu má opravdu docházet ke stálému sledování a zpracování těchto dat (například pomocí systémů umělé inteligence).

Dalším zajímavým aspektem, vyplývajícím z aktuálního množství dat a *networked privacy*, je fakt, že i cílené snahy o nevytváření digitální stopy (například vypnutím telefonu nebo používáním anonymizačních nástrojů) může být samo o sobě také součástí stopy – například používání prohlížeče zaměřeného na soukromí zároveň zvyšuje rozpoznatelnost uživatele skrze *browser fingerprinting*.

Zároveň se zdá, že (alespoň) mladí lidé kontrolu nad jimi vytvářeným digitálním obsahem více řeší a přizpůsobují tomu své chování (Livingstone, 2008).

2.2 Možnosti zneužití dat

Dalším z problémů týkajících se osobních dat a jejich sběru je možnost jejich zneužití. Pokusíme se na téma nahlédout ze dvou pohledů – možného zneužití dat zaměřeného na jednotlivce, a rizika zneužití s dopadem na společnost. Je však třeba mít na vědomí, že už jen z pohledu *networked privacy* toto rozdělení nemůže mít ostrou hranici, a je pouze orientační.
Pojem zneužití je v práci používán ve smyslu takovéého využití, jež lze vnímat jako negativní pro daného jedince či skupinu.

2.2.1 S dopadem na jednotlivce

Přemýšlení o osobních datech primárně vede k uvažování, jestli a jak mohou být zneužita proti mé osobě. Zároveň je častý i pocit, že „nemám co skrývat“, a tedy není třeba svou digitální stopu hlídat.

Tato sekce ukáže několik obecných rizik i konkrétních případů zneužití dat, které mohou tento pocit změnit.

Vloupání

S příchodem a rozmachem sociálních sítí se začaly objevovat případy vloupání, která zloději plánují za pomoci dat ze sociálních sítí. 75 % dopadených pachatelů vloupání se domnívá, že jiní pachatelé tato data využívají (Foster, 2015). To není překvapivé ve chvíli, kdy 50 % respondentů průzkumu na sociálních sítích uvedlo, že sdílí informaci o tom, že jsou na dovolené (Foster, 2015).

Krádeže identity

Kromě vloupání také narostl počet krádeží identity (ide, 2016). Používání dat ze sociálních sítí výrazně zjednodušilo používání cizí identity oproti vytváření falešné, a je používáno například v oblasti pojišťovacích podvodů.

Social engineering

Další možností zneužití dat je jejich použití pro další útoky, jako je *spear phishing* (phishingové útoky zaměřené na konkrétní osobu či instituci, využívající detailní znalosti tohoto cíle – oproti klasickým phishingovým útokům, které jsou plošné) a obecně v oblasti *social engineering*. Jednou z fází útoků spadající pod oblast *social engineering* je *information gathering* – získávání informací o cíli, a právě to

sociální sítě výrazně zjednodušily. Spear phishing útoky jsou často zaměřené na firmy, nástrojem pro sběr informací pak může být například profesní sociální síť LinkedIn (Beckers et al., 2017), nebo sběr informací probíhá pomocí nalezení specifických zájmů osoby (často dohledatelné na sociálních sítích), skrze které dojde k navázání kontaktu a získání důvěry (Hadnagy, 2011).

Neoprávněné sledování

Objevily se případy, kdy byly lokační data služby použity ke sledování osoby. V roce 2014 se ukázalo, že zaměstnanci společnosti Uber měli možnost sledovat lokace pasažérů (a tuto možnost využívali) – například známé osobnosti, novináře nebo svou bývalou přítelkyni (Morgan, 2017).

Neoprávněné zjišťování informací

Ve Spojených státech vyšetřování ukázalo, že zaměstnanci policie ve stovkách případů během dvou let nahlíželi do osobních záznamů, které nesouvisely s výkonem práce. Šlo například o data bývalých partnerů nebo novináře, který vydal kritický článek o místním policejním oddělení (Services, 2016).

Šíření fotografií dětí

Výrazně odlišným tématem, který však také ukazuje na rizika sdílení různých osobních dat, jsou opakované případy, kdy se fotografie dětí, sdílené jejich rodiči na sociálních sítích nebo sdílecích službách (ulozto.cz, rajce.cz), dostanou do oběhu ve skupinách pedofilů (Juna – Burýšek, 2020).

Úniky dat

Už běžnou součástí digitálního světa jsou úniky dat – jen v roce 2020 bylo zveřejněno téměř 4000 případů (Security, 2021). Tato data mohou být riziková z více důvodů, výraznou ukázkou je únik dat služby Ashley Madison – seznamovací aplikace pro zadané. Tato data byla využita pro vydírání uživatelů skrze e-maily obsahující citlivá data a vyhrožující jejich zveřejněním rodině a na sociálních sítích (Doffman, 2020).

Lokační data

Jedním z cenných druhů dat jsou lokační data. Při používání smartphonů v podstatě není možné si být jistý, že tato data nejsou sbírána. V čem může být tento sběr rizikový, upozorňuje například článek The New York Times (Thompson – Warzel, 2019), který ukazuje, jak je z lokačních dat a částečné znalosti o pohybu (například veřejně známé osoby) možné dosledovat celou její historii pohybu – tedy potenciálně i to, s kým a kde se setkávala, nebo kde bydlí (například u celebrit, které to cíleně tají).

Další případy – ve spojení s konkrétními postupy pro jejich zamezení – jsou rozebrány v kapitole Klíčové zásady uživatelské ochrany osobních dat.

Je možné si všimnout, že v několika z těchto případů se objevuje zmiňovaný princip *networked privacy* – tedy osoba, která data sdílí nemusí být stejná jako ta, která je terčem zneužití. To zvýrazňuje důležitost nahlížení na soukromí nejen z pohledu jednotlivce a jeho volby.

2.2.2 S dopadem na společnost

Henrik Skaug Sætra tvrdí, že nahlížení na soukromí pouze z pohledu jednotlivce a jeho svobodné volby se svými daty a soukromím naložit podle svého uvážení, není dostatečný (Sætra, 2020). Argumentuje tím, že volba jednotlivce v má tomto případě dopady i na ostatní – říká například, že si jednotlivec nemůže zachovat naprosté soukromí ve světě, kde nemají (i vlastní volbou) soukromí ostatní – soukromí jednotlivce je tedy neoddělitelně navázáno na soukromí ostatních: :

> „(...) it is impossible for me to be fully unknown in a world where everyone else is fully known.“ (Sætra, 2020)

Rozhodování jednotlivců pak vede k suboptimálním výsledkům pro společnost jako celek.

Soukromí pak nazývá *agregovaným veřejným statkem*, tedy něčím, co je poskytováno členům společnosti, a zároveň vyplývá ze společné aktivity většiny členů.

Tento pohled se dá stručně ukázat na příkladech konkrétních momentů, kdy se téma objevilo ve společenském povědomí.

Cambridge Analytica

Případ firmy Cambridge Analytica, která získala a zpracovala data z více než 50 milionů profilů na Facebooku (Cadwalladr – Graham-Harrison, 2018) pro použití v profilované politické reklamě, je pravděpodobně dosud největší kauza otevírající téma soukromí a osobních dat. Ač se po letech vyšetřování ukazuje, že byl pravděpodobně vliv menší, než se zdálo (BBC, 2020; Rathi, 2019), vyvolala mnoho otázek mimojiné o tom, jak můžou být osobní data využita pro politickou kampaň v dosud neviděném měřítku. Protože částí dat z Facebooku byly i vazby mezi jednotlivými lidmi, potvrzuje teze o *networked privacy* i Sætrův pohled na soukromí.

Čínský sběr dat

Česku bližší případ práce s osobními daty s potenciálním vlivem na celou společnost je firma Zhenhua Data Technology, která vytvářela na základě získaných veřejných a komerčně dostupných dat profily strategicky významných osob české společnosti. Šlo o politiky, pracovníky bezpečnostních složek či

podnikatele (Valášek – Horák, 2020). Česko bylo samozřejmě jenom jednou z mnoha zemí, o kterých tento sběr a analýza probíhaly, ukázal ale jasněji, jak může být s daty nakládáno. Zajímavé je, že i v tomto případě byla část dat z automatizovaného sběru ze sociální sítě Facebook, která tuto praktiku zakazuje. Firma Zhenhua byl pak dle mluvčí Facebooku ze sítě zablokována, ovšem ukazuje to na fakt, že ani po kauze Cambridge Analytica se nakládání s osobními daty a jejich ochrana u této firmy příliš nezměnila.

Citlivá lokační data

Příkladem ohrožení veřejných zájmů může být i případ, kdy fitness aplikace Strava zveřejnila mapu agregovaného pohybu uživatelů. Na mapě se tím objevilo umístění vojenských základen USA, například v Afghánistánu (Hern, 2018). Rozhodnutí jednotlivce (vojáka, jež využíval fitness aplikaci) zde tedy mělo jasný vliv na jeho okolí.

Shrnutí

Tato kapitola popsala, jaká jsou rizika plynoucí ze sběru osobních dat. Ač rozděluje případy na ty, které zasahují jednotlivce, nebo celou společnost, opakovaně se ukazuje, že toto rozdělení neodpovídá realitě, a je tedy třeba uvažovat o soukromí i skrze jiné pohledy.

Tím rozšířila pohled na problematiku o konkrétní příklady z reálného světa, a položila základ pro pohled na prevenci těchto rizik.

Kapitola 3

Hlavní zdroje dat digitální stopy uživatele

Úvod

Tato kapitola nabízí možnou kategorizaci dat podle toho, jak se uživatel podílel na jejich vytvoření, a podle jejich anonymity. Dále rozebírá konkrétní typy dat pro úplnější představu o tom, co všechno je ukládáno a zpracováváno.
Tato data budou využita v praktické části práce pro vytváření scénářů a modelových dat podle reálných typů dat a možnostech přístupu k nim.

3.1 Kategorie dat

Data, která vytváříme používáním technologií a digitálních služeb, lze kategorizovat několik způsoby. Dělení může být například:

- **aktivní** vs **pasivní** digitální stopu (Madden et al., 2007)
- případně doplněna o **vědomě nevědomou** digitální stopu (Fish, 2009)
- **identifikovatelná** vs **anonymní** data

3.1.1 Aktivní, pasivní a vědomě nevědomá digitální stopa

Toto dělení se dívá na data z pohledu uživatele, vědomí o jejich existenci, a možností vlivu.

Aktivní stopa

Aktivní stopou se myslí takový typ dat, který uživatel vědomě publikuje. Může jít o příspěvky, komentáře a reakce na sociálních sitích či osobních webech, fotky a jiné soubory nahrané na cloudová

úložiště, nebo vytvořené uživatelské účty (Madden et al., 2007).

V tomto pohledu nerozlišujeme, kdo je majitelem dat a jak s nimi kdo může nakládat, to závisí na podmínkách konkrétní platformy (omezené legislativou dané země).

Pasivní stopa

Pasivní stopa se skládá z typu dat, které uživatel vytváří svým používáním digitální platformy či služby, bez přímého sdílení či nahrávání. Nemusí si tedy vzniku těchto dat být vůbec vědom, zejména pokud má nedostatečnou představu o technologické stránkce používaných služeb. Obvykle jde o analytická data, používána pro lepší technický, bezpečnostní nebo marketingo-ekonomický efekt. Může jít o informace o zobrazení stránky či příspěvku, IP adresu a další technické parametry připojeného uživatele (například lokaci). Může jít i o kombinovaná data, například určení zájmů či demografické skupiny, vytvořené na základě jednotlivých dat.

Vědomě nevědomá stopa

Fish dále přidává kategorii vědomě nevědomé digitální stopy, která se skládá z dat aktivně vložených jinými uživateli (Fish, 2009). Může jít o fotografie – na sociálních sítích, ale například i z různých akcí (kde člověk zveřejnění dat musí v České Republice odsouhlasit), označení v příspěvcích nebo o data zveřejněná úřady. U části těchto dat může mít uživatel možnost ji zpětně omezit (například Facebook umožňuje odstranění označení).

3.1.2 Identifikovatelná a anonymní data

Toto dělení se dívá na data z pohledu právně-technologického.

Identifikovatelná data

Sem řadíme typ dat, které jsou přímo propojitelné s naší osobou. Jde pak o osobní údaje podle definice GDPR:

> „Pro účely tohoto nařízení se rozumí „osobními údaji“ veškeré informace o identifikované nebo identifikovatelné fyzické osobě (dále jen „subjekt údajů“); identifikovatelnou fyzickou osobou je fyzická osoba, kterou lze přímo či nepřímo identifikovat, zejména odkazem na určitý identifikátor, například jméno, identifikační číslo, lokační údaje, síťový identifikátor nebo na jeden či více zvláštních prvků fyzické, fyziologické, genetické, psychické, ekonomické, kulturní nebo společenské identity této fyzické osoby“ (Council of European Union, 2016)

Jde tedy o data, jež lze nějakým způsobem spojit s konkrétní osobou. S ohledem na rychlý vývoj v oblasti získávání dat, a zároveň malou mírou transparentnosti těchto procesů, je u dat těžké obecně říct, zda jsou identifikovatelná. Jak ukazuje dále příklad lokačních dat, i data která prošla procesem anonymizace mohou být ve skutečnosti identifikovatelná.

Anonymní data

Může jít o data anonymizovaná službou či mezivrstvou, nebo také o data vytvořená uživatelem s použitím anonymizačních nástrojů, jako je například VPN (virtuální privátní síť, používána pro skrytí IP adresy uživatele).

3.2 Konkrétní typy dat

3.2.1 Historie prohlížení

Jednou z významných součástí pasivní digitální stopy je historie prohlížení – tedy záznam navštívených webových stránek (případně webových requestů).

Samostatně jde o data anonymní (nejsou přímo spojená s naší osobou), ovšem v realitě to tak nemusí být.

Tato data může uchovávat internetový prohlížeč, a to buď lokálně nebo na cloudu. Například v případě prohlížeče Google Chrome se zapnutou synchronizací jsou tyto informace ukládány jako součást dat Google profilu.

Data o webové aktivitě má také poskytovatel internetového připojení (Internet Service Provider – ISP), který je zároveň má spojená s naší IP adresou, která je vázaná na konkrétní smlouvu o poskytování internetu. V České Republice si tato data může vyžádat policie, a prokazatelně to dělá (Vokuš, 2019).

Kromě toho je možné aktivitu uživatelů napříč weby sledovat pomocí cookies třetích stran a různých typů fingerprintingu.

Cookies

Jak bylo zmíněno v předchozí kapitole, cookies jsou soubory, které si stránka ukládá do počítače uživatele, aby ho mohla identifikovat při dalších požadavcích (zde mluvíme o *first-party cookies*). V současnosti mnoho webů obsahuje takzvané cookies třetích stran *third-party cookies*, které umožňují reklamním službám (například Google Ads) sledovat aktivitu napříč webem. Google v březnu 2021 oznámil postupný odchod od používání cookies třetích stran, které plánuje nahradit vlastní technologí, a ve svém prohlížeči Google Chrome je zakázat (a tím následovat další prohlížeče, jako Firefox či Safari).

Device a browser fingerprinting

S postupným legislativním tlakem na omezení rozsahu cookies se začaly služby přesouvat k používání *fingerprintingu*, tedy používání jakéhosi otisku zařízení nebo prohlížeče, ze kterého uživatel k webům přistupuje. Používané techniky jsou rozmanité, od získávání informací o prohlížeči a operačním systému (verze, jazyk, instalované doplňky a další), po *canvas fingerprint* využívající specifika v renderování webového prvku `canvas`, které se liší podle GPU nebo grafických ovladačů v daném zařízení.

Zjistit svůj browser fingerprint lze například pomocí služby Am I Unique. Služba ukazuje, kolik informací je prohlížeč schopen získat, a to i bez schválení uživatelem.

Behavioral profiling

Relativně nově používanou technikou je *behavioral profiling*, který se snaží rozlišit uživatele podle jejich chování – primárně charakteristik pohybu myši nebo způsobu psaní na klávesnici (Yang, 2010; Jorgensen – Yu, 2011; Ikuesan – Venter, 2019).
Tento způsob je stále ještě méně spolehlivý než dříve zmíněné, neboť využívá výrazně rozmanitější typ dat. V kombinaci s jinými nástroji, a s rostoucími možnosti vyhodnocování těchto dat (například využitím strojového učení), se však stává nezanedbatelnou možností v repertoáru nástrojů pro profilování a identifikaci uživatele.

3.2.2 Lokační data

Jak bylo zmíněno v předchozí kapitole, jedním z typů využívaných dat jsou data lokační.

Tato data mohou například poskytnout vhled, do jakých obchodů uživatel chodí, a nabídnout tak *remarketing* napříč fyzickým a digitálním světem.

Lokační data je v současnostni snadné získat, zejména díky množství mobilních zařízení připojených k internetu. I aplikace, které reálně data pro svoje funkce nepotřebují, o ně mohou uživatele požádat. Prodejem těchto dat pak mohou získat finance na samotný vývoj aplikace „zdarma“.
Je na místě říct, že možnosti prodeje dat se mohou výrazně lišit podle toho, v jaké zemi – a tedy pod jakou legislativou – se uživatel pohybuje.

Lokační data dále potenciálně sdílejí aplikace, které reálně data používají pro svoje služby – mapy (*Google Maps*, *Mapy.cz* a podobné), dopravy (*IDOS*, *Pubtran*, *PID Lítačka* a podobné) a dovozy (*DámeJídlo*, *Uber* a podobné).

Například Google Maps uchovává lokaci a přiřazuje ji ke konkrétním místům (obchody, úřady a podobně). Zároveň uchovává informaci o předpokládané formě aktivity (např. chůze, běh, kolo, auto, hromadná doprava) včetně důvěry (pravděpodobnosti) v predikci.

Mohlo by se zdát, že lokační data jsou anonymní (typicky obsahují nějaké náhodné ID zařízení, časovou značku a lokaci), v realitě jsou však velmi snadno deanonymizována. Profesor Ohm z Georgetown University říká, že přesná geolokační data je naprosto nemožné anonymizovat. Jediná méně anonymizovatelná data jsou podle něj genetická data:

> „Describing location data as anonymous is "a completely false claim" that has been debunked in multiple studies, Paul Ohm, a law professor and privacy researcher at the Georgetown University Law Center, told us. "Really precise, longitudinal geolocation information is absolutely impossible to anonymize."
>
> "D.N.A.," he added, "is probably the only thing that's harder to anonymize than precise geolocation information"." (Thompson – Warzel, 2019)

3.2.3 Data vkládaná na sociální sítě a další platformy

Nejvýznamnější složkou aktivní (veřejné) digitální stopy jsou pravděpodobně v současné době sociální sítě a jejich obsah. Jde o příspěvky, ale i komentáře, reakce (emotikony), projevení zájmu o událost a další vědomé interakce, jejichž výstupem je informace sdílená s dalšími uživateli.

Kromě toho, že jsou součástí celkového „balíku" dat, které služby mají, může být jejich riziko i v nastavení, kdo se k datům může dostat. Kvůli nevědomosti uživatele nebo chybou dané služby – například v roce 2018 byly příspěvky přibližně 14 milionů uživatelů veřejné místo soukromých (Frenkel, 2018) – se může stát, že jsou data veřejná pro jinou skupinu lidí, než si uživatel myslí.

A zároveň, jak již bylo uvedeno v první kapitole, jsou to data dobře využitelná do modelů, které dokáží o uživateli odvodit i informace, které vědomě nepíše, a která jsou mnohdy velmi citlivá.

Fotografie

Specifickou částí zveřejňovaných dat jsou fotografie. U nich může často docházet ke vzniku *vědomě nevědomé* digitální stopy, když se člověk objevuje na fotkách, které sdílí někdo jiný (a naopak, pokud sdílí fotografie, na kterých je někdo další). Pokrok ve strojové analýze obrazu už umožňuje, aby byly osoby na fotografii rozpoznány automaticky (Facebook díky tomu může u nahrané fotky navrhnout osoby k označení, Google Photos umožňuje zobrazit všechny fotografie dané osoby s velmi vysokou přesností). Zároveň je často možné z fotografie získat další informace z *metadat*. To jsou data, která vytváří fotoaparát (dnes obvykle smartphone), a můžou obsahovat informace jako čas a datum, lokaci, název zařízení nebo technické parametry jako délku expozice či ISO. Některé služby (například Facebook) při nahrávání fotografie citlivá metadata smaže, aby nebylo možné uložením fotografie získat i je, ale uživatel s tím při nahrávání fotografie obecně počítat nemůže. Kromě toho se s rozvojem strojového učení ukazuje, že bude nejspíš možné lokaci často velice přesně určit i bez lokačních metadat, a to až s přesností ulice či domu (Neil, 2016).

3.2.4 Další data sledovaná sociálními sítěmi a podobnými platformami

Kromě dat, které na sociální sítě uživatel vědomě vkládá, tyto služby sbírají a vytváří další data, napojená na jeho profil. Alespoň část je možné vidět vyžádáním těchto dat od dané služby. Toto právo v EU zajišťuje GDPR.
V datech z Facebooku je možné vidět typy dat, o kterých uživatel obvykle neví. Tato data nejsou nikde ve službě přímo zobrazována a nejsou třeba pro uživatelskou funkcionalitu služby:

- **friend_peer_group**
  Facebook si na základě dat o uživateli a jeho vazbách na další uživatele jeho profil zařadí do některé z předem definovaných skupin, například *Začínající dospělý život*, a tu pak využívá pro cílení reklamy a celkové nastavení toho, co uživatel vidí.
- **viewed**
  Facebook sleduje ve speciální kategorii
  - jaká videa, jakou část z nich, a kolik času celkově uživatel strávil u videí z Facebook View
  - jaké zboží na Facebook Marketplace si prohlížel
  - na jaké zobrazené reklamy reagoval

  Celkově zde jde o data, která jsou velmi cenná z pohledu reklam, monetizace a udržení uživatele ve službě.
- **ads_interests**
  Facebook k uživateli přiřazuje zájmy, které jsou pak využívány na reklamních aukcích.
- **unfollowed_pages**
  Facebook zachovává historii toho, jaké stránky uživatel přestal sledovat.
- **removed_friends**
  Stejně tak uchovává informaci o ukončených „přátelstvích“.
- **group interactions**
  Počet interakcí (příspěvků, komentářů, reakcí) ve skupinách, jejichž je uživatel členem.
- **people interactions**
  Počet interakcí s osobními profily.
- **your_topics**
  Podobně jak ads_topics – seznam témat, jež se Facebook domnívá, že uživatele zajímají, a podle nich přizpusobuje zobrazovaný obsah.

Podobně je možné se podívat na některá data, která vytváří a uchovává Google (samozřejmě podle toho, které služby uživatel používá). Zde jsou uvedena některá překvapivější, která se nezdají nutná pro základní funkcionalitu (v této sekci jsou uvedena data, která nebyla zařazena do jiné sekce této kapitoly). U každého bodu je pak uvedeno (pokud to je možné zjistit), skrze jakou službu tato data vznikají:

- **názvy tisknutých souborů**

  Služba *Google Cloud Print*, umožňující tisk přes internet, uchovává seznam s názvem tisknutých souborů.
- **informace o hrách**

  O uživatelích služby *Google Play Games* jsou uchovávána například data o tom, kdy poprvé a kdy naposledy hráli danou hru, a souhrnné herní statistiky. Kromě toho jsou v jiné sekci uloženy všechny záznamy spuštění her.
- **hudba**

  Služba *Google Play Music* uchovává mimojiné informaci o počtu přehrání jednotlivých skladeb, a seznam všech spuštění nahrávek s jejich časovou značkou. To platí i u služby *Google Podcasts*.
- **aplikace**

  Služba *Google Store* – tedy nejpoužívanější služba pro správu a instalaci aplikací na zařízeních s operačním systémem *Android* – kromě seznamu instalovaných aplikací zaznamenává údaje o zařízení (konkrétní model telefonu), na kterých byla instalována a telefonního operátora. Dále je uložena informace o každé instalaci a odinstalaci aplikací.
- **reklamy**

  Google uchovává informaci o všech reakcích (kliknutích) uživatele na zobrazené reklamy.
- **Android**

  Google uchovává seznam všech otevření aplikací. Vztahuje se na uživatele používající operační systém *Android* s přihlášeným profilem Google.
- **hlasové pokyny**

  Hlasové pokyny pro službu *Google Assistant* jsou ukládány jako originální zvukové nahrávky. To se děje i v případech, kdy začalo zařízení poslouchat „omylem“ – kdy uživatel neřekl klíčové slovo pro začátek poslechu, ale jiný zvuk tak byl mylně interpretován.
- **vyhledávání**

  Služba *Google Search* uchovává informaci o každém vyhledávání a o tom, kam z něj uživatel pokračoval

3.2.5 E-mail

Velká část online komunikace se stále děje přes e-mailové služby, zejména v případě komunikace pracovní. Ta může obsahovat mnoho (zejména pro firmu) citlivých informací, a jejich únik může mít velké následky (jak je vidět na příkladu uniklých e-mailů Hillary Clinton v prezidentské volbě v USA v roce 2016).

Kromě samotného obsahu e-mailu se dnes používají technologie, které umožňují sledovat uživatelovu interakci s přijatým e-mailem. Samotná technologie e-mailu toto sledování nenabízí (oproti různým chatovacím službám, nebo protokolu RCS, postupně nahrazujícímu SMS). Často používanou technikou je *pixel tracking*, který funguje pomocí vložení „neviditelného“ obrázku o velikosti jednoho pixelu, který

se při otevření e-mailu nahrává z unikátní adresy, a tento dotaz na danou adresu je možné zaznamenat. U tom, že je otevření e-mailu sledování, obvykle uživatel vůbec neví.

3.2.6 Finanční záznamy

V době, kdy velká většina finančních transakcí neprobíhá s fyzickými penězi (zejména v České Republice, která je k adopci nových technologií v oblasti převodu peněz velmi otevřená), se vytváří další část digitální stopy právě v této oblasti. Primárně mají veškeré záznamy k dispozici banky, u kterých transakce probíhají. Tyto informace podléhají bankovnímu tajemství, může si je však v oprávněných případech (dle § 8 zákona č. 141/1961 Sb., trestní řád) vyžádat policie, a to bez vědomí osoby, jejíž informace jsou nahlíženy. Výpisy z účtu můžou být vyžadovány i v jiných případech, třeba při bezpečnostní prověrce.

Znatelná část finančních transakcí se rovněž přesouvá do plateb telefonem, které probíhají přes dalšího prostředníka – data o plabách tedy nemá již jen banka, ale i další soukromá firma. Často používanými nástroji pro tyto platby jsou Alipay a WeChat (obojí primárně v Číně), Apple Pay, PayPal, Samsung Pay, Amazon Pay nebo Google Pay. Většinou tedy jde o firmy, které poskytují i další služby a historie transakcí pak může být dalším zdrojem dat o uživateli.

3.2.7 Zdravotní data

Další oblastí velmi citlivých dat jsou data zdravotní. S postupnou digitalizací zdravotnictví bude možné tato data čím dál více propojovat, což nese nesporné výhody ve snížení zdravotnické administrativy a urychlení péče, ale i potenciální rizika v případě úniku či zneužití těchto dat.

Kromě dat ve zdravotnictví se čím dál rozmanitější typy dat sbírají skrze *wearables* – nositelnou elektroniku, často nabízející různé fitness a zdravotní funkce. Aktuálně se tedy můžeme setkat se sběrem dat o pohybu, měřením tepu a dechu, nebo měření odporu kůže či množství kyslíku v krvi. Tato data jsou pak automaticky zpracovávána pro rozpoznávání typu a intenzity pohybu či sportu, doby a kvality spánku nebo míru stresu.

Dá se očekávat, že se budou objevovat další typy sledovaných dat, sensory budou přesnější a z dat (a to i historických) bude možné vyčíst více informací – probíhají mnohé výzkumy, které se snaží v datech najít návaznost na různé zdravotní jevy, příkladem může být studie propojující informace o variabilitě tepu s psychickým i celkovým zdravím (Coutts et al., 2020).

Tato data typicky opět sbírají soukromé firmy, často nabízející mnoho dalších služeb – uvést můžeme Google (který v roce 2019 koupil firmu Fitbit, výrazného výrobce fitness zařízení), Apple či Samsung.

Ještě výrazně problematičtější částí problematiky zdravotních dat může být v posledních letech velmi populární analýza DNA soukromými firmami. DNA je ze své podstaty extrémě osobní informace. Ani firmy psokytující tento typ služeb (a spravující genetická data, které jim klienti poskytují k dalšímu

využití) nejsou odolné vůči únikům dat, jak je vidět na případu firmy MyHeritage, které v roce 2018 unikla data 92 milionů účtů. V tomto případě nedošlo k úniku genetických dat, to je ovšem pravděpodobně jen otázka času. Aktuálně jsou genetická data užitečná zejména pro výzkum v oblasti medicíny a léčiv, s rostoucím pochopením genomu je těžké odhadnout, jakými způsoby by mohla být tato data využita či zneužita.

Shrnutí

Tato kapitola popsala možné způsoby kategorizace osobních dat a ukázala některé běžné typy osobních dat. Zároveň u jednotlivých dat opět ukazuje rizika s nimi spojená.

Kapitola 4

Klíčové zásady uživatelské ochrany osobních dat

Úvod

Ač se posouvá i právní ochrana uživatelů, která může provést větší zásahy do způsobů, kterými firmy s daty nakládají, každý uživatel má alespoň omezenou možnost kontroly nad svými daty a jejich ochranou. Tato kapitola nabízí pohledy a nástroje, které může uživatel použít pro regulaci vlastního soukromí v digitálním světě.

4.1 Obecné náhledy

Kromě přímých nástrojů a postupů je dobré si osvojit několik základních pohledů na osobní data a nakládání s nimi.

4.1.1 Co je na internetu, to je veřejné

Informace, které zveřejníme pomocí digitálních platforem (sociální sítě, chatovací aplikace a další) sice působí, že sdílíme jen s omezenou skupinou lidí, je však důležité si uvědomit, že je z jejich podstaty možné je snadno sdílet i mimo tento okruh. Opakovaně se to dělo ať chybou v systému, tak cíleným sdílením dat.

U informací, které dává uživatel na internet, je tedy třeba myslet na potenciální zveřejnění dat a rizika s tím spojená.

4.1.2 Propojenost

Ač může mít uživatel pocit, že nemá důvod se o své soukromí starat, je dobré si uvědomovat, že vytvářením a zveřejňováním dat zaprvé poskytujeme data pro zpřesňování modelů a nepřímo ovlivňujeme i další uživatele. Zároveň v některých situacích, kdy můžeme mít pocit, že sdílíme pouze svoje osobní data (například lokační data), v kombinaci s daty jiné osoby může vznikat úplně nová informace (například o tom, kdo se s kým setkává), kterou uživatel přímo nesdílel. A dopad na ostatní osoby je jasný ve chvíli, kdy přímo sdílíme informace o někom jiném (například fotografie, kontakty a adresy).

4.1.3 Budoucnost

Kromě toho je třeba myslet na to, že informace, které nyní vytvářím, budou existovat potenciálně navždy a mohou se mi vrátit v dalším životě – například při hledání zaměstnání, nebo pokud se stanu veřejně činnou osobou.

4.2 Zásady a nástroje

4.2.1 Omezení vědomého sdílení informací

Zejména sociální sítě vedou své uživatele k tomu, aby sdíleli mnoho informací ze svého života. Tyto informace však mohou uživatele ohrozit, například zvýšeným rizikem vloupání, krádeže identity, cíleného útoku (*social engineering*, *spear phishing*), nebo sdílení těchto informací někam, kde bychom je nechtěli (viz kapitolu *Rizikové dopady sběru osobních dat*).

Základní zásadou tedy je si být vědomý toho, jaké informace vědomě sdílím, včetně vědomí, že se mohou dostat kamkoli a kdykoli.

4.2.2 Pravidla pro aplikace a služby

Souhlasy cookies

Pokud chci omezit množství dalších dat, které vznikají mým běžným chováním, a které jsou propojitelné do jednoho celku (profilu), jednou z prvních možností je dbát na možnosti nastavení cookies a sběru dat, které díky GDPR v EU máme. Stránka je povinna nám nabídnout možnosti nastavení cookies, včetně odmítnutí těch, které nejsou esenciální pro funkcionalitu služby. Bohužel je však tato možnost stále často příliš složitá či nesrozumitelná, a mnoho uživatelů povolí vše. Jen některé služby nabízí snadné odmítnutí všech neesenciálních cookies.

Uložené cookies je také možné z počítače smazat, obvykle v nastavení prohlížeče. Některé prohlížeče

zaměřené na zvýšené soukromí dále umožňují jak omezit prvotní ukládání (zejména cookies třetích stran), tak jejich automatické mazání po zavření okna nebo po určité době.

Možnost smazání dat

Také je možné u služeb vyžádat mazání dat. Na toto mají opět nárok uživatelé v EU díky GDPR, a služby mají povinnost tuto možnost snadno nabízet. Často však jde o mazání celého účtu a není možné snadno smazat jen část dat. Případně to služby umožňují, ale běžný uživatel o tom nemusí ani vědět. Existují služby, které toto mohou za uživatele dělat automaticky, například aplikace Jumbo.

Do Not Track

S webovým požadavkem je možné poslat signál *Do Not Track*. Aby se tak dělo, je třeba toto nastavení zapnout v možnostech prohlížeče. Tento signál dává službám najevo, že uživatel nechce, aby byla jeho aktivita sledována. Aktuálně však není žádná zákonná povinnost ho jakkoli zpracovat, ani neexistuje konsensus na tom, jak by služby, které toto přání uživatele respektovat chtějí, měly postupovat. To se snad v budoucnu mění, neboť probíhají snahy tento postup sjednotit a dát mu zákonnou váhu (Brandom, 2021)

4.2.3 Anonymní prohlížení webu

Anonymní režim

Snad všechny běžně používané prohlížeče nabízí takzvaný *anonymní režim*. Je důležité si být vědom toho, co tento režim dělá a co ne. Základním principem je, že soubory cookie, a celkový kontext prohlížení, uchovává pouze po dobu, kdy je dané okno anonymního prohlížení otevřeno - potom, co okno smažu, jsou všechny cookies smazány. To ovšem uživatele nechrání před ostatními typy sběru dat a identifikace - dříve zmíněný *browser fingerprinting*, *behavioral profiling* nebo *tracking pixel* budou fungovat dále. Stejně tak není nijak pozměněno, jaké informace se dostávají s poskytovateli internetu. Bohužel ani na část ochrany soukromí, kterou by anonymní režim poskytovat měl, se nelze spolehnout. Opakovaně byly objeveny způsoby, jak díky chybám software sledovat uživatele používající anonymní režim, například pomocí cachování ikon stránek (Goodin, 2021).

Prohlížeče

Jak již bylo zmíněno, různé prohlížeče nabízí různou míru základního nastavení soukromí i různé možnosti pro uživatelské nastavení. V posledních letech vznikají prohlížeče, které této problematice dávají větší důraz – za zmínku stojí třeba Vivaldi nebo Brave. Dlouhodobě je také velkým advokátem soukromí Mozzila Foundation s jejich prohlížečem Firefox. Na druhé straně stojí prohlížeč Google

Chrome, který jako součást ekosystému služeb Google defaultně sbírá větší objem dat.

Specifickým případem prohlížeče je Tor Browser (případně další prohlížeče používající stejnou technologii), který umožňuje šifrovanou komunikaci, tedy chrání o před sledováním ze strany poskytovatele internetu, a stejně jako VPN umožňuje obcházení omezení dané například státem, ve kterém se uživatel nachází.

VPN

VPN (*virtual private network – virtuální osobní síť*) je technologií, jež umožňuje připojení ke službám skrze jiné servery, anonymizuje tedy naši IP adresu a s ní spojené informace (zejména lokaci). Je tedy možné je používat pro obcházení geografických omezení, či rizika sledování poskytovatelem (stejně jako u sítě Tor).

Doplňky

Proti některým konkrétním sledovacím praktikám se dá bránit specializovanými nástroji, dostupnými buď jako samostatné programy, nebo často jako doplňky do prohlížečů. Lze se takto bránit proti *canvas fingerprinting*u pomocí doplňků jako je Canvas Blocker nebo Canvas Fingerprint Defender, blokovat *tracking pixely* v Gmailu pomocí doplňku PixelBlock nebo e-mailových služeb se zabudovanou funkcí (například HEY), i proti *behavioral profiling*u pomocí doplňku přidávajícímu náhodnou prodlevu ke stiskům kláves. Kromě toho existuje mnoho doplňků, kombinujících více technik ochrany soukromí a blokace obsahu - mezi nejznámější patří AdBlock+, uBlock, Ghostery nebo Privacy Badger.

Vyhledávací enginy

Pokud se chceme vyhnout sběru dat při používání vyhledávačů, je třeba se poohlédnout po alternativách k nejpoužívanějším vyhledáváčům, jimiž jsou Google, Bing a v česku Seznam. Variantou, která neuchovává historii a nevytváří profil uživatele (a tím pádem nenabízí personalizované výsledky), je vyhledávač DuckDuckGo.

4.2.4 Finanční transakce

Je možné, i když komplikované, chránit svoje soukromí i v případě finančních transakcí. Běžné finanční operace probíhají přes banky, které naši identitu znají, možnost anonymní výměny peněz však přinesly v posledních letech kryptoměny. Ty umožňují založení anonymní peněženky, na kterou je možné převést například hotovost, a pak převádět peníze mezi peněženkami bez odhalení identity osoby.

4.2.5 Šifrování dat

Dalším způsobem omezení možností práce s daty, které vytváříme, je jejich šifrování. To se samozřejmě vztahuje jen na část možných digitálních služeb, zejména na komunikaci.

Šifrovaná komunikace

Jak u chatů, tak u e-mailu můžeme hledat varianty, kdy je komunikace šifrovaná. To znamená, že je chráněna před čtením třetí stranou. Je důležité si být vědom, že jsou různé způsoby šifrování, a jen některé zaručují, že komunikaci opravdu nemůže číst nikdo kromě komunikujících (tedy ani daná služba). Dále je podstatné, že ač tuto možnost některé služby (například WhatsApp) umožňují, šifrování neprobíhá vždy, ani nemusí být nastaveno jako základní možnost.

Nejznámější chatovací služby zaměřené na šifrovanou komunikaci jsou Signal a Telegram.

Šifrovanou komunikaci můžeme chtít i u e-mailové služby, pak je možné využít například ProtonMail.

4.2.6 Zabezpečení dat proti útočníkům

Ač se tím dostáváme více do oblasti bezpečnosti než soukromí, je nezbytné zmínit i základní zásady ochrany před útočníky, neboť i to může být způsob, jak se naše osobní data mohou dostat na veřejnost a být zneužita.

Ochrana zařízení

Primárně do důležité – a zároveň relativně jednoduché – chránit svá zařízení před útoky. Dva hlavní způsoby jsou:

Udržování aktuálního softwaru
Jedním z častých důvodu průniku do zařízení je neaktualizovaný software, který obsahuje známé bezpečnostní chyby. Typickým případem může být používání již nepodporované verze operačního systému a stará verze prohlížeče. V současné době se obvykle moderní osperační systémy i prohlížeče aktualizace vyhledávají sami a upozorňují (pokud přímo nevynucují) na aktualizace uživatele.

Antiviry
Zvýšenou ochranu pak uživatel získá použitím specializovaného software na odhalování rizikového kódu, například ve stažených souborech (což je jeden z častých způsobu napadení zařízení). Nabídka antivirových a podobných bezpečnostích programů je široká, zroveň novější verze operačního systému Microsoft Windows obsahují i zabudovaný nástroj Microsoft Defender Antivirus.

Phishing

Počátek phishingových útoků se datují do devadesátých let. Jde o konkrétní formu útoků typu *social engineering*, tedy typu útoků, který cílí uživatele a skrze něj se snaží prolomit zabezpečení. Phishing probíhá typicky formou e-mailů, které se vydávají za jiné osoby nebo instituce a chtějí po uživateli aby stáhl soubor či použil odkaz. Phishingové kampaně jsou stále oblíbenou formou útoků – v roce 2020 nahlásilo 75 % firem, že zaregistrovali phishingový útok, a v 57 % proběhl alespoň jeden úspěšně (Whitney, 2021).
Ochrana proti těmto útokům nemá jednoduché technické řešení, ale lze dát několik doporučení, jak phishingový útok rozpoznat:

- Firmy nebudou požadovat osobní data (heslo, číslo OP, číslo bankovní karty) v e-mailu ani po telefonu.
- Obecně je vhodné nestahovat soubory a neotevírat odkazy z e-mailu (ale ne vždy se tomu lze vyhnout).
- Při otevírání odkazu z e-mailu je vždy vhodné zkontrolovat, zda adresa opravdu odpovídá tomu, co očekáváme.
- Je bezpečnější zkopírovat adresu z e-mailu a vložit ji ručně do prolížeče, než kliknout na odkaz v e-mailu.
- Phishingové e-maily častěji obsahují gramatické nebo technické chyby.
- Phishingové e-maily častěji tlačí příjemce k rychlé akci (hrozí sankcemi, upozorňují na rizika).

Hesla a zabezpečení

Důležitou částí ochrany svých účtů, a tedy dat, která jsou v nich k dispozici, je dostatečné zabezpečení přihlašování. Téma hesel by vydalo na samostatnou práci, zmíním tak alespoň několik základních zásad a pravidel.

Bezpečné heslo Je důležité zvolit takové heslo, které je dostatečně složité prolomit. Útoky směřující k prolamování hesel obvykle využívají dvou technik (nebo jejich kombinace) – *brute force*, kdy útočník zkouší náhodné kombinace znaků, a *slovníkový útok*, který využívá známá slova (případně hesla známá z přechozích úniků).

To vede ke třem základním poučkám při tvorbě hesel:

- **Heslo by mělo být dostatečně dlouhé.**
  Tím je lépe chráněno před *brute force* útokem, protože počet kombinací u delších retězců je takový, že již není praktické vyzkoušet všechny. Doporučovaná délka se mění s růstem výkonu počítačů i s pokrokem hashovacích algoritmů, které jsou pro uchovávání hesel používány, hrubé doporučení může být kolem 10 znaků (pokud používáme všechny ASCII znaky) až 25 znaků (pokud používáme pouze velká a malá písmena)

- **Heslo musí být náhodné**
  Pokud heslo je celé nebo z většiny existujícím slovem (větou, souslovím), výrazně to snižuje jeho bezpečnost. Heslo by mělo být náhodné – buď řetězec náhodných znaků, nebo (lépe zapamatovatelné) kombinace více náhodně vybraných slov.
- **Heslo musí být unikátní**
  Pokud používáme stejné heslo pro více účtů, při úniku hesla je ohrožení větší. Je tedy zásadní mít odlišná hesla v různých službách.

Správci hesel Naplnit předchozí tři poučky a při tom si pamatovat hesla k desítkám až stovkám svých účtů není reálné. Proto velká část uživatelů tyto zásady nedodržuje (i v případě, že si jich jsou vědomi). Bezpečnostní experti proto obvykle doporučují používat nějakou formu správce hesel – místo, kde jsou hesla zapsána, ideálně v šifrované podobě pod hlavním heslem/klíčem.

Různé správce hesel již nabízejí v základní výbavě internetové prohlížeče, u nich však není vždy míra zabezpečení největší. Proto jsou i specializované nástroje, které typicky nabízí větší zabezpečení a větší škálu možností.

Vícefaktorová autentizace

Dalším způsobem, jak výrazně zvýšit zabezpečení svého účtu, je vícefaktorová autentizace. Ta obecně znamená, že pro přihlášení je třeba více než jedna věc (což by bylo obvykle heslo). Faktory, které jsou pro přihlášení použité, se dají rozdělit do následujících částí:

- **znalost** – *něco, co vím*
  Jde o heslo, PIN, případně o odpověď na bezpečnostní otázku.
- **vlastnictví** – *něco, co mám*
  To může být speciální bezpečnostní klíč, ale v současnosti běžně také mobilní zařízení.
- **biometrie** – *něco, co jsem*
  Se zlevňováním potřebných senzorů a technologií je běžnou součástí zařízení detektor otisku prstu, kamera schopná rozpoznat obličej, a rozpoznávání hlasu. Specializovaná zařízení se pak dají použít například pro kontrolu podle duhovky. Sem lze také zařadit dříve zmíněný *behavioral profiling* použitý pro autentizaci.
- **lokace** – *kde se nacházím*
  V některých případech lze jako další faktor používat lokaci, a to buď vázanou na konkrétní IP adresy (například zařízení v kanceláři), nebo (skrze mobilní zařízení) reálnou lokaci.

Mnoho služeb nabízí dvoufaktorovou autentizaci (2FA), obvykle v kombinaci hesla a zařízení (použití mobilního zařízení jako klíče při přihlašování na jiném zařízení) nebo biometrie (při přihlašování na zařízení, které tuto formu nabízí).

Některé služby dokonce tuto míru zabezpečnení vyžadují – ať už ze zákona (služby pro komunikaci s úřady, jako je v česku Datová schránka), nebo kvůli prioritizaci zabezpečení (e-mailová služba HEY).

Shrnutí

V této kapitole byly detailně rozebrány konkrétní přístupy a nástroje, které uživateli mohou pomoci zvýšit kontrolu nad digitální stopou a zabezpečit data. Ukázala, že možností kontroly je více, a jejich použití vychází z potřeb a požadavků každého uživatele.

Kapitola 5

Praktická část - Východiska

Úvod

Praktická část této balakářské práce se zabývá navržením a vytvořením prostředí (*Aplikace*) pro vzdělávání této tématiky. Je rozdělena do dvou kapitol. První kapitola popisuje východiska aplikace, druhá pak její technický návrh a řešení.

5.1 Zařazení problematiky digitální stopy do školních výukových materiálů

Úvod a vymezení pojmů

První oblastí východisek je zařazení tématu osobních dat a digitální stopy do kontextu vzdělávacích materiálů. Zaměřuje se na to, jakou roli a prostor v nich aktuálně (v kontextu narůstající důležitosti tématu ve společnosti a společenských debatách) zaujímá.

Pro snazší orientaci v kapitole zde shrnuji použité zkratky a kódy a jejich význam:

RVP – Rámcový vzdělávací plán, definován (NÚV, b) je následovně:

> „Rámcové vzdělávací programy (RVP) tvoří obecně závazný rámec pro tvorbu školních vzdělávacích programů škol všech oborů vzdělání v předškolním, základním, základním uměleckém, jazykovém a středním vzdělávání. Do vzdělávání v České republice byly zavedeny zákonem č. 561/2004 Sb., o předškolním, základním, středním, vyšším odborném a jiném vzdělávání (školský zákon).“ (NÚV, b)

Klasifikace SŠ (dle (NÚV, c)):

Obory s maturitou

SŠ (M) – úplné střední odborné vzdělání s maturitou (obory kategorie M)

„příprava má profesní charakter a délka studia je 4 roky. Po maturitě lze pokračovat ve vzdělávání na vysoké nebo vyšší odborné škole.“ (NÚV, c)

SŠ (L) – úplné střední odborné vzdělání s odborným výcvikem a maturitou (obory kategorie L)

„studium připravuje pro náročná dělnická povolání a nižší řídicí funkce. V denní formě je 4leté a jeho významnou součástí je odborný výcvik (obory vznikly z dřívějších 3letých učebních oborů). Absolventi získávají maturitní vysvědčení a mohou pokračovat ve vzdělávání na vysoké nebo vyšší odborné škole.“ (NÚV, c)

SŠ (K) – úplné střední všeobecné vzdělání (obory kategorie K)

„všeobecná příprava ve 4letých a víceletých gymnáziích je neprofesní a připravuje především pro vysokoškolské nebo vyšší odborné vzdělávání.“ (NÚV, c)

Obory s výučním listem

SŠ (H) – střední odborné vzdělání s výučním listem (obory kategorie H)

„tradiční učební obory s tříletou přípravou ve středních odborných učilištích. Po získání výučního listu lze pokračovat navazujícím nástavbovým studiem a získat i maturitu.“ (NÚV, c)

SŠ (E) – nižší střední odborné vzdělání (obory kategorie E)

„studium je tříleté nebo dvouleté, výstupem je výuční list. Obory mají nižší nároky v oblasti všeobecného i obecně odborného vzdělání a jsou určeny především pro žáky se speciálními vzdělávacími potřebami, např. pro absolventy dřívějších speciálních základních škol a žáky, kteří ukončili povinnou školní docházku v nižším než 9. ročníku základní školy. Obory připravují pro výkon jednoduchých prací v rámci dělnických povolání a ve službách.“ (NÚV, c)

5.1.1 Vymezení dokumentů

Pro prozkoumání, jak jsou témata zařazena ve vzdělání, budou primárním zdrojem Rámcové vzdělávací plány (dále RVP). Je nutné se dívat na jejich návaznost, podobnosti a rozdíly na různých stupních a zaměřeních vzdělávání.

Konkrétně tedy budou prozkoumány RVP pro základní vzdělávání, které definují vzdělávací oblast Informatiky a její cíle, a dále RVP pro gymnázia a vybrané odborné školy se zaměřením na Informatiku a Informační technologie.

Kromě budou rozebrány aktuální plány na revizi oblasti Informatiky a ICT, která může přinášet změny i v této oblasti, a zároveň ukazovat na možné trendy v pojetí vzdělávání v oblasti digitálního světa a technologií a s tím souvisejících témat.

Konkrétně tedy kapitola prozkoumá propojení na následující dokumenty:

- Rámcové vzdělávací plány
  - RVP základní školy (NÚV, 2021b)
  - RVP-G (kategorie L) (NÚV, 2021a)
  - RVP pro odborné školy – Informační technologie (MŠMT, b) (kategorie M)
  - RVP pro odborné školy – Informační služby(MŠMT, a) (kategorie M)
- Návrh revizí rámcových vzdělávacích programů v oblasti informatiky a informačních a komunikačních technologií (NÚV, a)

5.1.2 Rámcové vzdělávací plány

RVP základní školy

Vzdělávací oblast Informatika

Ve vymezení Cílového zaměření vzdělávací oblasti je v kontextu tématu důležitý tento bod

> „uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka“ (NÚV, 2021b)

konkrétně však tento bod není vymezen cíli ani učivem, které by s oblastí osobních dat a digitální stopy přímo souviselo.

RVP-G

Vzdělávací oblast Informatika a informační a komunikační technologie

Vzdělávací oblast Informatika a informační a komunikační technologie v RPV pro vyšší stupně vzdělávání navazuje na obast Informatika v RPV pro základní školy. V cílovém zaměření vzdělávací oblasti se tedy nachází totožný bod:

> „uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka“ (NÚV, 2021a)

V oblasti Zdroje a vyhledávání informací, Komunikace je v učivu bod

„informační etika, legislativa – ochrana autorských práv a osobních údajů“ (NÚV, 2021a)

navazující na výstup

„využívá informační a komunikační služby v souladu s etickými, bezpečnostními a legislativními požadavky“ (NÚV, 2021a)

Alespoň částečně tedy jde o problematiku osobních údajů, ačkoli více z pohledu legislativního a etického než z pohledu soukromí.

V dalších oblastech se pak téma neobjevuje.

RVP odborné školy – Informační technologie

V cílech vzdělávání můžeme najít následující body:

„**neohrožovali svým chováním v digitálním prostředí sebe, druhé**, ani technologie samotné“ (MŠMT, b)

„**uvědomovali si, že technologie ovlivňují společnost, a naopak chápali svou odpovědnost při používání technologií**“ (MŠMT, b)

Ve výsledcích vzdělávání pak najdeme konkrétní body:

„**chrání** digitální zařízení, digitální obsah i **osobní údaje** v digitálním prostředí před poškozením, přepisem/změnou či **zneužitím**; reaguje na změny v technologiích ovlivňujících bezpečnost“ (MŠMT, b)

„s vědomím souvislostí fyzického a digitálního světa **vytváří a spravuje jednu či více digitálních identit; kontroluje svou digitální stopu, ať už ji vytváří sám nebo někdo jiný, v případě potřeby dokáže používat služby internetu anonymně**“ (MŠMT, b)

RVP odborné školy – Informační služby

V tomto RVP se žádné body přímo související s tématikou nenachází.

5.1.3 Návrh revizí RVP v oblasti Informatiky a informačních a komunikačních technologií

Aktuálně se pracuje na revizi RVP v oblasti Informatiky a informačních a komunikačních technologií, je tedy vhodné se podívat, zda tato revize nějak mění zakotvení tématu digitální stopy a osobních dat

ve vzdělávání.

Jedno ze základních východisek návrhu revize je rozvoj digitální gramotnosti, v dokumentu definované jako:

> „Digitální gramotností rozumíme soubor digitálních kompetencí (vědomostí, dovedností, postojů, hodnot), které jedinec potřebuje k bezpečnému, sebejistému, kritickému a tvořivému využívání digitálních technologií při práci, při učení, ve volném čase i při svém zapojení do společenského života.“ (NÚV, a)

V oblastech digitální gramotnosti může být pak pro naše téma relevantní bod:

> „Vnímá a hodnotí potenciál i rizika zapojení digitálních technologií do různých procesů a v různých situacích a podle toho zodpovědně jedná.“ (NÚV, a)

Jak se to promítá přímo do očekávaných výstupů můžeme vidět v tabulce níže:

Na všechny typech SŠ nacházíme výstup:

> „**chrání** digitální zřízení, digitální obsah i **osobní údaje** v digitálním prostředí před poškozením či zneužitím“ (NÚV, a)

A pro školy v kategoriích K, L, M, H dále

> „**kontroluje** svou **digitální stopu**, ať už ji vytváří sám nebo někdo jiný, dokáže používat služby internetu anonymně“ (NÚV, a)

a pro SŠ (E) podobný bod

> „buduje svou digitální identitu a zajímá se, jak k ní přispívají ostatní“ (NÚV, a)

Shrnutí

Téma osobních dat a digitální stopy můžeme v aktuálních vzdělávacích materiálech najít v obecném vymezení v cíli vzdělávací oblasti Informatika (respektive Informatika a informační a komunikační technologie) v bodě:

> „uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka“ (NÚV, 2021a)

Konkrétní vzdělávací výsledky s tímto tématem se nachází pouze u RVP odborných škol oboru Informační technologie.

Zároveň se však toto téma výrazněji objevuje v návrhu revize, a lze tedy předpokládat, že se do RVP (a následně ŠVP jednotlivých škol) bude více promítat.

Vznik materiálů a prostředí pro zařazení této problematiky do výuky je tedy možné považovat za užitečný a do budoucna nutný.

5.2 Východiska

Jak vyplývá z předchozí sekce, téma digitální stopy a osobních dat se nejspíše bude po revizích více objevovat v RVP středních škol. Na tuto skupinu studentů tedy bude Aplikace cílit.

Prostředí může sloužit jako součást většího vzdělávacího oblouku, ovšem tato práce si nedává za cíl vytvoření metodik pro učitele, a tedy musí obsah Aplikace fungovat i sám o sobě.

Cílem je uživatele (studenty) v Aplikaci seznámit s tématem interaktivní, lákavou formou.

Zároveň však součástí Aplikace musí být kvalitní napojení na teorii i reálné příklady, včetně informování uživatele o tom, jak může svou digitální stopu spravovat.

5.2.1 Návrh aplikace z pohledu uživatele

Návrh aplikace je následující: Uživatel má možnost odehrát jednotlivé „Mise“, které představují fiktivní příběh. V každé Misi má uživatel ze simulovaných osobních dat různého typu vyřešit na začátku danou otázku. Zjednodušený use-case tedy je:

- uživatel si vybírá Misi
- uživatel je seznámen s příběhem Mise (symbolickým rámcem) a cílem - co je potřeba zjistit pro splnění Mise
- uživatel využívá dané datasety pro nalezení řešení
- po správném zadání odpovědi jsou uživateli zobrazeny doplňující informace - napojení na teorii, příklady z reálného světa, možnosti zabezpečení se v podobné situaci.

V sekci *Scénáře* budou popsány konkrétní možné scénáře, z nich část bude použita v prototypu.

5.2.2 Prototyp

Jak bylo řečeno, cílem práce je vytvoření prototypu. Ten tedy nemusí mít všechny funkce či grafické řešení aplikace, která by byla reálně veřejně použita ve vzdělávání, má za cíl pouze najít vhodnou

podobu, otestovat její technickou proveditelnost a náročnost, a získat zpětnou vazbu od vybraných testerů.

Prototyp bude obsahovat několik scénářů s různými typy dat.

5.2.3 Scénáře

Jak bylo popsáno, jádrem Aplikace jsou Mise - tedy jednotlivé příběhy, ve kterých se uživatel seznamuje s různými situacemi, týkajícími se osobních dat.

Aplikace bude navržena tak, aby bylo možné snadno další scénáře přidávat, a tím rozšiřovat její vzdělávací potenciál (například v zacílení na jiné věkové skupiny).

Při tvorbě scénářů vycházím z dat a rizik popsaných v předchozích kapitolách a navrhuji následující testovací scénáře, které představují různé typy a kategorie dat.

Některé scénáře ukazují činnost, jež je ilegální. Motivací pro jejich zachování je to, že možnost prožít si situaci z pohledu útočníka může vést k lepšímu prožití a přenesení do uvažování o vlastní ochraně. Zároveň u každého takového scénáře bude upozorněno, že jde o simulaci a jaké trestněprávní dopady by taková činnost měla v reálném světě.

Odhadnutí hesla

Cíl
Uhodnout heslo blízké osoby na sociální síť.
Používané zdroje dat
Příspěvky na sociální síti.
Typ dat
Veřejné / veřejné pro okruh lidí.
Dodatečné informace
Informace o tom, jak lidé tvoří hesla
Případy toho, kdy lidé měli veřejně informace, jež vedly k prolomení hesla.
Obecná doporučení, jak chránit svoje hesla
Upozornění na nelegálnost takové činnosti v reálním životě

V tomto scénáři je úkolem uhodnout heslo do sociální sítě. Scénář se odkazuje na témata vhodného zabezpečení svých účtů a sicuací, kdy člověkem sdílené informace napomáhají k prolomení jeho obran.

Jde o jednoduchý, začáteční scénář – pracuje s veřejnými daty z jednoho zdroje. Je cíleně zjednodušený oproti realitě.

Scénář se nezabývá hesly z pohledu kryptografického a obecně bezpečnostního, neboť to je již mimo

rozsah naší práce. Bylo by však pravděpodobně možné takový scénář do Aplikace přidat pro vzdělávání v této oblasti.

Plánování vloupání

Cíl

Naplánovat vloupání dané osoby – udělat si představu o jejím bydlišti a době, kdy bude dům prázdný.

Používané zdroje dat

Příspěvky na sociální síti Příspěvky z fitness sociální sítě

Stránky firmy

Typ dat

Veřejná data

Dodatečné informace

Případy z praxe

Obecná doporučení, jak se tomuto typu útoku bránit

Upozornění na nelegálnost takové činnosti v reálním životě

Tento scénář již ukazuje práci s více zdroji dat a klade na uživatele větší nároky v hledání částí informací.

Vyšetřování korupce

Cíl

U podezřelé osoby vyšetřit možné korupční vazby na jiné osoby.

Používané zdroje dat

Lokační data od operátorů

Výpisy z bankovního účtu

Výpisy hovorů a SMS zpráv

Typ dat

Soukromé – dostupné pouze poskytovatelům a v oprávněných případech policii

Dodatečné informace

Informace o sběru a uchovávání dat operátory

Informace o přesnosti lokačních dat

Informace o právních možnostech policie si tato data vyžádat

Výběr vhodné reklamy

Cíl

Vybrat, jaké reklamy zobrazit jakým uživatelům.

Používané zdroje dat

Příspěvky na sociálních sítích

Chování na sociálních sítích

Historie prohlížení

Lokační data

Typ dat

Soukromé – data sbírají aplikace.

Dodatečné informace

Napojení na teorii *attention economy*

Informace o možnostech nastavení ochrany soukromí v různých aplikacích.

Tento scénář simuluje fungování vyhodnocování dat algoritmy firem jako je Facebook nebo Google, a následnou reklamní aukci. Má potenciál, aby na něj bylo navázáno šířeji tématem personalizované reklamy a *attention economy*.

Shrnutí

Tato kapitola poskytla východiska pro tvorbu aplikace, a propojila dosud získané poznatky z této oblasti s potřebami školství.

Kapitola 6

Praktická část - návrh aplikace a technické řešení

6.1 Technické řešení

6.1.1 Výběr technologií

Prototyp byl vytvořen jako webové aplikace. Hlavními důvody pro toto rozhodnutí bylo:

- dostupnost pro uživatele i testery
  Není potřeba nic instalovat, stačí jakýkoli moderní webový prohlížeč.
- snadné vytvoření prototypu
- snadná rozšiřitelnost a kooperace na projektu
  Jako open-source projekt předpokládá možnosti další spolupráce, například s designéry. Úprava vizuálu je v případě webové aplikace čistě HTML+CSS(+JS), což je rozšířená dovednost, oproti jiným grafickým prostředím (jako Qt či herní enginy).

Konkrétně byl zvolen webový framework **Flask** (založeného na jazyce Python). Motivací byla osobní zkušenost s tímto frameworkem, a tedy dostatečná představa o realizovatelnosti tohoto typu aplikace v daném prostředí.

Na frontendové straně nebyl použit žádný rozsáhlý framework a pro někde nutné části Javascriptu byly zvoleny pouze microframeworky **stimulus.js** a **jQuery**.

Vývoj je verzován systémem **git** za použití veřejného repozitáře na službě GitHub.

6.1.2 Základní funkcionality

Nahrávání předem definovaných dat Mise

Východiska

Mise musí být schopné používat předem definovaná data. Tato data mohou být v různé formě.
V budoucnu by aplikace mohla nabízet možnosti, jak přidávat další data misí.

Řešení

Zvažováno bylo několik řešení:

- data v databázi – například SQLite (která může být oproti jiným přímo součástí projektu)
- data přímo v kódu
- data v externím strukturovaném souboru

Varianta databáze byla pro účely prototypu vyřadazena z důvodu složitější úpravy a nahlížení dat. Z dlouhodobého hlediska se jeví jako vhodná forma strukturovaného dokumentu (např. JSON) – má výhody ve snadném přidávání dalších scénářů, a jasného oddělení dat a funkcionality (kódu). Vyžaduje však vytvoření kódu pro převod těchto dat do objektové struktury, kterou používá samotná aplikace. Proto byla tato varianta zařazena pro prototyp jako *nice-to-have*, a v první verzi prototypu jsou data přímo v kódu v kontroleru mise.

V případě velkého objemu dat by forma nahrávání s externího souboru nemusela být nejvhodnější, a bude možná nutné přejít na variantu s databází, či jinou. Díky použití ORM knihovny SQLAlchemy, a abstrahování logiky do modelové/objektové vrstvy, to ovšem nebude znamenat velký zásah do kódu.

Zobrazení misi s jednotlivými záložkami

Východiska

V náhledu mise je třeba uživateli zobrazit následující obsah:

- Úvod – představení mise
- Data – několik různých typů dat
- Místo pro řešení – stránku nebo stránky, kde uživatel řeší daný úkol - zadává informaci (např. heslo), vybírá z uvedených možností a podobně
- Závěr – stránku s přehledem použitých zdrojů a doplňujících informací.

Řešení

V prototypu je zobrazování vyřešeno nahráním všech datových zdrojů při načtení stránky a zobrazování/skrývání pomocí jednoduchého javascriptového (stimulus.js) kontroleru. Kontroler funguje obecně pro libovolný počet datových zdrojů. Toto řešení má oproti jiným variantám jednoduchý kód

(přehledná šablona a malý přehledný javascriptový kontroler). Potenciální riziko je v tom, že pokud budou datové zdroje rozsáhlejší, bude první nahrání stránky pomalé. V takové situaci by bylo vhodné data nenahrávat všechna při prvotním načtení stránky, ale přidat asynchronní získání dat pomocí Fetch API. To se však nijak nevylučuje s vytvořeným kontrolerem, který má na starost pouze přepínání viditelnosti jednotlivých částí stránky.

Zobrazovat data různých typů

Východiska

Na základě dat z kapitoly *Hlavní zdroje dat digitální stopy uživatele* byly definovány základní typy dat, které se v náhledech misí mohou objevovat:

- Příspěvky na sociálních sítích
- Výměna zpráv mezi dvěma osobami
- Výpis dat
  seznam hovorů, výpis z bankovního účtu,...
- Datové body na mapě
- Souhrnné informace
  například informace, které uchovávají sociální sítě o jednotlivých uživatelích
- webové stránky

Kromě toho ale lze očekávat, že se budou v budoucnu objevovat další typy dat, je tedy důležité, aby bylo snadné tuto nabídku rozšiřovat, a aby bylo řešení dostatečně obecné (nebo zobecnitelné), aby bylo toto přidávání snadné.

Řešení

Pro jednotlivé typy dat byly vytvořeny samostatné modely, kontrolery a sadu šablon. Pro ukázku je zde struktura kódu pro zobrazování dat typu sociální síť:
Kontroler umožňuje zobrazovat několik základních stránek - osobní profil, sadu příspěvků (feed) a přihlašovací stránku. Zobrazování sady příspěvků je realizováno pomocí několik *partial* šablon - feed, post, comment. Jednotlivé šablony je pak snadné například designově upravovat, a zůstávají velice přehledné.

U většiny typů dat bylo třeba si vytvořit vlastní HTML+CSS kód, který nabízí očekávanou strukturu dat v přehledné podobě. Pro zobrazování lokačních dat byla využita javascriptová knihovna SMapy.

6.1.3 Další funkcionality

Mezi funkcionality, které nejsou potřebné pro fázi prototypu, ale je dobré na ně myslet při strukturování celé aplikace, patří:

- **Uživatelský účet**
  Aplikace by měla nabízet uchovávání informací o stavu jednotlivých misí, aby mohl uživatel navázat tam, kde přestal. Může to být řešeno uživatelským účtem s registrací, nebo identifikací pomocí cookies.
- **Přidávání dalších scénářů**
  Aplikace do budoucna počítá s možností vlastních scénářů. Systém zadávání zdrojových dat a jejich nahrávání tedy musí být možné vystavit ven. Varianty přidávání můžou být různě technicky složité, to bude záležet na další analýze používání aplikace v praxi.

6.1.4 Návrh aplikace

Aplikace má architekturu MVT (Model-View-Template), inspirovanou návrhem jiného webového frameworku v jazyce Python, a to Django.

Model

Aplikace plně využívá objektový návrh, a se všemi zdrojovými daty nakládá jako s objekty.
Jsou tedy definované modely/třídy jako `Facebook user`, `Facebook post`, `Location point` a další.
Kromě toho je objektem i každá Mise, která obsahuje jednotlivé `Mission Items`.

View

Vrstva View řeší získávání a zpracování dat, vykreslení šablony s těmito daty a navázání na konkrétní cesty (*route*).
V aplikaci je použito rozšíření `Flask-Classful`, které usnadňuje práci s views, například snadným vytvořením základních `CRUD` operací.

Template

Flask v základu využívá templatovací jazyk `Jinja2`, který nabízí omezené množství logiky v šabloně, a vede k tomu, aby většina aplikační logiky zůstávala mimo šablonu (tedy ve View).

6.1.5 Ověření a otestování aplikace

Při ověřování funkčnosti aplikace s reálnými uživateli budou zejména kontrolovány tyto oblasti:

- **Orientace v prostředí** Bude ověřováno, zda je uživateli dostatečně jasné, co se po něm chce, a dokáže v prostředí aplikace najít vše co potřebuje pro interakci s ním.

- **Používání datových zdrojů** S ohledem na to, že nahlížení datových zdrojů je v uživatelské interakci s aplikací hlavní činností, bude ověřováno, zda je jejich používání jasné, srozumitelné a umožňuje plnění misí.
- **Náročnost úkolů** Další testovanou oblastí bude náročnost úkolů v prezentovaných misích. Toto testování musí probíhat u každé mise, jež bude do aplikace přidána. Bude ověřováno, zda dokáže uživatel z dodaných informací vymyslet postup práce, zda má všechna potřebná data a naopak jestli není úkol příliš jednoduchý.
- **Míra zaujetí** Nakonec bude získán pohled uživatelů na to, zda je tato forma seznámení s tématikou zajímavá a poutavá.

Ze všech kontrolovaných oblastí lze očekávat výstupy, které poslouží k posunutí aplikace z fáze prototypu do produkční verze.

Závěr

Lorem Ipsum

Literatura

Identity theft rises sharply as fraudsters target social media. *Computer Fraud & Security*. 2016, 2016, 7, s. 1–3. ISSN 1361-3723. doi: https://doi.org/10.1016/S1361-3723(16)30048-3. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1361372316300483.

Azucar, D. – Marengo, D. – Settanni, M. Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*. 2018, 124, s. 150–159. ISSN 0191-8869. doi: https://doi.org/10.1016/j.paid.2017.12.018. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0191886917307328.

BBC. *Cambridge Analytica 'not involved' in Brexit referendum, says watchdog* [online]. 2020. [cit. 2021-04-02]. Dostupné z: https://www.bbc.com/news/uk-politics-54457407.

Beckers, K. et al. A Structured Comparison of Social Engineering Intelligence Gathering Tools. s. 232–246, 08 2017. doi: 10.1007/978-3-319-64483-7_15. ISBN 978-3-319-64482-0.

Brandom, R. *GLOBAL PRIVACY CONTROL WANTS TO SUCCEED WHERE DO NOT TRACK FAILED* [online]. 2021. [cit. 2021-04-14]. Dostupné z: https://www.theverge.com/2021/1/28/22252935/global-privacy-control-personal-data-tracking-ccpa-cpra-gdpr-duckduckgo.

Bryce, J. – Klang, M. Young people, disclosure of personal information and online privacy: Control, choice and consequences. *Information Security Technical Report*. 2009, 14, 3, s. 160–166. ISSN 1363-4127. doi: https://doi.org/10.1016/j.istr.2009.10.007. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1363412709000429. The Changing Shape of Privacy and Consent.

Cadwalladr, C. – Graham-Harrison, E. *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach* [online]. 2018. [cit. 2021-03-15]. Dostupné z: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Choi, H. – Park, J. – Jung, Y. The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*. 2018, 81, s. 42–51. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2017.12.001. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0747563217306817.

Council of European Union. Council regulation (EU) no 2016/679, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

Coutts, L. V. et al. Deep learning with wearable based heart rate variability for prediction of mental and general health. *Journal of Biomedical Informatics*. 2020, 112, s. 103610. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2020.103610. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1532046420302380.

Doffman, Z. *Ashley Madison Hack Returns To 'Haunt' Its Victims: 32 Million Users Now Watch And Wait* [online]. 2020. Dostupné z: https://www.forbes.com/sites/zakdoffman/2020/02/01/ashley-madison-hack-returns-to-haunt-its-victims-32-million-users-now-have-to-watch-and-wait/.

Editors. *US Digital Ad Spending Will Surpass Traditional in 2019* [online]. 2019. [cit. 2021-03-15]. Dostupné z: https://www.emarketer.com/content/us-digital-ad-spending-will-surpass-traditional-in-2019.

Fish, T. *My digital footprint: a two sided digital business model where your privacy will be someone else's business*. Futuretext, 2009. OCLC: 639933238. ISBN 9780955606984.

Foster, M. *Is The Way We Use Social Media Leaving Us Vulnerable To Burglary?* [online]. 2015. [cit. 2021-04-03]. Dostupné z: https://socialmediaweek.org/blog/2015/05/social-media-leaving-us-vulnerable-burglary/.

Foucault, M. *Power/knowledge: Selected interviews and other writings, 1972-1977*. Vintage, 1980.

Frenkel, S. *Facebook Bug Changed Privacy Settings of Up to 14 Million Users* [online]. 2018. [cit. 2021-04-02]. Dostupné z: https://www.nytimes.com/2018/06/07/technology/facebook-privacy-bug.html.

Goodin, D. *New browser-tracking hack works even when you flush caches or go incognito* [online]. 2021. [cit. 2021-04-14]. Dostupné z: https://arstechnica.com/information-technology/2021/02/new-browser-tracking-hack-works-even-when-you-flush-caches-or-go-incognito/.

Hadnagy, C. *Social engineering: the art of human hacking*. Wiley, 2011. OCLC: ocn635494717. ISBN 9780470639535 9781118028018 9781118029718 9781118029749.

Hern, A. *Fitness tracking app Strava gives away location of secret US army bases* [online]. 2018. [cit. 2021-04-02]. Dostupné z: https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

Hinds, J. – Williams, E. J. – Joinson, A. N. "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*. 2020, 143, s. 102498. ISSN 1071-5819. doi:

https://doi.org/10.1016/j.ijhcs.2020.102498. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1071581920301002.

Ikuesan, A. R. – Venter, H. S. Digital behavioral-fingerprint for user attribution in digital forensics: Are we there yet? *Digital Investigation*. 2019, 30, s. 73–89. ISSN 1742-2876. doi: https://doi.org/10.1016/j.diin.2019.07.003. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1742287619300945.

Jorgensen, Z. – Yu, T. On Mouse Dynamics as a Behavioral Biometric for Authentication. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '11, s. 476–482, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/1966913.1966983. Dostupné z: https://doi.org/10.1145/1966913.1966983. ISBN 9781450305648.

Juna, P. – Burýšek, J. *Ruský web pro pedofily vystavuje kradené fotky českých dětí* [online]. 2020. [cit. 2021-04-02]. Dostupné z: https://www.seznamzpravy.cz/clanek/rusky-web-pro-pedofily-vystavuje-kradene-fotky-ceskych-deti-89197.

Kemp, S. *Digital 2021: Global overview report* [online]. 2019. [cit. 2021-03-15]. Dostupné z: https://datareportal.com/reports/digital-2021-global-overview-report.

Livingstone, S. Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media & Society*. 2008, 10, 3, s. 393–411. doi: 10.1177/1461444808089415. Dostupné z: https://doi.org/10.1177/1461444808089415.

Madden, M. et al. *Digital Footprints - Online identity management and search in the age of transparency* [online]. 2007. [cit. 2021-04-02]. Dostupné z: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2007/PIP_Digital_Footprints.pdf.pdf.

Marwick, A. E. – boyd. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*. 2014, 16, 7, s. 1051–1067. doi: 10.1177/1461444814543995. Dostupné z: https://doi.org/10.1177/1461444814543995.

Morgan, R. *Uber settles federal probe over 'God View' spy software* [online]. 2017. [cit. 2021-04-02]. Dostupné z: https://nypost.com/2017/08/15/uber-settles-federal-probe-over-god-view-spy-software/.

MŠMT. *Publicistika, knihovnictví a informatika* [online]. a. [cit. 2021-03-15]. Dostupné z: https://www.edu.cz/rvp/ramcove-vzdelavaci-programy-stredniho-odborneho-vzdelavani-rvp-sov/obory-l-a-m/72-publicistika-knihovnictvi-a-informatika/.

MŠMT. *Informatické obory* [online]. b. [cit. 2021-03-15]. Dostupné z: https://www.edu.cz/rvp/ramcove-vzdelavaci-programy-stredniho-odborneho-vzdelavani-rvp-sov/obory-l-a-m/18-informaticke-obory/.

Neil, A. *Google AI Can Work Out Photo Locations Without Geotags* [online]. 2016. [cit. 2021-04-02]. Dostupné z: https://www.eteknix.com/google-ai-can-work-photo-locations-without-geotags/.

NÚV. *Návrh revizí ICT* [online]. a. [cit. 2021-03-15]. Dostupné z: http://www.nuv.cz/file/3362/.

NÚV. *Rámcové programy* [online]. b. [cit. 2021-03-15]. Dostupné z: http://www.nuv.cz/t/rvp.

NÚV. *RVP-G* [online]. 2021a. [cit. 2021-03-15]. Dostupné z: http://www.nuv.cz/file/159.

NÚV. *RVP ZŠ* [online]. 2021b. [cit. 2021-03-15]. Dostupné z: http://www.nuv.cz/file/4983/.

NÚV. *Střední vzdělávání* [online]. c. [cit. 2021-03-15]. Dostupné z: http://www.nuv.cz/t/stredni-vzdelavani.

Penland, J. *Browser Cookies: What Are They & Why Should You Care?* [online]. 2020. [cit. 2021-03-15]. Dostupné z: https://www.whoishostingthis.com/resources/cookies-guide/#:~:text=Cookies.

Pilton, C. – Faily, S. – Henriksen-Bulmer, J. Evaluating privacy - determining user privacy expectations on the web. *Computers & Security*. 2021, 105, s. 102241. ISSN 0167-4048. doi: https://doi.org/10.1016/j.cose.2021.102241. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0167404821000651.

Rathi, R. *Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential Election* [online]. 2019. [cit. 2021-03-15]. Dostupné z: https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d.

Research, P. *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information* [online]. 2019. [cit. 2021-03-15]. Dostupné z: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-inform

Scott, T. *Why is Internet such a mess* [online]. 2020. [cit. 2021-03-15]. Dostupné z: https://www.youtube.com/watch?v=OFRjZtYs3wY.

Security, R. B. *New Research: No. of Records Exposed Increased 141% in 2020* [online]. 2021. [cit. 2021-04-02]. Dostupné z: https://www.riskbasedsecurity.com/2021/01/21/new-research-no-of-records-exposed-increased-141-in-2020/#download.

Services, T. N. *AP investigation: Across U.S., police officers abuse confidential databases* [online]. 2016. [cit. 2021-04-02]. Dostupné z: https://www.chicagotribune.com/nation-world/ct-ap-police-database-abuse-20160928-story.html.

STATCOUNTER. *Search Engine Market Share Worldwide* [online]. 2021. [cit. 2021-03-15]. Dostupné z: https://gs.statcounter.com/search-engine-market-share.

SÆTRA, H. S. Privacy as an aggregate public good. *Technology in Society*. 2020, 63, s. 101422. ISSN 0160-791X. doi: https://doi.org/10.1016/j.techsoc.2020.101422. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0160791X20310381.

TANKOVSKA, H. *Number of monthly active Facebook users worldwide as of 4th quarter 2020* [online]. 2020. [cit. 2021-03-15]. Dostupné z: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

THOMPSON, S. A. – WARZEL, C. *Twelve Million Phones, One Dataset, Zero Privacy* [online]. 2019. [cit. 2021-04-05]. Dostupné z: https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.

VALÁŠEK, L. – HORÁK, J. *Policisté, politici i jejich děti. Čínský armádní dodavatel sbírá data stovek Čechů* [online]. 2020. [cit. 2021-03-15]. Dostupné z: https://zpravy.aktualne.cz/domaci/policiste-politici-i-jejich-rodina-cinsky-armadni-dodavatel/r~e3e18dfcf6c211ea842f0cc47ab5f122/.

VOKUŠ, P. J. *Poskytovatelé internetového připojení* [online]. 2019. [cit. 2021-04-02]. Dostupné z: https://www.policie.cz/clanek/poskytovatele-internetoveho-pripojeni.aspx.

WHITNEY, L. *How a successful phishing attack can hurt your organization* [online]. 2021. [cit. 2021-04-14]. Dostupné z: https://www.techrepublic.com/article/how-a-successful-phishing-attack-can-hurt-your-organization/#:~:text=Some57%25saidtheirorganization,andunsuccessful--in2020.

YANG, Y. C. Web user behavioral profiling for user identification. *Decision Support Systems*. 2010, 49, 3, s. 261–271. ISSN 0167-9236. doi: https://doi.org/10.1016/j.dss.2010.03.001. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0167923610000497.