

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



**Diplomová práce**

**Tvorba webových stránek  
podmíněných legislativními povinnostmi,  
webové stránky obcí**

**Bc. Vít Havránek**

**© 2023 ČZU v Praze**



## ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Vít Havránek

Informatika

Název práce

**Tvorba webových stránek podmíněných legislativními povinnostmi, webové stránky obcí**

Název anglicky

**Creation of websites subject to legislative obligations, websites of municipalities**

---

### Cíle práce

Diplomová práce je zaměřena na problematiku tvorby webových stránek pro obce, které jsou specifické legislativními nároky, které musí splňovat. Hlavním cílem je analýza současného stavu webových stránek a to především v plnění legislativních povinností.

Z hlavního cíle vychází následující dílčí cíle:

- vytvoření souhrnu teoretických principů požadované legislativy a ostatních náležitostí povinných i doporučených při tvorbě webových stránek pro obce,
- provedení analýzy reálných webových stránek obcí, skrze vhodně navržený algoritmus
- provedení uživatelského testování v rámci ověření strojového vyhodnocení splnění kritérií, které musí webové stránky splňovat,
- na základě provedené analýzy, stanovit rizikové body, kde dochází k porušení legislativních povinností či jiných standardů a následné stanovení doporučení pro eliminaci těchto rizik.

### Metodika

Metodika diplomové práce bude založena na studiu a analýze dostupných informačních zdrojů, existujících řešení v oblasti tvorby webových stránek pro obce. V rámci teoretické části budou zpracovány jednotlivé aspekty tvorby webových stránek od legislativně daných až po standardy, které by každý současný web měl dodržovat v rámci aktuálních trendů, například přístupnost webu, úřední deska, povinně zveřejňované údaje. Následně bude náhodně vybráno vhodně zvolené množství webových stránek obcí pro následnou komparaci, která bude realizována skrze algoritmus, který vyhodnotí splnění legislativních nároků webových stránek a dalších standardů. Dále bude provedeno uživatelské testování pro ověření strojem generovaných výsledků.

Následně bude provedena analýza rizik, stanovení nejvíce rizikových míst, kde dochází k porušování legislativních nařízení a doporučených standardů. Výstupem budou dedukovaná doporučení mající za cíl eliminovat tyto rizika, jak ze strany tvůrce, tak ze strany následných uživatelů. Z výsledků analýzy bude vycházet závěr práce.

## Doporučený rozsah práce

60 – 80 stran

## Klíčová slova

Legislativa, webové stránky, algoritmus, použitelnost, přístupnost, obce

---

## Doporučené zdroje informací

Akgul, Yakup. App and Website Accessibility Developments and Compliance Strategies. Engineering Science Reference (an Imprint of IGI Global), 2022. ISBN 9781799878490.

BUREŠ, Miroslav, Miroslav RENDA, Michal DOLEŽEL, Peter SVOBODA, Zdeněk GRÖSSL, Martin KOMÁREK, Ondřej MACEK a Radoslav MLYNÁŘ. Efektivní testování softwaru: klíčové otázky pro efektivitu testovacího procesu. Praha: Grada, 2016. Profesionál. ISBN 978-80-247-5594-6.

GASSTON, Peter. Moderní web. Přeložil Ondřej BAŠE. Brno: Computer Press, 2015. ISBN 978-80-251-4345-2.

JANEČKOVÁ, Eva. GDPR: řešení problémů v praxi obcí. Praha: Grada Publishing, 2019. Právo pro praxi. ISBN 978-80-247-2925-1.

---

## Předběžný termín obhajoby

2022/23 LS – PEF

## Vedoucí práce

Ing. Václav Lohr, Ph.D.

## Garantující pracoviště

Katedra informačních technologií

Elektronicky schváleno dne 27. 9. 2022

**doc. Ing. Jiří Vaněk, Ph.D.**

Vedoucí katedry

Elektronicky schváleno dne 28. 11. 2022

**doc. Ing. Tomáš Šubrt, Ph.D.**

Děkan

V Praze dne 30. 11. 2023



### **Čestné prohlášení**

Prohlašuji, že svou diplomovou práci "Tvorba webových stránek podmíněných legislativními povinnostmi, webové stránky obcí" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 30. 11. 2023 \_\_\_\_\_

## **Poděkování**

Rád bych touto cestou poděkoval vedoucímu diplomové Ing. Václavu Lohrovi, Ph.D. za vedení mé diplomové práce a odbornou pomoc, cenné rady a za jeho čas, který mi věnoval pro potřeby zpracování diplomové práce.

# Tvorba webových stránek podmíněných legislativními povinnostmi, webové stránky obcí

## Abstrakt

Diplomová práce se zabývá problematikou legislativních povinností, které musí být splněny v rámci realizace webových stránek, především jejího obsahu. V teoretické části jsou zpracovány základní informace jednotlivých legislativních požadavků. Další části jsou věnovány metodám zpracování dat, v rámci které dále rozvedena technika web scrapingu. Poslední část teoretické části je věnována uživatelskému testování. Praktická část se zabývá zpracováním legislativních požadavků do testovatelné podoby, která je aplikována pomocí principů web scrapingu na reálné webové stránky pro účely ověření jejich splnění. Následně je provedena analýza získaných dat a stanovení výsledků zkoumání. Pro ověření výsledků je provedeno uživatelské testování. Následně provedená analýza výsledků testování v porovnání s výsledky předchozího zkoumání. Výsledkem diplomové práce je zjištění současné úrovně plnění legislativních požadavků v rámci testovaných webů a dedikovaná doporučení, jak nalezené nedostatky eliminovat a dále jim předcházet.

**Klíčová slova:** Legislativa, webové stránky, algoritmus, použitelnost, přístupnost, obce

# **Creation of websites subject to legislative obligations, municipal websites**

## **Abstract**

The diploma thesis deals with the issue of legislative obligations that must be met in the implementation of the website, especially its content. In the theoretical part the basic information of the individual legislative requirements is elaborated. Other parts are devoted to the methods of data processing, within which the technique of web scraping is further developed. The last part of the theoretical part is devoted to user testing. The practical part deals with the processing of the legislative requirements into a testable form, which is applied using the principles of web scraping to real websites to verify their compliance. Subsequently, the analysis of the obtained data and the determination of the results of the examination are carried out. User testing is performed to verify the results. Subsequently, the analysis of the testing results compared with the results of the previous exploration. The result of the thesis is the determination of the current level of compliance with legislative requirements within the tested websites and dedicated recommendations on how to eliminate and further prevent the identified deficiencies

**Keywords:** Legislation, websites, algorithm, usability, accessibility, municipalities

# Obsah

<b>1 Úvod.....</b>	<b>9</b>
<b>2 Cíl práce a metodika .....</b>	<b>10</b>
2.1 Cíl práce .....	10
2.2 Metodika .....	10
<b>3 Teoretická východiska .....</b>	<b>11</b>
3.1 Legislativní nároky na webové stránky.....	11
3.1.1 Přístupnost webových stránek .....	11
3.1.1.1 Zákon č. 99/2019 Sb., o přístupnosti internetových stránek .....	11
3.1.1.2 WCAG .....	12
3.1.1.3 Prohlášení o přístupnosti webových stránek .....	14
3.1.2 Zákon č. 106/1999 Sb. o svobodném přístupu k informacím .....	14
3.1.2.1 Povinně zveřejňované informace .....	16
3.1.2.2 Úřední deska.....	16
3.1.3 Obecné nařízení o ochraně osobních údajů .....	17
3.1.4 Zákon č. 171/2023 Sb. o ochraně oznamovatelů .....	18
3.2 Web scraping.....	19
3.2.1 Metody web scrapingu.....	20
3.2.1.1 HTML parsing .....	20
3.2.1.2 DOM parsing .....	21
3.2.1.3 API Scraping .....	21
3.2.1.4 Headless Browser Scraping.....	22
3.2.1.5 Aplikace a nástroje pro webscraping.....	22
3.2.2 Právní aspekt.....	23
3.2.3 Autorská práva .....	23
3.3 Zpracování dat.....	24
3.3.1 Získávání dat.....	24
3.3.2 Ukládání dat.....	25
3.3.3 Čištění dat .....	25
3.3.4 Transformace dat .....	25
3.3.5 Analýza dat .....	26
3.3.6 Prezentace dat .....	27
3.4 Uživatelské testování .....	27
3.4.1 Postup uživatelského testování .....	28
3.4.2 Metody uživatelského testování.....	29

3.4.2.1	Laboratorní testování .....	29
3.4.2.2	Vzdálené testování .....	30
3.4.2.3	Task analysis .....	30
3.4.2.4	Metody s využitím strojového měření .....	31
<b>4</b>	<b>Praktická část .....</b>	<b>33</b>
4.1	Stanovení požadavků .....	33
4.1.1	Plně definované .....	33
4.1.2	Částečně definované .....	34
4.1.3	Definice na základě zdrojového kódu .....	35
4.1.4	Metody ověření požadavků .....	35
4.2	Sběr a zpracování dat .....	35
4.2.1	Řešení obsahu skrze klíčové výrazy .....	37
4.2.2	Řešení přístupnosti na základě ověřování kódu .....	40
4.2.3	Řešení přístupnosti z hlediska souborů .....	42
4.2.4	Specifické úlohy .....	45
4.3	Analýza .....	47
4.3.1	Přítomnost povinného obsahu .....	47
4.3.2	Povinně zveřejňované informace .....	49
4.3.3	Prohlášení přístupnosti webových stránek .....	50
4.3.4	Přístupnost grafického obsahu .....	52
4.3.5	Přístupnost souborů .....	54
4.3.6	Formuláře .....	55
4.4	Uživatelské testování .....	56
4.4.1	Dotazník na tvorbu obsahu .....	57
4.4.1.1	Struktura dotazníku .....	57
4.4.1.2	Vyhodnocení dotazníku .....	58
4.4.2	Testování vkládání příspěvků .....	62
4.4.3	Prototyp .....	62
4.4.4	Testování hledání obsahu .....	63
4.4.5	Průběh testování a zvolené metody .....	63
4.4.6	Analýza získaných dat .....	64
4.4.7	Výsledky testování .....	64
<b>5</b>	<b>Zhodnocení výsledků a doporučení .....</b>	<b>66</b>
5.1	Výsledky analytické části .....	66
5.2	Výsledky uživatelského testování .....	67
5.3	Dedukované shrnutí výsledků .....	68
5.4	Doporučené postupy pro eliminaci nedostatků .....	69
5.4.1	Obecná doporučení .....	69

5.4.2	Konkrétní doporučení .....	70
5.5	Zhodnocení použitých metod .....	71
<b>6</b>	<b>Závěr.....</b>	<b>72</b>
<b>7</b>	<b>Seznam použitých zdrojů .....</b>	<b>75</b>
<b>8</b>	<b>Přílohy .....</b>	<b>80</b>
8.1	Příloha 1 - diagram předpokládaného řešení aktuality .....	81
8.2	Příloha 2 - Diagram předpokládaného řešení přidání nového oznámení .....	82
8.3	Příloha 3 - Testovací scénář .....	83
8.4	Příloha 4a – Dotazník.....	84
8.5	Příloha 4b – dotazník .....	85

## Seznam obrázků

Obrázek 1	Proces webscrapingu .....	20
Obrázek 2	Wireframe testovacího prototypu .....	63

## Seznam zdrojových kódů

Zdrojový kód 1	Extrakce obsahu webové stránky .....	36
Zdrojový kód 2	Ukládání dat CSV souboru .....	36
Zdrojový kód 3	Úprava získaného obsahu .....	38
Zdrojový kód 4	Funkce ověřování použití klíčových slov .....	38
Zdrojový kód 5	Funkce provádějící kompletní analýzu .....	39
Zdrojový kód 6	Funkce provádějící analýzu grafického obsahu .....	41
Zdrojový kód 7	Funkce provádějící analýzu přístupnosti PDF souborů .....	44
Zdrojový kód 8	Funkce provádějící analýzu formuláře.....	46

## Seznam grafů

Graf 1	Ověření přítomnosti povinných stránek .....	48
Graf 2	Nedostupné povinné stránky .....	49
Graf 3	Vyhodnocení plnění kritérií povinně zveřejňovaných informací .....	50
Graf 4	Vyhodnocení plnění kritérií prohlášení přístupnosti .....	51
Graf 5	Nejčastěji nesplněné klíčové výrazy v prohlášení přístupnosti .....	52
Graf 6	Vyhodnocení míry přístupnosti grafického obsahu.....	53
Graf 7	Vyhodnocení míry přístupnosti souborů .....	55
Graf 8	Vyhodnocení ošetření formulářů v rámci GDPR .....	56
Graf 9	Autor prohlášení přístupnosti .....	58

Graf 10 Editor prohlášení přístupnosti .....	59
Graf 11 Autor obsahu prohlášení GDPR .....	59
Graf 12 Editor obsahu prohlášení GDPR.....	60
Graf 13 Autor obsahu povinně zveřejňovaných informací .....	60
Graf 14 Editor obsahu povinně zveřejňovaných informací .....	61
Graf 15 Způsob přidávání//vytváření rutinního obsahu na webové stránky .....	61



# 1 Úvod

Webové stránky jsou v současné době jedním z nejpoužívanějších zdrojů informací. Slouží nejen k zabavení a edukaci, ale roste i jejich význam na poli veřejné správy. Díky úrovni dnešních technologií odpadá nutnost cestování na přílehlý úřad při potřebě získání jakýkoliv informací v rámci působnosti obce. Veškeré tyto informace by měly být zahrnuty v rámci webové prezentace obce. Avšak to, že obec disponuje webovou prezentací ještě neznamená, že je web kvalitně zpracován a všechny informace jsou snadno dohledatelné a ve správné podobě. V tuto chvíli na řadu přichází legislativní požadavky a nároky, jejichž úkolem je zajištění toho, aby všechny podstatné informace byly snadno dohledatelné a v použitelné podobě. Mimo to mají zajistit transparentnost informací a minimalizaci bariér v jejich získávání. Jednotlivé požadavky mají různorodý rozsah a vliv na samotnou realizaci webových stránek, avšak v kompletnosti toho všeho plní vesměs jednotný cíl, kterým je zajištění dostupných informací a pro jakéhokoliv uživatele bez ohledu na jeho případný handicap.

Motivací této diplomové práce je ověřit, zda obce skutečně plní tyto požadavky, které jsou dány legislativou, případně v jaké míře. Ač webové stránky obcí netrhají návštěvnické rekordy, tak jsou důležitým nástrojem jak obcím pro publikování veškerých informací, které se dané obce týkají, tak i běžným občanům pro zjištění důležitých informací. Kvalitní webové stránky, které splňují legislativní požadavky mají vysoký potenciál zpříjemnit život občanům a úředníkům ulehčit práci v případě zodpovídání dotazů, na které by měly být odpovědi snadno dohledatelné skrze webové stránky. Důležitou součástí takto komplexně pojaté problematiky není pouze analyzování úrovně plnění požadavků, ale také skrze dedukci poskytnou doporučení, která by mohla zjištěné nedostatky eliminovat.

## **2 Cíl práce a metodika**

### **2.1 Cíl práce**

Diplomová práce je zaměřena na problematiku tvorby webových stránek pro obce, které jsou specifické legislativními nároky, které musí splňovat. Hlavním cílem je analýza současného stavu webových stránek a to především v plnění legislativních povinností.

Z hlavního cíle vychází následující dílčí cíle:

- a. vytvoření souhrnu teoretických principů požadované legislativy a ostatních náležitostí povinných i doporučovaných při tvorbě webových stránek pro obce,
- b. provedení analýzy reálných webových stránek obcí, skrze vhodně navržený algoritmus
- c. provedení uživatelského testování v rámci ověření strojového vyhodnocení splnění kritérií, které musí webové stránky splňovat,
- d. na základě provedené analýzy, stanovit rizikové body, kde dochází k porušení legislativních povinností či jiných standardů a následné stanovení doporučení pro eliminaci těchto rizik.

### **2.2 Metodika**

Metodika diplomové práce bude založena na studiu a analýze současných legislativních požadavků kladených na správu obsah webových stránek obcí. Následovat bude studium vhodných metod pro algoritmičké zpracování procesu ověřování plnění legislativních požadavků a dále uživatelského testování. Teoretická východiska budou zajištěna studiem odborné literatury, článků a dalších dostupných zdrojů s důrazem na jejich aktuálnost a odbornost.

V praktické části bude zpracováno algoritmičké řešení, které bude strojově analyzovat a vyhodnocovat plnění legislativních požadavků skrze vhodně zvolená kritéria s ohledem na zjištěné poznatky z teoretické části. Řešení bude aplikováno na náhodně zvolené webové stránky, kdy bude provedena následná analýza výsledků v kombinaci s komparací za účelem zjištění současného stavu plnění požadavků napříč různými webovými stránkami. Následně bude zpracováno uživatelské testování s důrazem na ověření výsledků získaných v analytické části skrze navržený algoritmus.

Na základě získaných výsledků budou synteticky sestaveny rizikové body, ve který dochází k nedodržování legislativních požadavků. Ve finální části budou stavena dedukovaná doporučení, jejichž využitím a dodržováním by mělo vést k minimalizaci chybovosti eliminaci případů porušování legislativních požadavků.

## **3 Teoretická východiska**

### **3.1 Legislativní nároky na webové stránky**

#### **3.1.1 Přístupnost webových stránek**

V rámci problematiky přístupnosti webových stránek vystupují převážně hendikepovaní uživatelé. Typickým zástupcem této skupiny je člověk se zrakovým postižením, který k prohlížení webových stránek využívá aplikaci třetí strany pro předčítání webového obsahu. Vzhledem k tomu, že informace veřejné správy musí být všem snadno dostupné a dosažitelné, je předmětné a důležité se zabývat tvorbou webových stránek tak, aby jejich obsah byl snad dosažitelný i pro tyto uživatele. (Pavlíček, 2009; Vránková, 2023)

Nutno zmínit, že ve skupině hendikepovaných uživatelů nejsou pouze ti, kteří jsou odkázáni na využívání aplikací pro předčítání, ale tato skupina zahrnuje i částečně hendikepované uživatele, kterých se týkat například barvoslepost, proto je důležité pracovat nejen s obsahem, ale i barevným kontrastem. (Vránková, 2023)

Přístupné webové stránky lze chápat jako takové, které nekladou na uživatele žádné bariéry, jejich ovládání je přívětivé, obsah je snadno srozumitelný a dosažitelný z jakéhokoliv typu zařízení. Mimo klasické počítače a chytré telefony zde vystupují tzv. asistivní technologie, které pomáhají hendikepovanými uživatelům s nejrůznějším typem postižení. V rámci problematiky přístupnosti se nejčastěji jedná o zrakově postižené. Nejčastějším příkladem asistivních technologií jsou softwarové čtečky, které převádějí textový obsah webových stránek na zvukový výstup neboli předčítají obsah. (Pavlíček, 2009; Vránková, 2023)

##### **3.1.1.1 Zákon č. 99/2019 Sb., o přístupnosti internetových stránek**

V České legislativě se touto problematikou zabývá Zákon č. 99/2019 Sb., o přístupnosti internetových stránek a mobilních aplikací a o změně zákona č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů, ve znění pozdějších předpisů. Tento zákon vychází ze směrnice Evropského parlamentu a Rady (EU) 2016/2102 o přístupnosti webových internetových stránek a mobilních aplikací subjektů veřejného sektoru. (Vránková, 2023)

Zákon vychází především z metodiky WCAG zabývající se kritérii, která by měla být při tvorbě přístupných webových stránek naplněna a udává pravidla a principy pro jejich dosažení. Dále zákon udává tzv. povinné subjekty. Jedná se o výčet subjektů, na které se zákon vztahuje. Největší skupinu tvoří územní samosprávný celek neboli obce a školy. V případě, že povinné subjekty nejsou schopny zcela splnit všechny požadavky přístupnosti, ať z důvodu finanční či časové náročnosti, mohou uplatnit výjimku v rámci tzv. nepřiměřené zátěže. Zákon vymezuje jasná pravidla pro uplatnění této výjimky. Výjimka nemůže být vztahena na celé webové stránky, ale pouze na konkrétní část. Dále musí být dostatečně zdůvodněna a musí být zajištěno náhradní řešení. Základní myšlenkou je, že pro uživatele musí být informace stále dosažitelné. Drobným příkladem může být vložení informačního letáku o nadcházející akci, který je strojově nečitelný. Jasnou alternativou je vložení nečitelného letáku, doplněného o textové shrnutí informací na něm obsažených. (Vránková; Zákon č. 99/2019 Sb.)

### 3.1.1.2 WCAG

Web Content Accessibility Guidelines je metodika zpravována mezinárodním konsorciem W3C neboli World Wide Web Consortium, která se zabývá tvorbou webových standardů. Jedná se o nejuznávanější a nejpoužívanější metodiku v rámci přístupnosti, od které jsou často odvozovány interní metodiky či zákonné povinnosti. V současné chvíli je aktuální verze WCAG 2.1, která byla vydána roku 2018. Hlavní důraz je kladen především na koncového uživatele tak, aby finální web umožňoval pohodlné užívání a snadnou dosažitelnost informací bez zbytečných bariér. (W3C konsorcium 2018)

Metodika definuje 4 základní principy přístupnosti, které obsahují jednotlivá pravidla poukazující na to, jak by se měl klást konkrétní důraz při tvorbě webových stránek a obsahu samotného. Těmito principy jsou:

- Vnímatelnost – obsah a webové prezentace by měly být tvořeny tak, aby je mohli vnímat všichni uživatelé bez ohledu na postižení nebo technické vybavení
- Ovladatelnost – uživatelské rozhraní a jeho prvky by měly být plně a snadno ovladatelné z jakéhokoliv typu zařízení
- Srozumitelnost – obsah, ovládací prvky a navigace by měly být plně srozumitelné, uživatel by měl vše snadno pochopit

- Robustnost – tvorba obsahu by měla klást důraz na co nejlepší kompatibilitu se současnými, ale i budoucími technologiemi, ze kterých mohou uživatelé na web přistupovat (W3C konsorcium, 2018; Pavlíček, 2019)

Pravidel je celkově 13 a jsou rozděleny v rámci uvedených principů. Pro příklad lze uvést pravidlo Vnímatelnosti tzv. textové alternativy, kdy ke každému netextovému obsahu je zapotřebí zajistit textovou alternativu. Toto se dotýká například map, fotografií a jiného grafického obsahu, který by měl být vždy doplněn o popis vyjadřující co obsah znázorňuje. Z principu Ovladatelnosti se jedná například o pravidlo zajištění přístupu celého webu pouze z klávesnice na což navazuje pravidlo usnadnění používání webových stránek i za použití asistivních technologií. Pravidlo klade důraz na to, aby bylo snadné a pohodlné web používat v rámci různých vstupních zařízení. Pravidla Srozumitelnosti se zabývají především tím, aby byl kladen důraz na čitelnost a srozumitelnost obsahu a pomoc uživatelům vyvarovat se chyb, případně jim pomoci s reakcí a případným opravením. Toto lze znázornit na procesu vyplňování kontaktního formuláře. Ten by měl být navržen tak, aby bylo plně jasné, co je zapotřebí do jednotlivých polí vyplnit a v případě chybného vyplnění by měl být uživatel jasně upozorněn na možné chyby či nevyplnění povinného pole. Samotný princip Robustnosti již zahrnuje pouze jediné pravidlo, které klade důraz na maximální možnou kompatibilitu se současnými zařízeními, ale i budoucími technologiemi. Nesmí být opomenuty asistivní technologie. (W3C konsorcium 2018; Dobrý web, s.r.o.)

Pro všechna obsažená pravidla existují tzv. testovatelná kritéria úspěšnosti, která jsou aplikována především v momentě, kdy je zapotřebí otestovat současný stav webových stránek vzhledem k přístupnosti a stanovit případný rozsah zátěže. Kritéria jsou rozdělaná do tří úrovní dle priority A neboli musí být splněno, AA mělo by být splněno a AAA může být splněno. Ideálním stavem jsou webové stránky, které splňují všechna kritéria včetně skupiny AAA, na takovém webu jsou informace snadno dosažitelné a případné akce nevyžadují zdolávání složitých bariér. Na opačném konci jsou kritéria A, která lze označit za naprosté minimum. Příkladem kritéria A je například používání barev, kdy by měl být dodržen kontrastní poměr mezi textem a pozadím 3:1. V rámci metodický pokynů Ministerstva vnitra České republiky k zákonu o přístupnosti webových stránek je předpokládáno, že přístupnost webových stránek je dosaženo splněním kritérií z úrovní A a AA. (W3C konsorcium 2018; Dobrý web, s.r.o.)

### 3.1.1.3 Prohlášení o přístupnosti webových stránek

Mimo splnění kritérií přístupnosti taktéž udává zákon o přístupnosti webových stránek povinnost zveřejnění tzv. Prohlášení o přístupnosti webových stránek s odkazem na směrnici Evropské unie o přístupnosti webových stránek. Samotné prohlášení musí obsahovat stav souladu, který udává, v jakém stavu jsou webové stránky vůči přístupnosti a musí zde být uvedené příslušné zákony a směrnice přístupnosti. Typicky se jedná o prohlášení v plném souladu s přístupností nebo v částečném souladu v případě, že webové stránky disponují drobnými nedostatky, které však neznemožňují nalezení informací a pohybu na webu hendikepovaným uživatelům. V případě částečného souladu musí být všechny nedostatky uvedeny s odůvodněním jejich výskytu. (Zákon č. 99/2019 Sb.)

Prohlášení o přístupnosti dále musí obsahovat datum zpracování s uvedením zpracovatele a odkazem na kritéria, dle kterých bylo postupováno. Standardně se jedná o metodické pokyny k zákonu č. 99/2019 Sb., ve znění pozdějších předpisů a standardu nejnovější standardu WCAG v aktuální verzi 2.1. V závěru musí být uvedeny kontakty pro případnou zpětnou vazbu v návaznosti na problémy s funkčností webových stránek či podmínky v rámci nedostatků přístupnosti. (Zákon č. 99/2019 Sb., Vranková, 2023)

### 3.1.2 Zákon č. 106/1999 Sb. o svobodném přístupu k informacím

Zákon č. 106/1999 Sb. o svobodném přístupu k informacím udává povinným subjektům způsoby a povinnosti v oblasti poskytování informací spjatých s jejich organizací a činností. Ty jsou definovány jako státní orgány, územní samosprávné celky a jejich orgány a veřejné instituce. Zákon se nedotýká pouze prostoru webových stránek, ale zastřešuje celou problematiku poskytování informací napříč všemi úrovněmi. V rámci základních pojmů zákon vymezuje informaci jako obsah, který je zaznamenán na jakémkoliv nosiči. Může se jednat o fyzickou listinu nebo elektronický formát, a to nejen textový, ale i zvukový, audiovizuální nebo obrazový. Avšak je stanoveno, že informací není zamýšlen počítačový program. Důležitým vymezením je pojem tzv. zveřejněné informace:

*„Zveřejněnou informací pro účel tohoto zákona je taková informace, která může být vždy znovu vyhledána a získána, zejména vydaná tiskem nebo na jiném nosiči dat umožňujícím*

*zápis a uchování informace, vystavená na úřední desce, s možností dálkového přístupu nebo umístěná v knihovně poskytující veřejné knihovnické a informační služby podle knihovního zákona.*“ (Zákon č. 106/1999 Sb., 2023)

Tento bod je důležitý z hlediska přístupnosti daných informací. V rámci webových stránek by informace měly být publikovány tak, aby byly snadno dosažitelné a použitelné v rámci vzdáleného přístupu z jakéhokoliv zařízení. (Zákon č. 106/1999 Sb., 2023)

Zákon stanovuje 2 cesty zveřejnění informací, a to na základě žádosti anebo zveřejněním. Webových stránek se týká hlavně možnost zveřejnění, kde zákon dále rozšiřuje pojem informace následovně:

*„Informace poskytovaná zveřejněním se poskytuje za standardních podmínek užití a ve všech formátech a jazycích, ve kterých byla vytvořena; při zveřejnění takové informace v elektronické podobě musí být jeden z těchto formátů otevřený a, je-li to možné, též strojově čitelný. Je-li to možné a vhodné, zveřejní povinný subjekt spolu s informací též metadata, která se k ní vztahují. Formát i metadata musí splňovat otevřené formální normy.*“

(Zákon č. 106/1999 Sb., 2023)

V definici je jasně dáno, že formát dat má být takový, který nebude klást na uživatele další překážky v rámci jeho prohlížení či případného dalšího využití.

Samotné webové stránky by měly využívat standardní technologie, které umožňují prohlížení v jakémkoliv typu zařízení za využití jakéhokoliv internetového prohlížeče. Konkrétně se může jednat o značkovací jazyky XML nebo HTML. Na úrovních jednotlivých dokumentů lze předpokládat například obrázkové formáty PNG a JPEG, audio formáty FLAC a MP4 a v neposlední řadě dokumentový formát PDF, který je ideální pro publikaci textových souborů. Na rozdíl od spousty různorodých formátů pro textové soubory, nevyžaduje pro otevření externí aplikaci, ale lze jej otevřít přímo v internetovém prohlížeči. (Odbor hlavního architekta eGovernmentu)

Zákon dále vymezuje informace spjaté s působností a organizační strukturou subjektu, které je subjekt povinen zveřejnit. Tuto část v současnosti normalizuje vyhláška č. 515/2020 Sb. o struktuře informací zveřejňovaných o povinném subjektu a o osnově popisů úkonů vykonávajících v rámci agendy. (Zákon č. 106/1999 Sb., 2023)

### **3.1.2.1 Povinně zveřejňované informace**

V rámci zákona č. 106/1999 Sb. o svobodném přístupu a pozdější vyhlášky č. 515/2020 Sb. o struktuře informací zveřejňovaných o povinném subjektu a o osnově popisů úkonů vykonávajících v rámci agendy s účinností od 26. 12. 2020, musí povinné subjekty umístit na své webové stránky souhrn tzv. Povinně zveřejňovaných informací. Zjednodušeně řečeno se jedná o stránku, která obsahuje veškeré povinné informace. Současná struktura daná legislativní normou je shrnuta do 14 bodů. Obsah zahrnuje název subjektu, kontaktní údaje, organizační strukturu, datovou schránku. Dále jsou zde zastoupeny informace, jakým způsobem subjekt přijímá žádosti o informace, poskytuje informace a s tím spojené případné poplatky. V poslední radě jsou zastoupeny právní předpisy platné v rámci subjekty, licenční smlouvy a taktéž výroční zprávy. (Vyhláška č. 515/2020 Sb., 2021)

### **3.1.2.2 Úřední deska**

Na základě zákona č. 500/2004 Sb., správní řád a ve znění pozdějších předpisů jsou správní orgány, mezi které patří právě obce, zřídit úřední desku. Ta musí být neustále komukoliv veřejně přístupná. Zákon přímo nevymezuje pojem elektronické úřední desky, avšak přímo udává povinnost zveřejnění informací nejen fyzicky, ale i způsobem umožňující dálkový přístup. Zákon jasně udává, že úřední deska musí být přístupná v jakýkoliv den a v jakoukoliv hodinu. Informace zveřejňované v rámci fyzické a elektronické úřední desky nesmí být v rozporu. (Preuss, 2023; Zákon č. 500/2004 Sb. 2021)

Obsah na úřední desce je vyvěšován formou jednotlivých oznámení, které lze chápat jako ucelené bloky obsahující informace k jedné dané skutečnosti. Zákon jasně udává, že zveřejněn musí být vždy celý obsah oznámení a jednotlivých dokumentů, nesmí se jednat o pouhý seznam dokumentů či výtah informací. Avšak není povinností zveřejnit přímo originální dokument, postačujícím je zveřejnění vlastního textu, realizace dokumentu tak, aby byl obsah totožný s originálem. Každé oznámení na úřední desce musí disponovat datem zveřejnění a v případě, že dojde či došlo k jeho odebrání z úřední desky, tak i datem sejmutí. Obsahem a povinnostmi v rámci zveřejňování dokumentů, smluv a dalších informací se dále zabývá zákon č. 128/2000 Sb. o obcích. (Preuss, Ondřej, 2023; Zákon č. 500/2004 Sb. 2021)



V rámci webových aplikací jsou úřední desky často opatřeny dalšími funkcemi, které napomáhají k dalšímu šíření informací například formou emailů či SMS zpráv anebo funkce pro propojení s jinými aplikacemi například RSS kanál, avšak i zde je jasně dáno, že informace musí být dále prezentovány v takové formě, aby nedošlo k rozporu s původním oznámením na úřední desce. (Preuss, 2023)

Rostoucím trendem je nyní řešení kombinací fyzické úřední desky, jež je zbudovaná v blízkosti souvisejícího úřadu a řešení skrze webové stránky či specializovanou webovou aplikaci. Nejpokročilejším způsobem realizace je sjednocení obou řešení v jedno skrze vybudování interaktivních digitálních panelů, které často umožňují promítat obsah přímo z webové aplikace v rámci fyzického řešení a umožňují je ovládat přímo na místě. Tímto řešením odpadá složitosti spojené s udržováním totožného obsahu na obou typech úřední desky. (Preuss, 2023; Škop, 2022)

### **3.1.3 Obecné nařízení o ochraně osobních údajů**

Celým názvem Nařízení Evropského parlamentu a Rady (EU) č. 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů) je nařízením Evropské unie, které spolu se zákonem č. 110/2019 Sb. o zpracování osobních údajů se zabývají problematikou zpracovávání osobních údajů, který zahrnuje celý proces získání dat od jejich vyžádání, zaslání, následné uložení a jejich zpracování. Nařízení se zabývá právy a povinnostmi ve 3 rovinách, a to přesněji na úrovni subjektu údajů neboli uživatele jenž údaje poskytuje, dále správce, který osobní údaje získává, následně zpracovává a vyhodnocuje a poslední řadě může správce do procesu zapojit tzv. zpracovatele, který provede zpracování údajů. (Úřad pro ochranu osobních údajů; Pracovní skupina WP29, 2017)

Webové stránky jsou v mnoha případech prostředkem získávání různorodých údajů včetně těch osobních. Příkladem může být často využívaný kontaktní formulář, ve kterém jsou zpravidla vyžadovány kontaktní údaje včetně jména. Zde se již jedná o osobní údaje a správce je povinen postupovat dle zákona. V rámci nařízení je třeba stanovit jasný účel potřeby získání dat a je stanoveno několik důvodů, kdy lze data získávat bez výslovného souhlasu uživatele jedná se například o zpracování údajů, které je nezbytné pro splnění smluvního vztahu nebo v rámci ověření údajů před uzavřením smlouvy. Důvodů zákon definuje více, avšak v rámci webových stránek dochází ve velké většině k důvodům,

kteře vyžadují souhlas subjektu. Správce nesmí bez souhlasu subjektu data sbírat a dále zpracovávat. V případě, že subjekt souhlas udělí, může jej kdykoliv odvolat a na základě žádosti o odvolání souhlasu musí správce získané údaje vymazat ze své databáze. Správce musí uvést za jakým účelem osobní údaje vyžaduje a jak je dále zpracovává. (Úřad pro ochranu osobních údajů; Kouba, 2021)

V rámci nařízení jsou stanoveny zásady, které uvádějí, jakým způsobem mají být osobní údaje získávány, upřesňují jejich rozsah a využití. Při pořizování musí být dbáno na minimalistický rozsah dotazovaných údajů. Vyžadovat lze pouze údaje, které přímo souvisí s důvodem jejich pořizení. Pořizené osobní údaje musí být přesné a v rámci možností by mělo dojít k jejich ověření. Údaje, které správce od subjektu získá nesmí být v nesouladu s původním důvodem jejich zisku a nesmí dojít ani k využití v rámci jiného cíle. Samotné údaje smí správce uchovávat pouze po nezbytně nutnou dobu a v rámci jejich uložení musí být dosahováno co nejvyšší možné úrovně zabezpečení. (Úřad pro ochranu osobních údajů)

Subjekt neboli běžný uživatel, který data poskytuje by měl mít jasný přístup k informacím, jak je s jeho osobními údaji nakládáno, a to nejen v rámci zpracování, ale i jejich uložení a následného zabezpečení. Tyto údaje, které vychází přímo ze zásad je správce povinen poskytnout, tak aby byly snadno dosažitelné, a to ještě před samotným poskytnutím údajů. V rámci webových stránek musí být každý formulář opatřen informací za jakým účelem a osobní údaje zpracovávány s odkazem prohlášení, které je zpracováno v rámci zásad ochrany osobních údajů. V prohlášení je správce povinen uvést, jakým způsobem osobní údaje zpracovává, o jaké údaje se přesně jedná, za jakými důvody a cíli je zpracovává, po jakou dobu je uchovává, kdo je zpracovatelem, jak lze poskytnutý souhlas odvolat a práva subjektu. Samotný zákon nepředepisuje přesné znění, pouze odkazuje na stanové zásady. (Úřad pro ochranu osobních údajů; Kouba, 2021)

#### **3.1.4 Zákon č. 171/2023 Sb. o ochraně oznamovatelů**

V reakci na směrnici Evropského parlamentu a Rady 2019/1937 o ochraně osob, které oznamují porušení práva Unie byl implementován zákon č. 171.2023 Sb. o ochraně oznamovatelů. Tato problematika je též označována anglickým výrazem Whistleblowing. Zákon se zaměřuje na proces podání oznámení a ochranu oznamovatelů v předmětu nelegální či protizákonné činnosti v rámci povinných subjektů, mezi které patří například, firma čítající více jak 50 zaměstnanců nebo obce mající více než 10 000 obyvatel. Dále se

zabývá ochranou oznamovatelů před tzv. odvetnými opatřeními, které by mohl zaměstnavatele po zjištění nahlášení případné trestné činnosti na oznamovatele aplikovat. Například by se mohlo jednat o rozvázání pracovního poměru nebo změny ve finančním ohodnocení atd. Pro zabezpečení ochrany totožnosti oznamovatele musí být zajištěn přístup k oznámení pouze pověřené osobě, která nesmí umožnit nahlížení 3. osobě. Údaje lze poskytnout k nahlížení další osobě pouze se souhlasem oznamovatele. (Zákon 171/2023 Sb., 2023)

Zákon přesně vymezuje typy trestné činnosti, jež se mají oznámení týkat, lze například uvést protiprávní jednání se znaky jakéhokoliv trestného činu, se znaky přestupku s pokutou jejíž horní hranice je alespoň 100 000 Kč nebo porušení zákona o ochraně oznamovatelů atd. Je jasně stanoveno, že ač povinný subjekt musí disponovat systémem pro podávání oznámení, tak lze samotné oznámení podat ústně či osobně přímo pověřené osobě. (Zákon 171/2023 Sb., 2023; Chvátalová, 2023)

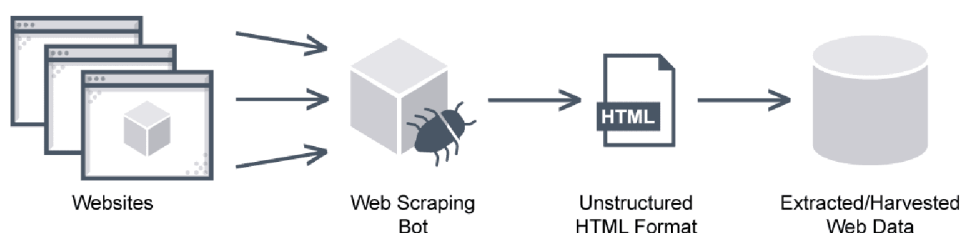
Povinné subjekty jsou povinny určit jednu či více osob, které budou pověřeny zpracováním obdržených oznámení s důrazem na její nestrannost v případném konfliktu. Dále musí zavést vnitřní oznamovací systém, který umožní podání oznámení a spolu s ním elektronicky uveřejnit informace popisující proces a způsob zpracování oznámení spolu s kontaktními údaji na pověřenou osobu. Obce, jež se povinnost týká mohou využít externí řešení skrze specializovanou aplikaci nebo řešení implementovat přímo do webových stránek, avšak v obou případech musí uvést povinné informace týkající se procesu zpracování oznámení a kontaktní údaje na pověřenou osobu. Nabízeným řešením může být webová stránka poskytující povinné údaje s formulářem pro podání oznámení. (Zákon 171/2023 Sb., 2023; Chvátalová, 2023)

## **3.2 Web scraping**

Web scraping je technika zaměřující se na získávání dat z webových stránek za účelem dalšího zpracování s využitím automatizovaného strojového řešení od jednoduchých skriptů až po sofistikovanější boty. Složitost řešení vždy závisí od řešené problematiky. Samotná data na webových stránkách bývají často v nestrukturované podobě a v rámci extrakce dochází i ke konverzi dat do strukturovaného formátu, který je vhodný pro další analýzu a vyhodnocování dat. Nutno zmínit, že cílem web scrapingu není datová analýza či jiný způsoby vyhodnocení dat, ale pouze jejich získání a uložení do žádoucího formátu určeného pro další zpracování. (Karatas, 2023; Harkiran78, 2023)

Způsobů uplatnění web scrapingu je v dnešní době opravdu mnoho ať už na poli osobního, výzkumného či obchodního využití. Jedním z častých způsobů užití je například sledování vývoje cen vybraných produktů na daných e-shopech nebo získávání kontaktních údajů napříč webovými stránkami, v neposlední řadě sběr dat za účelem specifické analýzy v rámci výzkumných prací. (Karatas, 2023; Harkiran78, 2023)

Obrázek 1 Proces webscrapingu



Zdroj: <https://avinetworks.com/glossary/web-scraping/>

### 3.2.1 Metody web scrapingu

Metod, jak web scraping aplikovat pro potřeby zisku dat je mnoho a závisí primárně na struktuře zkoumané webové stránky, požadovaném výstupu a znalostech tvůrce. V rámci rozsáhlejších projektů jsou často voleny kombinace těchto metod.

#### 3.2.1.1 HTML parsing

Jedná se o metodu, která je založená na analyzování zdrojového HTML kódu webových stránek. Požadovaná data jsou extrahována přímo z HTML elementů na základě zvolených parametrů. HTML parsing spočívá v získání zdrojového kódu, které je nejčastěji realizováno formou HTTP požadavku. Následně je provedena identifikace a výběr dat na základě stanovených cílů skrze volbu potřebných tagů a selektorů. Následuje fáze zpracování a extrakce dat, kdy dochází k potřebným úpravám v rámci řešené úlohy. Příkladem může být očištění dat od prázdných či nepotřebných hodnot. Když jsou data úspěšně extrahována a upravena dle potřeb, dochází k finálnímu uložení do zvoleného formátu, může se jednat například o textový soubor nebo databázi. (Karatas, 2023)

HTML parsing je často aplikován skrze frameworky programovacích jazyků. Mezi nejznámější patří BeautifulSoup a lxml pro Python nebo HtmlAgilityPack pro C#.

Metoda je vhodná zejména pro statické webové stránky, které nemění strukturu HTML kódu. (Karatas, 2023)

### **3.2.1.2 DOM parsing**

Metoda Dom parsing je svým zpracováním obdobná jako předchozí HTML parsing, avšak liší ve formátu, ze kterého se provádí extrakce dat. Nejdříve je získán zdrojový kód požadovaných webových stránek a následně je převeden do stromové struktury DOM. Následné zpracování je obdobné jako u HTML parsing jen probíhá právě v převedené DOM struktuře, a nikoliv přímo ve zdrojovém kódu. (Karatas, 2023)

### **3.2.1.3 API Scraping**

Metoda API Scrapingu je založena na extrakci dat pomocí API rozhraní, ze kterého je prováděna. Nedochozí tak ke sběru dat přímo ze zdrojového HTML kódu jako tomu je u předchozích metod. Hlavním předpokladem je dispozice API na straně webových stránek, na kterých bude web scraping prováděn. (Karatas, 2023; Apify)

Celý proces začíná identifikací API na straně webové stránky, některé mohou poskytovat API přizpůsobené přímo pro web scraping, například Facebook API, Amazon API atd. Následuje odeslání HTTP požadavku skrze konkrétní koncový bod v API, anglicky endpoint, který specifikuje určitou operaci skrze URL adresu. V rámci těchto požadavků lze efektivně získávat data bez složité filtrace či definice složitých cest napříč strukturou zdrojového kódu. Tomu předchází samotné funkce API, které jsou často přizpůsobovány pro efektivní výměnu dat. Skrze odpovědi na HTTP požadavek jsou často obdržena data již ve strukturované podobě, nejčastěji ve formátech JSON, HTML či CSV a je možné je dále využívat, avšak v rámci principů web scrapingu bývá většinou zapotřebí tyto data ještě upravit a případně extrahovat do vhodnější formátu. (Apify)

Výhodou API scrapingu je efektivní zpracovávání dynamický webových stránek, které mění obsah. Není závislý na struktuře zdrojového HTML kódu a CSS jako předchozí metody, a proto případné zásahy ve zdrojovém kódu nemusí znamenat potřebu reakci na straně API scraperu. Většina nevýhod je spojená v závislosti na realizaci API na straně webové stránky, kde nemusí být API poskytováno. Přístup může být limitován případně sledován, a proto nelze tuto metodu aplikovat v každém případě. (Apify)

### 3.2.1.4 Headless Browser Scraping

Headless Browser Scraping je metoda, která využívá při web scrapingu bezhlavé prohlížeče, anglicky právě Headless browsers. Jedná se o typ internetového prohlížeče, který nemá vlastní grafické rozhraní. Prohlížení webových stránek a provádění následných akcí je realizováno skrze příkazový řádek nebo s využitím automatizovaného skriptu. Při procházení a řešení úloh odpadá potřeba načítání webových stránek ať už s přechodem na jinou podstránku anebo vyvoláním změn či dialogových oken přímo na stránce. Veškeré operace se tak dějí na pozadí a zadavatel obdrží rovnou výsledek. (Karatas, 2023; ZenRows, Inc., 2022)

Hlavní výhodou je úspora času, která je úměrná rozsáhlosti řešené úlohy. V rámci web scrapingu je tato metoda snadno aplikovatelná i na dynamické webové stránky. Avšak vzhledem k lehce neintuitivnímu přístupu a vyšší míře potřebných znalostí jsou kladeny vysoké nároky na implementaci, která může být nejen náročná, ale i zdlouhavá. Využití metody je vhodné především pro rozsáhlejší úlohy a pro sběr velkého množství dat. (ZenRows, Inc.; 2022)

### 3.2.1.5 Aplikace a nástroje pro webscraping

V současnosti je dostupného velké množství externích řešení, která jsou realizována skrze samostatné aplikace nebo rozšíření do internetových prohlížečů. Poskytovatelé nabízí interaktivní prostředí, kde mohou uživatelé přímým kliknutím vybrat potřebné prvky pro sběr dat. Odpadá nutnost znalosti programovacích jazyků, avšak pro efektivní sestavení algoritmu je zapotřebí alespoň základní znalosti HTML a CSS. Většina poskytovatelů nabízí buď omezené verze zdarma nebo možnost pro vyzkoušení po omezenou dobu. Avšak volné dostupné verze jsou dostatečné spíše pro drobné potřeby. (Karatas, 2023)

Metody využívající externích, předpřipravených nástrojů jsou cíleny hlavně pro potřeby uživatelů s minimální znalostí programovacích jazyků. Každá z aplikací stanovuje vlastní pole působnosti, proto pokročilejší uživatelé mohou naopak svazovat při řešení úloh. Využití je vhodné spíše pro drobné úkoly. (Karatas, 2023)

### 3.2.2 Právní aspekt

Vzhledem k tomu, že web scraping se začal ve velké míře rozšiřovat a využívat teprve v nedávné době, tak je jeho využití v současnosti otázkou právních debat. Při získávání a následného zpracovávání dat z webových stránek může docházet k právním konfliktům z hlediska autorského zákona, zpracovávání a uchovávání citlivých dat v rámci GDPR či porušení smluvních podmínek souvisejících s využíváním obsahu dané webové stránky. S odkazem na tyto právní aspekty je patrné, že samotný proces získávání dat není protiprávním procesem, avšak případný konflikt může vzniknout při následném zpracovávání dat, jejich uchovávání nebo další distribucí v rámci vlastních řešení. V případě protiprávního sporu, kdy jsou autor web scrapingové aplikace a zpracoval dat 2 různé osoby, tak vina padá právě na zpracovatele dat, za předpokladu, že vytvořená aplikace není jasně určena pro protiprávní činnost. Na základě toho lze stanovit, že web scraping samotný je pouhým nástrojem a jako takový není v současné chvíli nezákonný, až teprve uchovávání, zpracování a využití získaných dat může nabýt podstaty protiprávního skutku. (Čech, 2022)

### 3.2.3 Autorská práva

Na obsah webových stránek se vztahuje zákon č. 121/200 Sb., zkráceně Autorský zákon, který definuje právní podstatu autorského díla, práva autora, možnosti jeho užití a další specifika. Zákon se zabývá komplexní problematikou autorských práv napříč všemi odvětvími a činnostmi, kde dochází ke vzniku nových autorských děl. Zákon definuje autorské dílo jako výsledek tvůrčí činnosti autora nebo tvůrčím zpracováním jiného autorského díla, kdy vzniká tzv. dílo odvozené. V rámci webových stránek se nejedná pouze o obsahovou část, která může zahrnovat například fotografie, videa atd, ale i tvůrčí část ve smyslu zdrojového kódu. Zákon taktéž definuje, co autorským dílem není, v rámci webových stránek lze zmínit hlavně zpravodajské informace, úřední díla, veřejné informace atd. (Zákon č. 121/2000 Sb., 2023; FAPI Business s.r.o., 2021)

Při web scrapingu může docházet ke kolizím v rámci využití obsahu webových stránek. Zde záleží, o jaký obsah se přesně jedná, zda se na něj autorský zákon vztahuje případně zda autor uplatňuje v rámci své webové stránky smluvní či licenční nároky pro využití jeho obsahu. Jako příklad lze uvést hojně využívanou skupinu licencí od organizace Creative Commons, které definují, za jaké podmínky lze dílo využívat. Mezi základní typy

patří například licence CC BY, v rámci, které lze obsah libovolně zpracovat s uvedením zdroje. Dále CC BY-NC, kdy lze obsah využívat pouze v rámci nekomerčních účelů s uvedením autora nebo CC BY – ND, kde musí být opět uveden autor, obsah lze využívat, ale nelze jej modifikovat. (CZ.NIC. 2014)

Autorský zákon se primárně zaměřuje na případy užití autorského díla pro komerční účely. Mimo ně definuje i pojem tzv. volného užití, kdy lze autorské dílo využít bez souhlasu autora. Jedná se o případy osobního využití, kde nedochází k obchodnímu nebo hospodářskému obohacením. Předpokladem je, že dílo není dále šířeno. Dalším způsobem je využití díla bez souhlasu autora je uplatněním tzv. zákonné licence. Jde o případy, kdy je užívání díla spojeno s veřejným zájmem, a nikoliv s vlastním obohacením. Jedná se o využití ve zpravodajských službách, vzdělání, výzkumu atd. (Zákon č. 121/2000 Sb., 2023)

### **3.3 Zpracování dat**

Zpracování dat, anglicky označováno též jako Data procesing, je metoda shromažďování dat s následnou transformací do využitelné podoby požadované informace. Na vstupu je datový soubor, který obsahuje data v nezpracované formě, často se jedná o nestrukturované soubory s minimální organizací, ze kterých lze těžko vyčíst potřebné informace. Metoda je koncipována do 6 základních fází na základě typu operací s daty. (Banu a Pedamkar, 2023; Duggal, 2023)

#### **3.3.1 Získávání dat**

První fáze je zaměřená na sběr nezpracovaných dat, která mohou mít různorodé vlastnosti, formáty atd. Důraz by měl být kladen na zajištění vhodných zdrojů a omezení se na sběr pouze potřebné skupiny dat. Samotný proces by měl být řešen co nejefektivnější metodou, která zajistí především rychlost a přesnost sběru dat. Efektivita procesu získávání dat se odráží nejen na samotném procesu z hlediska času a následné kapacitě uložených dat, ale i v rámci následujících částí metody zpracování dat, kdy je se získanou datovou základnou dále nakládáno. V případě, kdy zvolení nevhodných zdrojů a bez omezení se na potřebná data narůstá složitost řešení v rámci dalších zpracování, díky čemuž roste doba zpracování a zvětšuje se i riziko chyb. (Banu a Pedamkar, 2023; Duggal, 2023)



### **3.3.2 Ukládání dat**

Fáze ukládání dat je závislá na přímé aplikaci metody. Data mohou být uložena ihned po sběru a až následně zpracovávána. Dále lze skrze sofistikovanějších řešení s využitím komplexní automatizace může dojít k ukládání až po aplikování dalších fází metody. Nejčastějším takovýmto případem je automatické očištění dat ihned po sběru s následným uložením. Podstatou tohoto kroku je nejen volba vhodného uložení s ohledem na velikost potřebnou pro zpracovávána data, ale i definování případného rozložení uložení, aby nedocházelo k záměně dat v rámci jednotlivých fází zpracování. Důležitou součástí je zvolení datových formátů, ve kterých data budou uložena a do kterých budou dále transformována s ohledem na průběh řešení a stanovený výsledek. (Banu a Pedamkar, 2023, Duggal, 2023)

### **3.3.3 Čištění dat**

Čištění dat je důležitým procesem, ve kterém dochází k filtrování za účelem odstranění nežádoucích dat. Hlavním cílem je odstranění nesprávných, nadbytečných či neúplných dat za účelem usnadnění následného zpracování a dosažení kvalitních výsledků. Vyžaduje-li to řešená problematika dochází i ke třídění dat do ucelených skupin v rámci potřeb úlohy. Vzhledem k datové dispozici a stanoveným cílům lze uplatnit širokou škálu metod pro efektivní čištění dat. Lze uvést například kontrolu a odstranění duplicit, odstranění redundantních dat, převod či opravu formátu dat, validace dat atd. Metod, jak provést efektivní čištění dat je velké množství a vždy záleží na dané úloze a volbě řešení, zda je metoda vhodná pro použití. V případě, že je čištění dat vynecháno či provedeno špatně, může dojít k nemalému zkreslení výsledků celého zpracování anebo znemožnění výsledků dosáhnou. (Duggal, 2023; Khandelwal, 2023)

### **3.3.4 Transformace dat**

Transformace dat se primárně zaměřuje na celistvé soubory a samotnou strukturu dat, než na konkrétní data jako tomu bylo ve fázi čištění, po které mají být data již připravena pro další zpracování, a především následnou analýzu. I přestože jsou data již zbavena nežádoucích hodnot, neznamená to, že jsou připravena pro potřeby analýzy. Data jsou většinou před transformací stále v nevhodném formátu nebo struktuře neumožňující efektivní zpracování a vyhodnocování. Skrze transformaci jsou

uplatňovány například metody agregace dat, normalizace dat, redukce dimenzionality atd. Cílem transformace je převedení datových souborů do funkčních a výsledných formátů pro potřeby analýzy a zjišťování výsledků. Po transformaci by měly být soubory ve finální podobě, bez potřeby dalších úprav. (Duggal, 2023; Khandelwal, 2023)

### 3.3.5 Analýza dat

Analýzu dat lze interpretovat jako proces zkoumání datového souboru s cílem odhalit užitečné informace a případné vztahy mezi nimi pro podporu rozhodovacích procesů nebo zodpovězení stanovených otázek. K dosažení výsledků lze využít mnoho různorodých metod založených na statistice, logice či strojovém učení. Volba vhodné metody a nástrojů vždy závisí na stanovených cílech a datové dispozici. Vzhledem k formátu analyzovaných dat jsou metody rozlišovány na kvantitativní a kvalitativní.

(Frankenfield, 2023; Kelley, 2023)

Kvantitativní datová analýza je zaměřena pouze na měřitelné číselné hodnoty. Nejčastěji jsou uplatňovány statistické metody, jež se soustředí na nalezení spojitostí mezi daty. Jedná se například o analýzu časových řad, deskriptivní analýzu, shlukovou analýzu, korelační analýzu, regresní analýzu atd. Mezi nejčastějšími cíli kvantitativní datové analýzy patří testování statistických hypotéz neboli jejich potvrzení či vyvrácení nebo zjištění průměrných hodnot a trendů zkoumaného souboru. (Frankenfield, 2023; Kelley, 2023)

Kvalitativní datová analýza se zaměřuje hlavně na porozumění významu dat a jejich spojitostí na základě textových či grafických informací, nejčastěji slov a symbolů. Mezi základní metody patří například analýza obsahu, která se zaměřuje na systematické zkoumání textu s cílem na odhalení klíčových slov, témat a souvislostí v daném souboru. Dále Grounded Theory, která cílí na vysvětlení zkoumané události, analýzou předchozích případů výskytu. Ještě lze zmínit často využívané principy narativní analýzy zaměřující se na práci s daty, které byly získány prostřednictvím průzkumů, rozhovorů či jiným obdobným typem. (Frankenfield, 2023; Kelley, 2023)

Datovou analýzu lze dále dělit nejen podle zpracovávaných hodnot, ale taktéž dle účelu, v rámci kterého jsou následně uplatňovány specifické metody. Mezi nejpoužívanější typy patří:

- Deskriptivní analýza – využívá převážně deskriptivních statistický metod pro odhalení klíčových charakteristik datového souboru s cílem získání základního přehledu v datech či stanovení trendu
- Diagnostická analýza – hlavním cílem je odhalení vztahů mezi daty a příčiny nastání, zkoumaných jevů převážně s využitím regresní a korelační analýzy
- Průzkumný analýza – zaměřuje se na pochopení souvislostí dat bez předem daných hypotéz se zaměřením na odhalování vtaů mezi daty s využitím vizualizace a různorodých statistický metod
- Textová analýza – na rozdíl od předešlých typů se zaměřuje na analyzování textových dat skrze extrakci klíčových slov, klasifikaci textu a zkoumání vzájemných souvislostí. (Frankenfield, 2023; Kelley, 2023)

### **3.3.6 Prezentace dat**

Po provedení analýzy dat následuje konečná prezentace výsledných informací, které byly v rámci předešlého zpracování zjištěny. Způsob prezentace se odvíjí od typu výsledků, požadovaného výsledného formátu a v neposlední řadě na od volby nástrojů pro analýzu dat. Nejčastěji jsou data prezentována skrze grafickou vizualizaci formou jednoduchý grafů, animací či tabulek. Ze závěrečné prezentace musí být jasně patrné dosažené výsledky. (Banu a Pedamkar, 2023; Khandelwal, 2023)

## **3.4 Uživatelské testování**

Uživatelské testování je proces, ve kterém dochází k testování zvolených aplikací, služeb nebo prototypů skutečnými uživateli s cílem odhalení chyb, posouzení použitelnosti a dalších náležitostí spojených s užíváním testované aplikace. Testování je zaměřeno k odhalení problémů, které mohou vzniknout právě využíváním aplikace běžnými uživateli. Dochází k ověření hypotéz o funkčnosti a použitelnosti testované aplikace, dále k identifikaci rizikových bodů, které mohou být překážkou v samotné použitelnosti a vést uživatele k nechtěným operacím a chybám. Díky začlenění uživatele do procesu lze definovat jeho chování v rámci testované aplikace a skrze získání zpětné vazby odhalit další nedostatky aplikace. (Hotjar Ltd., 2023; Torsten)

K účelům testování jsou často využívány prototypy aplikací, které obsahují pouze nezbytně nutné funkce, rozložení ovládacích prvků a minimalistický design, pro zkoumání

uživatelského chování a použitelnosti daného řešení. Díky tomuto přístupu uplatněného v prvních fázích vývoje lze ušetřit nemalé finanční částky, které by byly spojeny s odhalením nepoužitelného řešení a chyb v případě již dokončené aplikace. Další častou aplikací je využití testování na stávající řešení s cílem odhalení uživatelských návyků, nežádoucího chování či úrovně celkové použitelnosti. (Omniconvert, 2023)

### **3.4.1 Postup uživatelského testování**

S ohledem na stanové cíle, času, respondenty a dostupné technologie lze aplikovat různorodé metody, které se liší převážně v samotném přístupu k testování, avšak pro efektivní proces testování by měl být vždy dodržen, alespoň rámcový postup. Úvod testování se zaměřuje na stanovení cílů zkoumání. S cíli úzce souvisí následný rozsah a způsob provedení samotného testování. Právě proto musí být definovány na samotném začátku. V návaznosti na stanovené cíle jsou v případě potřeby vytvářeny testovací objekty, což jsou například aplikace nebo webové stránky, na kterých bude probíhat následné testování. Testovacím objektem mohou být jak hotové, již zaběhnuté aplikace, tak i speciálně navržené prototypy, které jsou uzpůsobeny pro potřeby testování. Následuje výběr testovací metody, dle které bude při testování postupováno. Rozboru těchto metod je věnována následující podkapitola. (Torsten; Omniconvert, 2023)

V momentě, kdy jsou jasně stanoveny cíle, připravený objekt testování a zvolena vhodná metoda, dochází k sestavení testovacích scénářů a případného harmonogramu. Testovací scénáře určují jaké operace má uživatel při testování provádět. Tyto scénáře lze rozdělit do dvou základní skupin, na uzavřené, jež definují jasný postup krok po kroku, co má uživatel dělat s minimální volností. Druhou možností jsou otevřené, které dávají uživateli podstatně větší volnost při testování a jsou více zaměřeny na uživatelovo chování a jeho svobodnou mysl při postupování v rámci testů. (Omniconvert, 2023)

Následuj výběr uživatelů, který by měl být taktéž specifický s ohledem na cíle testování. S ohledem na potřeby testování je vhodné stanovit cílovou skupinu, ze které jsou následně uživatelé vybíráni. Mezi důležité faktory, která mohou hrát roli jsou například věk, zaměstnání nebo vzdělání. Mimo vlastnosti samotných uživatelů je třeba zvážit i počet uživatelů s ohledem na rozsah testování a stanovené cíle. (Omniconvert, 2023)

Po zajištění veškerých přípravných procesů je zahájeno samotný proces testování, který je prováděn s ohledem na zvolenou metodu. Testování může probíhat za fyzicky přítomnosti uživatelů nebo vzdáleně a v obou případech lze aplikovat metodu řízeného

testování s moderátorem, který celý proces vede a kontroluje. Během testování jsou sbírána data pro následnou analýzu. (Omniconvert, 2023)

Závěrečnou fází testování je analyzování dat získaných během testování a jejich následné vyhodnocení. Typ získaných dat záleží na aplikované metodě a vyžívaných technologiích. Po provedení datové analýzy jsou výsledky interpretovány dle potřeby testování a případné prezentace. (Omniconvert, 2023)

### **3.4.2 Metody uživatelského testování**

Výběr vhodné metody pro testování závisí především na stanovených cílech, předpokládaného rozsahu testování, zvoleném testovaném subjektu a technologiích, jež budou následně uplatňovány a využívány pro potřeby testování. V rámci komplexního testování mohou být uplatněny i kombinace jednotlivých metod pro dosažení co nejlepších výsledků. Metod, které lze aplikovat v praxi je velké množství, proto bude demonstrativně popsány pouze některé, jejichž využití lze uplatnit u webových stránek a aplikacích. (Hotjar Ltd., 2023)

#### **3.4.2.1 Laboratorní testování**

Laboratorní testování je založeno na průběhu v laboratorních podmínkách, které jsou přizpůsobeny pro potřeby zkoumání. Je vyžadována osobní přítomnost uživatelů. Důležitým prvkem, který je často v rámci této metody uplatňován je moderátor, který uživatele vede v průběhu testování a může jej přizpůsobovat vzniklým situacím. Moderátor může v rámci potřeby uživatele kontrolovat, čímž zajistí vyšší kvalitu výsledků a získává okamžitou zpětnou vazbu, kterou může být strojově obtížně zaznamenávána. Takovýto přístup k testování je označován jako moderovaný a jednou z metod testování. (Torsten)

Testování v laboratorních podmínkách často implementuje další metody, nejčastěji využitím přístrojového měření, kdy je zkoumáno uživatelské chování skrze zaznamenávání biometrických dat nebo pohybu kurzoru a klikání. Příkladem lze zmínit využití technologie Eyetrackingu, která je založená na monitorování pohybu očí uživatele a následném analyzování jeho pozornosti. (Hotjar Ltd., 2022)

Hlavními výhodami je možné využití široké škály metod a technik, které pomohou k přesnější analýze výsledků a samotná režie v podobě moderátora, díky které lze efektivněji testování vést a získávat další poznatky. Mezi nevýhody lze zařadit omezení na

danou lokalitu, což může stěžovat výběr vhodných uživatelů pro testování. Laboratorní testování je též kapacitně velice omezeno a nelze zapojit velké množství uživatelů. Nutné je i zmínit, že jsou uživatelé vytrženi z komfortní zóny a aplikaci testují v pro ně neznámém prostředí, což může vést ke zkreslení výsledků. (Torsten; Hotjar Ltd., 2022)

### **3.4.2.2 Vzdálené testování**

Vzdálené testování je pravým opakem testování laboratorního. Testovací scénáře a testovaný subjekt jsou navrženy pro možnosti vzdáleného testování bez fyzické přítomnosti uživatelů, kteří mohou testovat aplikace z pohodlí domova. V rámci metody jsou aplikovány nástroje pro vzdálené připojení, umožňující propojení s uživatelem a následné sledování činností, jež jsou prováděny při průchodu testování. (Carr, 2022)

Metoda může být aplikována v rámci vzdáleného připojení s moderátorem, ale i bez něj. Vzhledem k řešené úloze lze taktéž využít mnoho metod a nástrojů jako u laboratorního testování, avšak je zapotřebí počítat s limity zařízení na straně uživatele. Pro potřeby vzdáleného testování je často aplikováno zaznamenávání celého průběhu testování včetně záznamu z videokamery, kterou by měl uživatel disponovat. (Schade, 2013)

Výhodou je samotný vzdálený přístup, který umožňuje zapojení široké škály uživatelů a odpadá bariéra v rámci soustředění s na jednu lokaci. S tím souvisí i absolvování procesu v domácích podmínkách na známém zařízení, kdy uživatel není v rámci vystavení stresu náchylný na atypické chování. Nevýhody opět plynou ze vzdáleného přístupu, kdy mohou průběh testování narušovat stav internetu. Další nevýhody plynou z případných limitací uživatelského rozhraní. (Carr, 2022; Schade, 2013)

### **3.4.2.3 Task analysis**

Task analysis, do českého jazyka přeloženo jako analýza úkolů je metoda, která se soustředí na zjištění a porozumění toho, jak uživatelé plní zadané úkoly, jaké způsoby řešení volí, jak v plnění postupují a jaké mohou nastávat problémy v rámci jejich postupu. Případně může sloužit k odhalení spojitosti, které vedou k nežádoucímu chování uživatele. Princip metody je v zadávání úkolů nebo otázek uživateli, který se snaží zadání splnit. Následně dochází k analýze uživatelského postupu a dosažených výsledků v rámci cílů testování. Typickým příkladem může být nalezení specifické informace na webových

stránkách, uživateli není stanoven postup, ten volí si sám. Právě zvolený postup je následně podroben analýze. (Schade, 2013; User Interviews Inc.)

Metodu lze rozdělit do 3 samostatných fází. V 1. fázi jsou stanoveny cíle, od kterých se následně odvozují testované úkoly. Tato fáze je také zaměřena na sběr informací ohledně testovaných úkolů. Je zapotřebí jednoznačně analyzovat, čeho se úkoly přesně týkají a jakým způsobem by měly být řešeny. Tyto informace lze zjistit s využitím metod v rámci interakcí s uživateli, například rozhovorem, dále může být uplatněno sledování uživatelské aktivity a neposlední řadě lze využít simulaci, kdy jsou jednotlivé úkoly procházeny přímo aplikací v rámci pochopení její funkčnosti a možných postupů řešení. (Schade, 2013)

Následuje 2. fáze analýzy úkolů, kdy dochází k rozložení jednotlivých úkolů na jednotlivé úlohy či operace, které vedou ke splnění stanovených úkolů. K tomuto úkonu jsou často využívány metody grafických znázornění, nejčastěji diagram analýzy úloh. V rámci diagramu dochází k rozložení úkolů na dílčí drobné úkoly, které jsou následně rozděleny do operací nutných pro jejich splnění. Vzniká tak rámec, definovaný postup, který by měl uživatel v ideálním případě dodržet v rámci testování. Při detailnější analýze lze zahrnout očekávané možnosti jiných řešení v případě komplikace. Příkladem může být proces přihlášení uživatele, kdy uživatel zapomene heslo. Zde dochází ke zmapování zcela jiného postupu k úspěšnému přihlášení. Cílem prvních 2 fází je detailní porozumění testované problematice a vytvoření očekávaných postupů za účelem analýzy. (Schade, 2013; User Interviews Inc.)

Poslední fáze je věnována samotnému testování, kdy jsou uživateli předloženy jasně specifikované úkoly a případné informace nutné k testovacímu procesu. Všechny kroky uživatele jsou dále sledovány s využitím dalších metod a následně vyhodnoceny s ohledem na druhou část, sestavené diagramy a očekávané postupy řešení. (Schade, 2013)

#### **3.4.2.4 Metody s využitím strojového měření**

Metody využívající strojového měření a sledování uživatelské činnosti jsou většinou součástí jiných metod, a to hlavně z důvodu jejich zaměření se primárně na sběr dat. Jejich zaměření je velice specifické a v případě samostatného využití nejsou schopny poskytnout komplexní data a odpovědi v rámci testování, avšak jsou účinnými nástroji pro zjištění dat v rámci konkrétních situací a problematik, řešených v následných analýzách. Pro demonstraci následuje popis 3 nejpoužívanějších metod.

Eyetracking je metodou, která se zaměřuje na sledování pohybu očí uživatele. Tohoto je dosaženo skrze zařízení, která monitorují pohyb zornic v rámci pohybu očí uživatele při sledování obrazovky počítače. Přínosem metody je zjištění toho, kde a jakým způsobem je upoutávána uživatelská pozornost v rámci průchodu testování. Lze například odhalit rušivé prvky v aplikaci, které odvádí uživatelskou pozornost. (Casado, 2020)

Heatmaps, v českém překladu teplotní mapy, využívají metodu pro sledování kurzoru myši na obrazovce. Podobně jako Eyetracking jsou teplotní mapy zaměřeny na odhalení způsobu orientace a pohybu v rámci testované aplikace a také se s jejich pomocí analyzují prvky, které přitahují uživatelskou pozornost. Název je odvozený od způsobu vizualizace, kdy specializovaný software na základě monitorování barevně vykreslí a odliší prostory aplikace, na základě uživatelské aktivity. (Casado, 2020)

Záznamy relací jsou metodou, která využívá nahrání záznamů uživatelského postupu napříč testováním. V případě specializovaného softwaru lze aplikovat znázornění pohybu kurzoru myši v rámci relace, kdy je jednoznačně vidět, jak uživatel postupoval a co zaujalo jeho případnou pozornost v reálném čase. Případně lze využít i Eye-tracking. Metoda je sice časově náročnější pro následné zpracování, ale poskytuje detailní informace o uživatelském postupu. (Casado, 2020)



## 4 Praktická část

### 4.1 Stanovení požadavků

Problematika legislativních požadavků v rámci webových stránek obcí je velice obsáhla. Jednotlivé zákony, vyhlášky a směrnice řeší dané problematiky komplexně, a to nejen v rámci publikovaného obsahu a jeho formy, ale i procesů a dalších povinností mimo prostor webových stránek. Tyto procesy je dobré znát z důvodu pochopení problematiky, avšak v rámci výzkumné části nejsou zahrnuty, protože je buď nelze algoritmicky otestovat nebo nejsou odraženy v rámci obsahu webových stránek, který je hlavním předmětem zkoumání této práce v souladu se stanovenými cíli.

Na základě specifikace, kterou legislativa udává, lze jednotlivé požadavky rozdělit do několika skupin, charakteristickými dle způsobu povinného provedení. Rozdělení požadavků je účelné především z důvodu následného získávání dat, které bude pro každou skupinu vyžadovat odlišné provedení, které musí počítat s případnou mírou variability v jejich řešení.

#### 4.1.1 Plně definované

Plně definované požadavky jsou charakteristické svou úplnou definicí v rámci legislativy a nesmí docházet k žádné odchylce. Mají přesně definovanou strukturu nebo její část toho, jak mají vypadat a jaké informace mají obsahovat. V rámci této skupiny lze uplatnit model 1:1, kdy je možné porovnat se požadavky přesně dle vytvořeného vzoru.

Konkrétně se jedná o stránku Povinně zveřejňované údaje, které má přesně stanovenou strukturu vyhláškou č. 515/2020 Sb., ta je dle vyhlášky koncipována do 14 bodů a pro každý bod je jasně specifikovány jaké informace v něm mají být obsaženy.

Leč v legislativě není přesně definován pojem elektronické úřední desky, ale pouze úřední desky obecně, tak i přesto lze úřední desku zařadit do této skupiny. Důvodem je jasně specifikované vlastnosti jednotlivých oznámení, ta by měla mimo oznamovací části disponovat především daty zveřejnění a sejmutí.

Dále je zařazena část GDPR, která se týká formulářů na webových stránkách. Ty musí být ze zákona opatřeny zaškrtačacím polem pro udělení souhlasu se zpracováním osobních údajů v případě, že budou osobní údaje využívány pro další činnost, například zaslání emailových nabídek. Pokud je formulář využíván pro jednorázovou činnost, například podání dotazu či oznámení, tak postačuje pouze sdělení informující na

zpracování osobních údajů. Obě verze musí následně odkazovat na celistvé prohlášení GDPR, kde jsou přesně popsány procesy ohledně zpracování obdržených údajů.

#### 4.1.2 Částečně definované

Jako částečně definované požadavky lze chápat, kdy legislativa sice stanovuje jasná požadavky, avšak v rámci jejich plnění je subjektu umožněná jejich volnost. To znamená, že struktura řešení požadavků může být v praxi velice variabilní a proces jejího ověřování byl značně neefektivní, tudíž by mohlo docházet k nesprávným výsledkům. Do této skupiny spadají zejména různé formy prohlášení, která musí obsahovat specifické informace, avšak jejich interpretace už je závislá na řešení subjektu. Vhodným řešením je ověřování požadavků skrze využití klíčových slov a výrazů, které by měly být v rámci realizace vždy uvedeny.

Mimo splnění požadavků na přístupnost musí být v rámci legislativy zpracováno i tzv. Prohlášení o přístupnosti webové stránky. Zákon sice nastiňuje hrubou strukturu prohlášení, avšak spíše se zaměřuje na informace, které by měly být v rámci oznámení obsaženy, proto nelze uplatnit model 1:1. Prohlášení musí odkazovat na související zákony, dále na úroveň souladu s odkazem na metodiku využitou při tvorbě webových stránek. Dále musí být zahrnut popis stránky, kontakt na odpovědnou osobu, datum zpracování prohlášení a sekce nepřístupný obsah s odůvodněním jeho výskytu.

Dalším požadavkem této skupiny je prohlášení v rámci ochrany osobních údajů neboli GDPR. Legislativa opět nevymezuje přesnou strukturu informací, avšak definuje, jaké informace by měly být v prohlášení obsaženy. Vzhledem k tomu, že je problematika GDPR velice obsáhlá hlavně z pohledu nastavení nutných procesů v rámci zpracování osobních údajů, které se v rámci subjektů mohou značně lišit, tak zde dochází ke značným rozdílům ve zpracování, protože musí být tyto procesy zohledněny i v následném prohlášení. Prohlášení musí obsahovat informace o zpracovateli dat a způsobu jejich zpracování, dále způsob zabezpečení, kontakt na odpovědnou osobu, postupy.

Dalším zástupcem je oznámení o ochraně oznamovatelů. Samotné oznámení vykazuje stejné charakteristiky jako předešlá prohlášení. Není vyžadována přesná struktura, avšak je vyžadováno zahrnutí informací ohledně způsobu podání oznámení, definice oznamovatele, kontaktů na pověřenou osobu, postupů zpracování, ochrany oznamovatele a údajů. Nutno zmínit, že tato povinnost se týká pouze webových stránek obcí s počtem obyvatel nad 10 000. Avšak s ohledem na povinnost zahrnutí tohoto obsahu do webové prezentace

až k 15. 12. 2023, není věcné toto ověřovat, protože v případě absence se v době zpracovávání této práce nejedná o přestupek.

#### **4.1.3 Definice na základě zdrojového kódu**

V poslední skupině jsou zahrnuty požadavky, které nejsou na první pohled vidět a jejich podstatu tkví ve správné zpracování zdrojového kódu. Proces ověření bude zaměřen převážně na zkoumání zdrojového kódu, kde bude cíleno na konkrétní tagy a jejich atributy.

Zástupcem této skupiny je samotná přístupnost webových stránek, konkrétně se jedná o grafické prvky, které musí být opatřeny alternativním textem. Nejčastěji se jedná o obrázky, iframy nebo samostatná videa. Hlavním důvodem přítomnosti alternativního textu v rámci zdrojového kódu je využití asistivních technologií u nevidomých uživatelů, kdy alternativní text plní funkci prostředku poskytnutí informace, což je v rámci obrázku prezentováno.

#### **4.1.4 Metody ověření požadavků**

Vzhledem k definovaným skupinám požadavků, které mají svá specifika v rámci variability možných řešení s ohledem na legislativu, jsou uplatňovány zejména následující metody:

- Testování obsahu pomocí klíčových slov a výrazů
- Testování kódu v rámci jednotlivých značek a atributů
- Testování přístupnosti dokumentu

## **4.2 Sběr a zpracování dat**

Vzhledem zkoumaného obsahu je v rámci procesu sběru dat zvolen web scraping, konkrétně metoda HTML parsingu. Důvod výběru této specifické metody je nejen její jednoduchost a přehlednost v rámci zkoumaných požadavků, ale především statický charakter webových stránek obcí. Předkladem je, že takovéto webové stránky mají naplňovat hlavně informační charakter bez zbytečných a složitých funkcí a prvků, které jsou často nasazovány skrze dynamická řešení.

Pro konkrétní řešení je zvolen programovací jazyk python, mezi jehož hlavní výhody patří nejen jednoduchost a přehlednost zdrojového kódu, ale hlavně vysoká úroveň

podpory a přítomnost velkého množství knihoven. Pro potřeby algoritmu jsou implementována knihovny BeautifulSoup4, které je zaměřená přímo na webscraping a Requets, které je zaměřená na odesílání HTTP požadavků. Dále je implementována knihovna CSV, pomocí které je zajištěna práce se soubory ve formátu csv a to z důvodu ukládání dat do vhodného formátu, pro potřeby následné analýzy.

Ukázka zdrojového kódu níže prezentuje využití obou knihoven. V rámci knihovny Requets je uplatněn příkaz requests.get, který realizuje odeslání HTTP požadavku pro získání zdrojového kódu webová stránky. Následně dochází k nasazení knihovny BeatifulSoup, kdy jsou v rámci zdrojového kódu vyhledány všechny obrázky pomocí tagu img

#### Zdrojový kód 1 Extrakce obsahu webové stránky

```
source = requests.get('https://www.pef.czu.cz/cs').text
soup = BeautifulSoup(source, 'xml')

img = soup.find_all('img')
print(img)
```

Zdroj: vlastní zpracování

S ohledem na co největší možnou efektivitu algoritmu jsou řešeny procesy sběru, filtrování a zpracování dat řešeny v rámci jednoho průchodu. To znamená, že data jsou po zisku ihned odfiltrována a zpracována v rámci vhodně navržených funkcí a až následně uložena. Díky tomu způsobu realizace odpadá nutnost ukládání a načítání několika verzí dat a v rámci jednoho průchodu budou data rovnou uložena do formátu csv, který je vhodný pro další zpracování v rámci analýzy.

Nejčastější případ využití ukládání dat je prezentován v následující ukázce zdrojového kódu. Zde dochází k otevření již vytvořeného souboru v rámci příkazu open, který specifikuje o jaký soubor se jedná a dále parametry pro práci se souborem. Skrze další příkazy jsou definovány hodnoty a jejich samotný zápis do souboru. Ukázkový kód realizuje zapsání definovaných hodnot do nového řádku souboru.

#### Zdrojový kód 2 Ukládání dat CSV souboru

```
csv_file_path = 'pristupnost_kod.csv'
with open(csv_file_path, mode='a', newline="", encoding='utf-8') as file:
    writer = csv.writer(file)
    writer.writerow([x[2], vysledek["ano"], vysledek["ne"], vysledek["spatne"]])
```

Zdroj: vlastní zpracování

#### 4.2.1 Řešení obsahu skrze klíčové výrazy

Řešení použití klíčových slov a výrazů lze při správném provedení využít v rámci většiny úloh, které se zabývají ověřením obsahu v rámci jednoduchého algoritmu. Úskalím tohoto způsobu řešení je zohlednění využití různých synonym a případného skloňování slov či změny ve slovosledu. V rámci komplexnějších a rozsáhlejších úloh se nabízí využití textových modelů umělé inteligence, avšak v rámci této práce byl zvolen způsob ověření skrze prostý algoritmus, v rámci kterého je zapotřebí vhodně stanovit soubor klíčových slov.

Použití klíčových výrazů je využito pro potřeby ověřování existence stránek z mapy stránek a navigace. Dále pro ověření prezentovaných informací v rámci prohlášení o přístupnosti webových stránek, prohlášení o ochraně osobních údajů a struktury povinně zveřejňovaných údajů. Jednotlivé realizace mají podobný charakter řešení, a proto bude prezentován zdrojový kód v rámci jednoho řešení.

První část zdrojového kódu provádí převedení surového zdrojového kódu do užitečné podoby potřebné pro testování. Vzhledem k tomu, že velikost písmen nehraje roli, jsou všechna velká písmena převedena na malá pro jednodušší testování. Při prvním spuštění algoritmu došlo k špatnému vyhodnocování použití klíčových slov, kdy při odebrání HTML značek v rámci převodu kódu na pouhý text došlo ke spojení některý textových řetězců vlivem HTML struktury. Dalším případem bylo využití tzv. nezalomitelných mezer, které jsou na úrovni kódu reprezentovány řetězcem „&nbsp;“. Oba tyto případy jsou ošetřeny ve zdrojovém kódu nahrazením či přidáním mezery.

### Zdrojový kód 3 Úprava získaného obsahu

```
def PrevedText(obsah):
    ctext = ""
    if obsah:
        for odstavec in obsah:
            ctext += " " + odstavec.get_text(separator=' ', strip=True)
        ctext = ctext.lower()
        ctext = ctext.replace("&nbsp;", " ")
    return ctext
```

Zdroj: vlastní zpracování

Na druhé ukázce zdrojového kódu je znázorněno samotné filtrování. Pro doplnění potřebných funkcí byla využita knihovna „re“. Nejdříve jsou inicializovány proměnné pro počítání počtu použitých a nepoužitých klíčových slov spolu s proměnnou, do které se zapisují slova, jež nejsou při aplikaci zjištěna, použita. Následně jsou procházena všechna klíčová slova i v rámci jejich synonymum a je hledán jejich výskyt v získaném textu. V případě nalezení dochází k zvětšení počtu požití o hodnotu jedna, v opačném případě dojde ke zvýšení počtu nepoužitých o 1 a uložení slova do seznamu nepoužitých. Pokud je nalezeno jedno ze synonym je automaticky vyhodnoceno použití všech slov rámci zadané varianty synonymum.

### Zdrojový kód 4 Funkce ověřování použití klíčových slov

```
def ProjdiKlics(text, klicova_slova):
    pslova = 0
    nslova = 0
    SeznamNslov = []
    for klicove_slovo in klicova_slova:
        synonyma = klicove_slovo.split("/")
        for synonymum in synonyma:
            nalezeno = re.search(r'\b' + re.escape(synonymum.lower()) + r'\b', text)
            if nalezeno:
                pslova += 1
                break
        else:
            nslova += 1
            SeznamNslov.append(klicove_slovo)
    return pslova, nslova, SeznamNslov
```

Zdroj: vlastní zpracování

Vzhledem k tomu, že by ruční zpracování jedné stránky po druhé bylo neefektivní je algoritmus upraven tak, aby zajistil projití zadaného seznamu adres a všechny je vyhodnotil. Vytvořená funkce je zobrazována na zdrojovém kódu níže. Prezentovaná funkce využívá již předešlé zmiňované funkce pro filtrování a zpracování obsahu. Její přínos je hlavně v tom, že v prochází jednotlivé adresy ze seznamu adresy a provádí všechny úkony nutné pro zpracování. Tedy navázání a extrakci dat pomocí knihoven Request a BeautifulSoup. Následně dochází k využití funkcí pro úpravu zdrojového kódu, převedení na text a ověření použití klíčových slov. Ve finální části dochází k využití funkce pro zapsání výsledných dat do csv souboru.

Zdrojový kód 5 Funkce provádějící kompletní analýzu

```
def AnalyzujWeb(seznam_adres, klicova_slova):
    for adresa in seznam_adres:
        source = requests.get(adresa)
        source.encoding = 'utf-8'
        soup = BeautifulSoup(source.text, 'html.parser')
        obsah_stranky = soup.find_all('div')
        ctext = PrevedText(obsah_stranky)
        pouzita, nepouzita, nepouzita_klicova_slova = ProjdiKlics(ctext, klicova_slova)
        uloz(adresa, pouzita, nepouzita, nepouzita_klicova_slova)
```

Zdroj: vlastní zpracování

Při aplikování algoritmu se ukázalo být toto řešení uspokojivé pro potřeby úlohy, a to přesně v případě mapy stránek, povinně zveřejňovaných informací a prohlášení o přístupnosti. Vzhledem k zjištěné míře odlišností při aplikování algoritmu byly prováděny změny v klíčových výrazech a v případě nutnosti zařazena další synonyma. Avšak v rámci ověřování prohlášení o ochraně osobních údajů vykazovalo testování velkou míru chybovosti, a to s ohledem vysokou úroveň různorodosti řešení. Rozsah informací a forma se v rámci jednotlivých řešení natolik lišila, že nebylo možné vytvořit jednotné řešení, které by uspokojivě analyzovalo, zda je dané řešení v rámci požadavků na legislativu uspokojivé. Tomu nasvědčuje i fakt, že některé subjekty realizují umístění informace

ohledně ochrany osobních údajů skrze množství stažitelných souborů. Vhodnost řešení tak nelze testovat s využitím jednoduchých algoritmů s uspokojivou úrovní efektivity.

#### 4.2.2 Řešení přístupnosti na základě ověřování kódu

V rámci testování přístupnosti webové stránky jsou navrženy 2 řešení, která se zabývají 2 odlišnými částmi. První z nich je algoritmus zaměřený na prověřování povinných atributů, které musí HTML značky obsahovat za účelem popisu obsahu na nepřístupnému grafickém prvku. Jedná se především o obrázky, fotografie a iframy. Každý takovýto prvkem musí obsahovat alternativní text, který informace o obrázku obsahuje. K tomuto účelu slouží především atribut alt, který je pro alternativní text přímo určenou. Algoritmus počítá i s alternativou využití atributu title, který sice přímo neplní funkce atributu alt, ale při vhodně zvoleném názvu může splnit podmínku přístupnosti. Důležité je zajištění přístupnosti v rámci grafického obsahu.

Při ověřování přístupnosti toho obsahu, je kladen důraz na zahrnutí alespoň jedné fotogalerie, jednoho iframu, kterým jsou často realizovány náhledy do map, dále stránky s obyčejným obsahem, nejčastěji články o historii a alespoň jedna aktualita obsahující fotografie. Důvodem je zachycení a ověřování co možná nejvíce vzájemně podobného typu obsahu, pro co nejlepší výsledky. Zahrnutím jakéhokoliv různorodého obsahu by mohlo dojít ke značnému zkreslení výsledků, i přestože by měl být testovaný obsah dostupný ze 100 % bez výjimek.

Navržený algoritmus je realizován skrze jednu hlavní funkci, v rámci které je rovnou implementován i sběr dat. Ten je blíže specifikován a soustředěn pouze na obsahovou část webové stránky z důvodu toho, že hlavičky a patičky webových stránek často obsahují grafický obsah a při zahrnutí těchto částí do sběru by došlo k znehodnocení výsledků díky jejich mnohonásobnému započítání do výsledků, protože v rámci algoritmu dochází k testování vícero podstránek. Ve stejném příkazu, přesněji „`soup.find('div', {'id': 'obsah'}).find_all('img')`“ je dále specifikována značka pro prvky, kterou jsou v rámci průchodu testovány. V rámci ukázky je to značka `img` reprezentující obrázky a fotografie. Následné zpracování je realizováno pomocí `for` cyklu, který nejdříve ověřuje přítomnost atributů, v rámci prezentace kódu níže se jedná o `alt`. V případě, že se jedná o `alt` dochází ke kontrole, zda atribut není prázdný případně neobsahuje přípony souborů `JPG` nebo `PNG`. Pokud je atribut obsažen, není prázdný nebo chybně vyplněný, dochází skrze přičtení hodnoty 1 do proměnné „`ano`“, reprezentující počet přístupných obrázků.



V opačném případě dochází k přičtení hodnoty 1 do proměnné „ne“, která reprezentuje prvky, které tagem vůbec nedisponují nebo proměnné „spatne“, která reprezentuje počet špatně vyplněných nebo prázdných atributů. Výsledky jsou následně ukládány do seznamu výsledek, který je dále zpracováván v rámci ukládání výsledků. Vzhledem k rozdílné struktuře je zapotřebí vždy zohlednit id nebo class divu, který slouží pro zabalení obsahové části podstránky a dále v případě změny zkoumaných atributů tyto změny zavést do algoritmu.

Zdrojový kód 6 Funkce provádějící analýzu grafického obsahu

```
def Obrazky(adresa, ano, ne, spatne):
    source = requests.get(adresa).text
    soup = BeautifulSoup(source, 'html.parser')
    obrobsahu = soup.find('div', {'id': 'obsah'}).find_all('img')

    for obrazek in obrobsahu:
        try:
            alt = obrazek["alt"]
            altBoolean = True
            if ".jpg" in alt or ".png" in alt or alt == "":
                spravnyAlt = False
                spatne = spatne + 1
                ne = ne - 1
            else:
                spravnyAlt = True
        except Exception as e:
            altBoolean = False
            spravnyAlt = True
        if spravnyAlt and (altBoolean):
            ano = ano + 1
        else:
            ne = ne + 1
    vysledek = {
        "ano": ano,
        "ne": ne,
        "spatne": spatne
    }
    return vysledek
```

Zdroj: vlastní zpracování

### 4.2.3 Řešení přístupnosti z hlediska souborů

Velké množství obsahu je na webové stránky přidáváno formou stažitelných souborů. Nejčastěji jsou takto přidávány soubory v rámci úřední desky. Dále se může jednat o různorodé vyhlášky, zákony, účetní uzávěrky, záznamy ze zasedání zastupitelstva atd. Význam přístupnosti těchto souborů je tak neméně důležitý jako je tomu u prostého textu a případného grafického obsahu, a proto je důležité tyto dokumenty poskytovat v čitelné formě. Vzhledem přístupnosti by se v ideálním případě mělo jednat o formát PDF, který klade minimální bariéry pro jeho zobrazení na rozdíl do jiných formátů, kdy je například vyžadována externí aplikace pro možnost zobrazení obsahu. Pro potřeby kódu bylo zapotřebí implementovat další 2 knihovny a PDFPlumber a PDFSyntaxError, které jsou zaměřeny na práci s PDF soubory a případnými chybami.

V první části kódu, který je uveden níže je demonstrována funkce „StahniAnalyzuj“, která stáhne zkoumaný soubor do definované složky pomocí proměnné adresář a název stanoví ze spojení zjištěného názvu a textu staženo. Následně dochází k otevření souboru a pomocí knihovny PDFPlumber a otestování, zda se na první stránce nachází text. Podmínka ohledně otestování první stránky je stanoveno z důvodu zvýšení rychlosti testování, které je při stahování různorodých souborů a jejich následné analýze náročnější z hlediska času. Předpokladem přístupnosti je 100% přístupnost informací, tudíž soubor, který je přístupný pouze z části, lze považovat za nepřístupný, protože neposkytuje handicapovaným uživatelům stejné informace. Výsledek toho, zda se v dokumentu nachází text je uložen do proměnné „text“ a následné vyhodnocení probíhá v rámci další funkce „ProjdiUloz“, avšak v případě, že se zkoumaný soubor není v formátu PDF, dochází k chybové hlášce a soubor není dále analyzován ani zohledňován ve výsledcích. Mezi hlavní důvody této realizace patří složitost řešení v rámci jiných typů souborů a dále předpoklad, že by měl být upřednostňován právě formát PDF. Dále předpoklad, že soubory ve formátech DOC a Excel, které bývají po formátu PDF nejčastěji viditelné na webových stránkách, obsahují přístupný obsah vzhledem ke své povaze.

Ve druhé části kódu je definována funkce „ProjdiUloz“, která prochází všechny nalezené odkazy. Vzhledem k tomu, že odkazy na soubory jsou ve zdrojovém kódu definovány pouze částečně v atributu href. Je zapotřebí takovéto odkazy nejdříve spojit s odkazem na hlavní doménu webových stránek, aby byly funkční a bylo možné soubory získat. Po vytvoření funkčního odkazu je následně zavolána funkce „StahniAnalyzuj“, která provede analyzování zkoumaného dokumentu a vrátí výsledek ve formě proměnné text, která je následně testována, zda je prázdná v případě nenalezení textu nebo neprázdná v případě nalezeného textu. V případě, každých z variant dochází v momentě jejich nastání k přičtení hodnoty 1 v rámci stavu výsledku pro zjištění celkového počtu přístupných a nepřístupných souborů.

## Zdrojový kód 7 Funkce provádějící analýzu přístupnosti PDF souborů

```
def StahniAnalyzuj(odkzsoub, nazevsoub, adresar):
    try:
        source = requests.get(odkzsoub)
        pdfko = os.path.join(adresar, f"{nazevsoub}stazeno.pdf")
        with open(pdfko, "wb") as file:
            file.write(source.content)
        with pdfplumber.open(pdfko) as pdf:
            if len(pdf.pages) > 0:
                first_page = pdf.pages[0]
                text = first_page.extract_text()
                return pdfko, text
            else:
                return pdfko, None
    except requests.exceptions.RequestException as e:
        print(f"Stážení {odkzsoub} se nezdařilo. Chyba: {e}")
        return None, None
    except PDFSyntaxError as e:
        print(f"Extrakce textu z {odkzsoub} se nezdařila. Chyba: {e}")
        return pdfko, None

def ProjdiUloz(seznam_adres, domena, adresar, csv_writer):
    pristupnosuma = 0
    nepristupnosuma = 0
    for url_stranky in seznam_adres:
        odkazy = HledejOdkazy(url_stranky)
        pocetpristupno = 0
        pocetnepristupno = 0
        for odkaz in odkazy:
            odkaz = odkaz['href']
            odkzsoub = domena + odkaz
            nazevsoub = odkaz.split("/")[-1].replace("?", "_").replace("&", "_")
            pdfko, text = StahniAnalyzuj(odkzsoub, nazevsoub, adresar)
            if pdfko:
                if text:
                    pocetpristupno += 1
                else:
                    pocetnepristupno += 1
            pristupnosuma += pocetpristupno
            nepristupnosuma += pocetnepristupno
    csv_writer.writerow([domena, pristupnosuma, nepristupnosuma])
```

Zdroj: vlastní zpracování

#### 4.2.4 Specifické úlohy

Specifickou úlohou lze nazvat problematiku GDPR v rámci dostupných zdrojů a zákonem daných povinností bylo taktéž realizováno řešení skrze klíčové výrazy pro ověření obsahu, který by mělo prohlášení GDPR obsahovat, avšak zde došlo na rozdíl od ostatních úloh k velké míře chybovosti. Při podrobnější ruční analýze v rámci webových stránek za účelem minimalizace chybovosti, které bylo nutné provést v určitém rozsahu u všech řešení, bylo zjištěno, že způsoby realizace se velice odlišují. Zákon v rámci prohlášení neudává přesnou strukturu informací, které musí být zveřejněny, pouze definuje okruhy, jež by se měly týkat. V praxi tak dochází k různorodým interpretacím, který mimo textovou formu často zahrnují zveřejnění pomocí souborů ke stažení. Zjištěna byla i vysoká míra rozdílnost v rozsahu prohlášení. Zvoleným typem řešení v rámci této práce nelze dosáhnout spolehlivých výsledků, a právě proto je zkoumána hlavně přítomnost a dostupnost toho prohlášení v rámci webových stránek.

Další specifickou úlohou je zákon o ochraně oznamovatelů, v jehož důsledku musí být webové stránky doplněny o prohlášení, které poskytují všechny potřebné informace týkající se této problematiky. Nutno zmínit, že plná účinnost zákona a s tím spojená povinnost doplnění webových stránek je platná od 15. 12. 2023, za předpokladu, že se tato povinnost obce týká. Z tohoto důvodu nebyla problematika do zpracování zahrnuta.

Avšak v rámci GDPR lze zahrnout zpracování formuláře, protože ze zákona je povinností v případě sběru osobních údajů poskytnout údaje, které jsou předmětem prohlášení GDPR, subjektu, který údaje poskytuje. V rámci toho byl navržený algoritmus, jehož hlavní část je prezentována níže. Funkce „AnalyzujFormu“ provádí analýzu formuláře na základě otestování přítomnosti hlášky, která by měla uživatele upozorňovat nebo směřovat na prohlášení GDPR a odkazu na něj. Následně jsou stanoveny 4 možné stavy, které mohou nastat na základě přítomnosti a nepřítomnosti hlášky a odkazu, v rámci, který je vrácena číselná hodnota reprezentující tento stav. Tato funkce je následně využita v následující funkci „PojdiWeb“, ve které jsou nejdříve inicializovány výstupní proměnné. Následně dochází ke sběru dat, který je soustředěn pouze na formuláře, které obsahují prvek s atributem „submit“. Po získání dat je volána funkce „AnalyzujFormu“, které provede vyhodnocení a vrátí číselnou hodnotu, na základě které jsou v rámci výsledné varianty přiřčeny hodnoty 1 v rámci proměnných, které byly inicializovány na začátku funkce. Tyto hodnoty jsou následně uloženy do CSV souboru.

## Zdrojový kód 8 Funkce provádějící analýzu formuláře

```
def AnalyzujFormu(form):
    hlaskaBoolean = ("osobní" in form.text and "údaje") or "souhlas" or "zprac" in form.text
    odkazBoolean = form.find("a") is not None
    if hlaskaBoolean and odkazBoolean:
        return 1
    elif not hlaskaBoolean and odkazBoolean:
        return 2
    elif hlaskaBoolean and not odkazBoolean:
        return 3
    else:
        return 4

def ProjdiWeb(adresa):
    obsahujeHlasku = 0
    neobsahujeHlasku = 0
    obsahujeOdkaz = 0
    neobsahujeOdkaz = 0

    try:
        source = requests.get(adresa).text
        soup = BeautifulSoup(source, 'html.parser')
        form = soup.find(["input", "button"], attrs={"type": "submit"}).findParent("form")
        vysledek = AnalyzujFormu(form)
        if vysledek == 1:
            obsahujeHlasku += 1
            obsahujeOdkaz += 1
        elif vysledek == 2:
            neobsahujeHlasku += 1
            obsahujeOdkaz += 1
        elif vysledek == 3:
            obsahujeHlasku += 1
            neobsahujeOdkaz += 1
        elif vysledek == 4:
            neobsahujeHlasku += 1
            neobsahujeOdkaz += 1
    except Exception as e:
        print(f'není form pro {adresa}')
    return adresa, obsahujeHlasku, neobsahujeHlasku, obsahujeOdkaz, neobsahujeOdkaz
```

Zdroj: vlastní zpracování

## 4.3 Analýza

Vzhledem k tomu, že využitá metoda web scrapingu zajišťuje principy nejen sběru dat, ale i následného čištění, filtrace a uložení, je možné rovnou přistoupit k analytické části získaných dat k čemuž byla využita aplikace Microsoft Excel. Důvodem výběru je přehledné prostředí, které poskytuje veškeré potřebné nástroje pro zpracování dat v rámci této diplomové práce. Data lze nejen zpracovat pomocí matematických funkcí, ale i vizualizovat skrze grafy, což je velice užitečným nástrojem. Výhodou je i snadná, intuitivní implementace dat ve formátu CSV.

Při analýze byla uplatňována zejména metoda deskriptivní statistické analýzy dat, která je v rámci cílů práce dostačující pro získání potřebných informací o současném stavu plnění legislativních povinností v rámci obsahu webových stránek. Uplatňovány jsou zejména průměrné hodnoty, extrémy a vizualizace s použitím grafů, které jasně a přehledně prezentují výsledky analýzy.

### 4.3.1 Přítomnost povinného obsahu

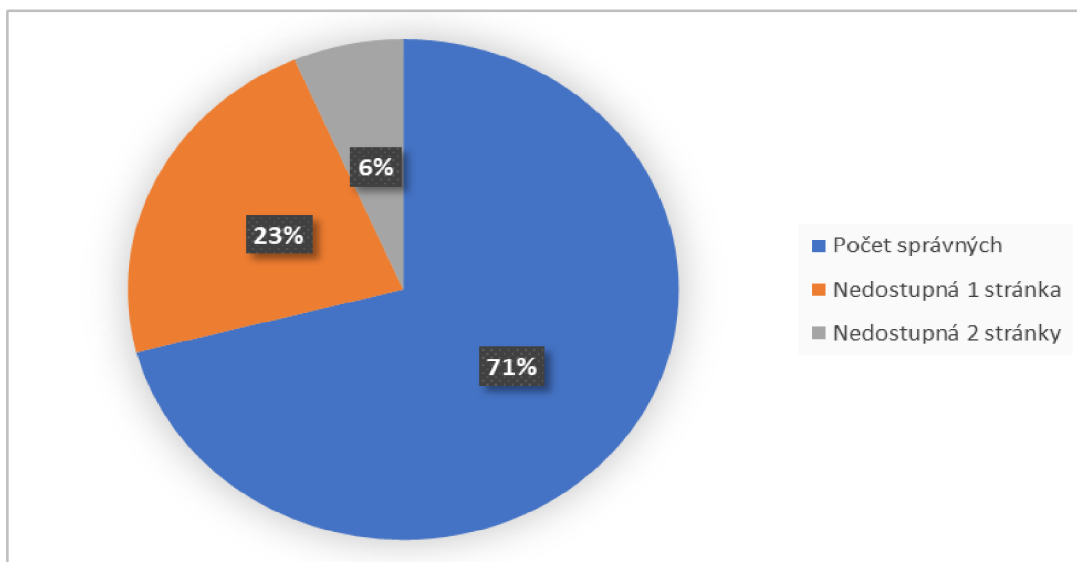
Základním předpokladem je dohledatelnost povinného obsahu, který by měl být snadno dohledatelný buď skrze navigaci neboli menu, případně mapu stránek, která může být využita právě handicapovanými uživateli, protože poskytuje seznam všech stránek v stromové struktuře bez dalších prvků, které by mohly být zbytečně předčítány. Sběr dat zahrnoval obě tyto cesty k přístupu včetně umístění v patičce a hlavičce webových stránek. Všechny zahrnuté informace by měly být snadno dosažitelné a samostatně prezentované. V rámci analýzy lze připustit možnost, že je obsah na webu přítomen, avšak vzhledem k tomu, že není dosažitelný přímo z navigace, mapy stránky, zápatí a záhlaví, tak jej lze nazvat za hůře až těžko dosažitelným a tím pádem nesplňující kritérium snadné dosažitelnosti, a proto je v rámci vyhodnocení takovýto případ vyhodnocen negativně.

Nutno zmínit shrnutí povinně zveřejňovaných údajů v rámci jedné povinné stránky, která je ze zákona nezávazně stejnojmenně, dle této řešené problematiky. Ze zákona se však nejedná pouze o tuto stránku, ale i o všechny informace, které zahrnuje. Často jsou některé body řešeny samostatnou stránkou a v rámci povinné stránky jsou zahrnuty odkazem. Jedná se například o elektronickou podatelnu, poskytování informací, vyhlášky, zákony atd. Avšak vzhledem k tomu, že zákon ukládá povinnost tyto všechny informace sjednocovat v rámci stránky Povinně zveřejňované údaje, která má zákonem danou strukturu, není předmětné zkoumat přítomnost jednotlivých stránek zvlášť, protože v praxi dochází

k odlišnostem v řešeních a tak některé informace nemusí být prezentován samostatnou stránkou, ale zahrnuty pouze v rámci stránky Povinně zveřejňované údaje a naopak, čímž by se zvedla potenciaální chybovost ověřování v rámci algoritmického testování. Právě proto je testována přítomnost právě zákonem dané stránky Povinně zveřejňované údaje, dále stránek Prohlášení o přístupnosti, GDPR a úřední desky.

Z celkového počtu 62 obcí splnilo povinnost všech 4 bodů celkem 44 obcí. Nejhorším výsledkem jsou případy 2 nedostupných stránek v rámci jedné webové stránky, což bylo zjištěn celkem u 4 obcí. Nejčastěji chybějící či nedostupnou stránkou je prohlášení GDPR a to přesně v 14 případech. Následující grafy prezentují vyhodnocení v rámci procentuálního zastoupení webových stránek v rámci zkoumané dostupnosti a následně počet případů nenalezení jednotlivých stránek.

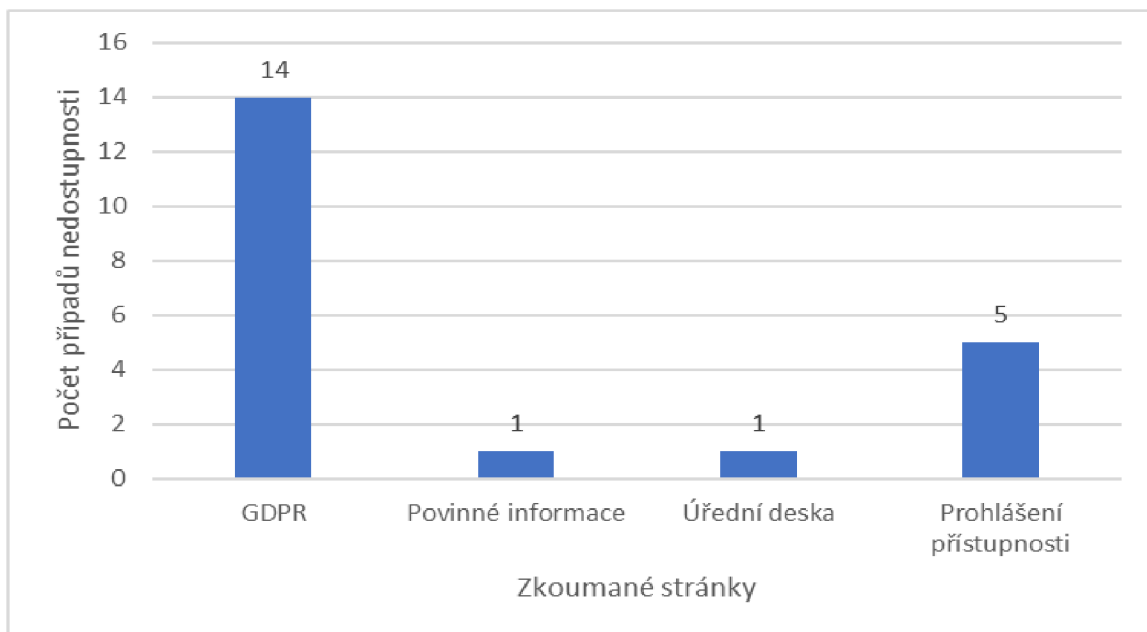
*Graf 1 Ověření přítomnosti povinných stránek*



Zdroj: vlastní zpracování



Graf 2 Nedostupné povinné stránky



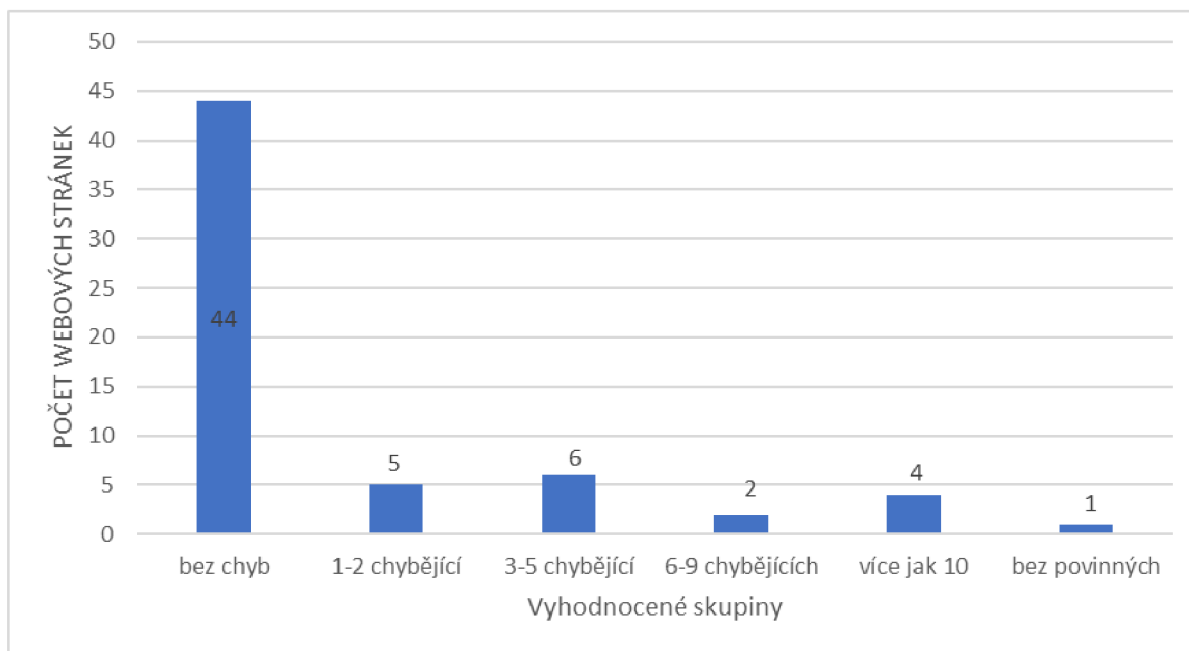
Zdroj: vlastní zpracování

#### 4.3.2 Povinně zveřejňované informace

Každé webové stránky obce musí bezpodmínečně obsahovat shrnující stránku, která obsahuje všechny povinně zveřejňované informace. Tato podstránka má přesně danou strukturu, která je rozčleněna celkem do 14 bodů, kdy některé mají ještě jednotlivé podsekce. Tato struktura musí být bezpodmínečně dodržena. Povoleny jsou drobné odchylky v interpretaci, kdy můžeme narazit na drobné změny v názvech. Tyto odchylky byly při analýze stávajících řešení zohledněny při sběru a filtraci dat. Příkladem lze uvést často odlišný bod 7, kdy je nejčastěji aplikován zkratkovitý název DIČ, ale častou alternativou je název Plátce daně z přidané hodnoty. Tyto alternativy nijak neporušují zákonem stanovené podmínky, ale bylo je zapotřebí zohlednit při sběru a zpracování dat, za účelem minimalizace případné chybovosti.

V rámci provedené analýzy 44 obcí obstálo bez chyby, přesněji obsahovaly všechna povinná kritéria, který stanoveno celkem 30. Dále byla zjištěna 1 webová stránka, která povinnou podstránkou vůbec nedisponuje a další 4 vykázaly velkou míru chybovosti, tj. nesplnily více jak 10 kritérií. Graf níže zobrazuje procentuální shrnutí počtu obcí v rámci jednotlivých případů vyhodnocení.

Graf 3 Vyhodnocení plnění kritérií povinně zveřejňovaných informací



Zdroj: vlastní zpracování

### 4.3.3 Prohlášení přístupnosti webových stránek

Analýzy prohlášení o přístupnosti webových stránek je realizováno obdobným způsobem jako předchozí analýza stránky Povinně zveřejňované údaje. Zde je nutno zmínit, že legislativa týkající se samotného prohlášení není tak důrazná jako je tomu ve zmíněném předchozím případě. Jsou udány okruhy a informace, které musí prohlášení obsahovat, avšak konkrétní zpracování již zahrnuje určitou míru volnosti. V rámci kritérií bylo zapotřebí detailnějšího studia jednotlivých realizací, tak aby došlo k ověření pouze takových informací, které musí být vždy obsaženy v prohlášení a jejich nejčastějších textových alternativ v rámci zpracování. Pro potřeby ověřování bylo stanoveno celkem 17 klíčových výrazů s ohledem na dostupné vzory a získané realizace s důrazem na co nejefektivnější způsob ověření.

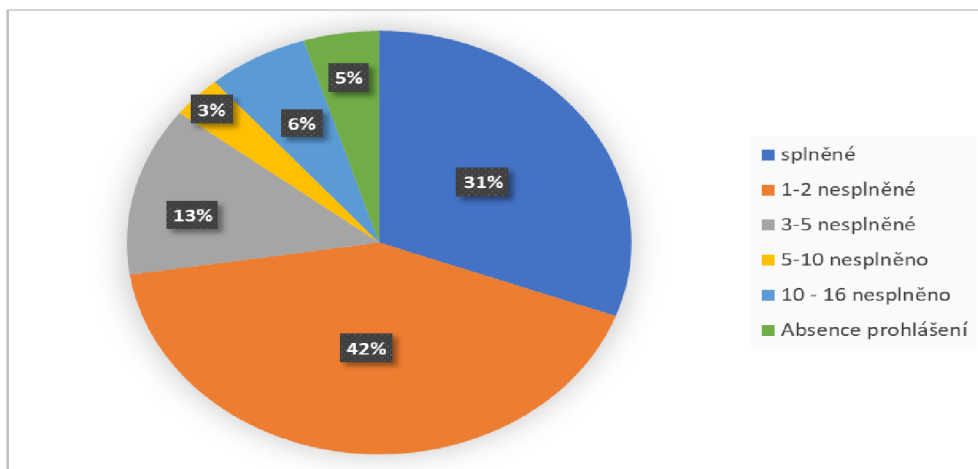
Z výsledku je patrné, že celkem 19 webových stránek splnilo všech 17 bodů. Kladně lze taktéž zhodnotit 26 webových stránek, které zaznamenali pouze 1-2 chybějící informace, což lze vzhledem k volnosti interpretace považovat taktéž jako případ splnění. Ve 3 případech nebylo prohlášení nalezeno vůbec, což lze považovat za hrubý nedostatek vůči legislativě. Po absenci prohlášení lze uvést jako nejhorší případy 13 a 12 nepřítomných klíčových výrazů. Jedná se celkem o 3 webové stránky, které při bližší

analýze textu disponují pouze shrnujícím textem, který však postrádá většinu povinných informací, které by toto prohlášení mělo obsahovat.

Při zahrnutí pouze případů, které prohlášením disponovaly, lze uvést, že průměrný počet splněných, nalezených klíčových výrazů na webovou stránku je 15. Nejčastějším klíčovým výrazem, který nebyl naplněn je „nepřístupný obsah“ a to v 38 případech. V rámci legislativy je nutné v prohlášení samostatně shrnout případy nepřístupného obsahu, dále zdůvodnit jeho existenci a nabídnou případné alternativy jeho zpřístupnění. V rámci metodik a dostupných vzorů je tento obsah zpracováván v samostatné sekci reprezentované názvem „Nepřístupný obsah“. Při detailnější analýze textu, bylo odhaleno, že se touto problematikou prohlášení částečně zabývají v obecné části textu, avšak z odkazem na legislativu, lze takovýto případ považovat za nedostatečný. Níže prezentované grafy prezentují nejčastěji nedodržená kritéria a dále počet webových stránek procentuálního zastoupení vyhodnocených skupin webových stránek, dle splnění výskytu analyzovaných informací.

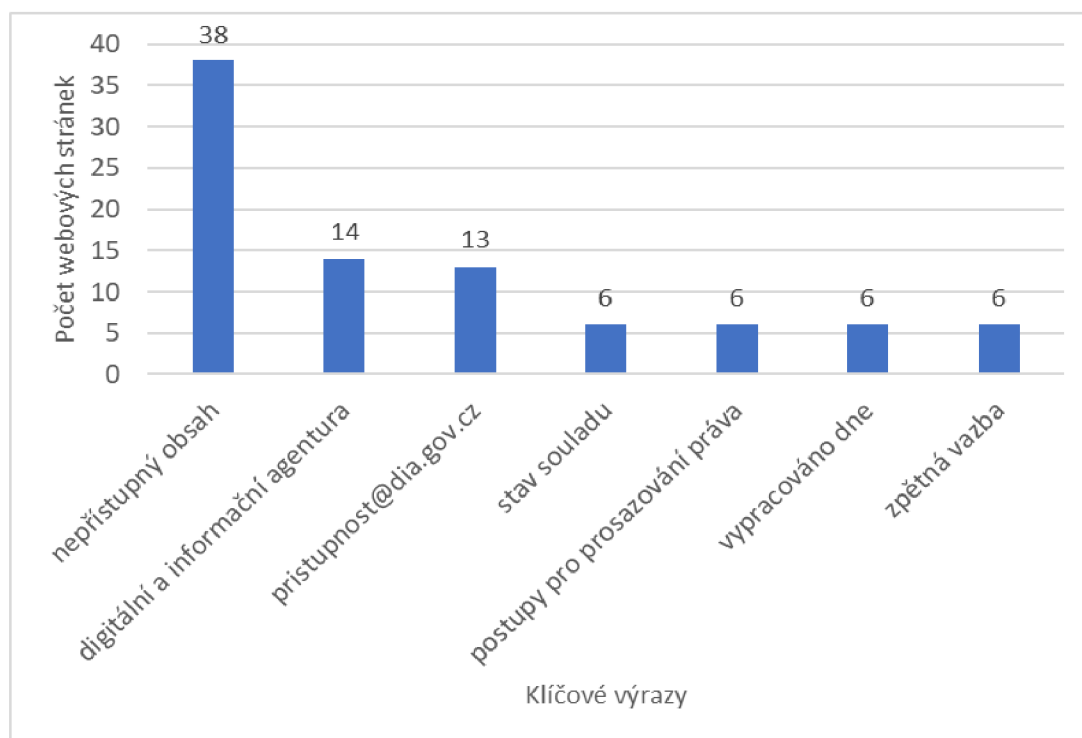
Další skupinou nejčastěji neobsažených klíčových výrazů jsou „digitální a informační agentura“ a „pristupnost@dia.gov.cz“, první nenalezen ve 14 případech a druhý ve 13. Obě klíčová slova souvisí s obsahem ohledně možné podání stížností na pověřený státní orgán v záležitosti porušování přístupnosti. Tyto informace musí být taktéž v prohlášení ze zákona obsaženy a v případě jejich nezveřejnění se může jednat taktéž o porušení zákonem daných požadavků na prohlášení. Uživatel musí být jasně informovat o možnostech kontaktování jak na stranu správce, tak i pověřený státní orgán.

*Graf 4 Vyhodnocení plnění kritérií prohlášení přístupnosti*



Zdroj: vlastní zpracování

Graf 5 Nejčastěji nesplněné klíčové výrazy v prohlášení přístupnosti



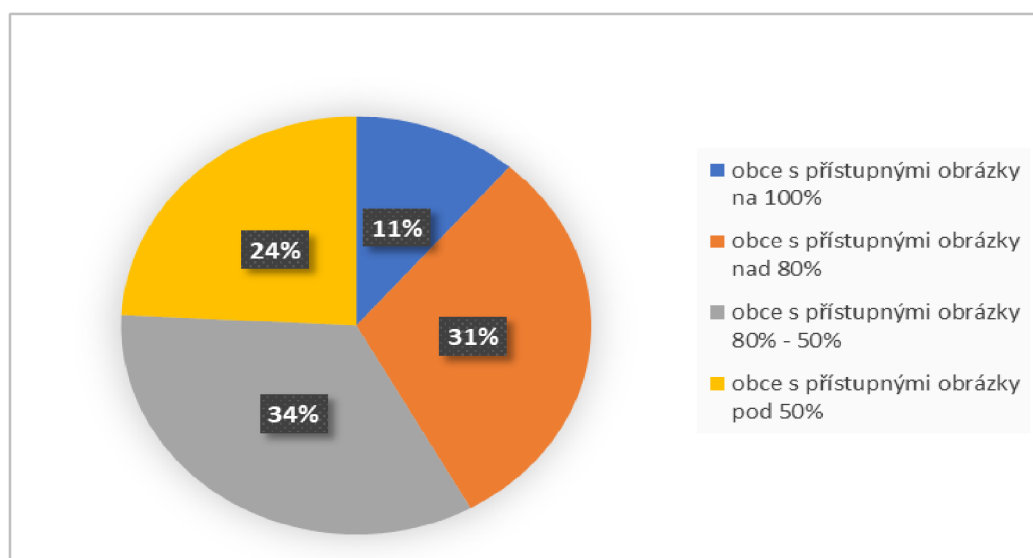
Zdroj: vlastní zpracování

#### 4.3.4 Přístupnost grafického obsahu

Analýza grafického obsahu, vychází ze sběru dat skrze, který došlo k získávání obrázku a jejich následné analyzování a rozřídění do 3 kategorií, přesněji na přístupné, které obsahují neprázdný alternativní text. Dále do kategorie bez alternativního textu, která zahrnuje prvky, které vůbec nemají atributy pro alternativní text definovány. Poslední kategorie zahrnuje grafické obrázky, které sice atribut pro alternativní text obsahují, ale buď je prázdný nebo nedošlo k jeho vědomému vyplnění a obsahuje klasické přípony pro grafický obsah jako jsou JPG a PNG. Tento kritérium je nastaveno z toho důvodu, že často jsou názvy souborů v nesmyslném tvaru, který je například generován automaticky skrze pořízení v rámci skenování, fotografování atd. Alternativní text by měl vyplňován vědomě s důrazem na obsah a z těchto důvodů je takový alternativní text vyhodnocen jako špatný. Nutno poznamenat, že v rámci sběru dat a prvotní analýzy za účelem efektivního sběru dat a jejich následného zpracování byla zpozorována dominance případů prázdného atributu nad variantou automaticky převzatého názvu. Z toho lze dedukovat celkovou převahu této varianty napříč webovými stránkami. Pro následné vyhodnocení byly 2. a 3. kategorie sloučeny do jedné, ve které jsou obsaženy nepřístupné soubory.

V rámci procesu sběru dat, bylo zjištěno a následně analyzováno celkem 3382 případů výskytu grafického obsahu. I přes snahu zajištění podobného počtu vzorků pro každou webovou stránku, se zde nachází odlišnosti v počtech získaných grafických prvků na stránku, kdy minimem je 16 prvků v rámci jedné webové stránky a maximem 108, průměrný počet zjištěných je 55 prvků na 1 webovou stránku. Průměrný počet přístupných souborů na webovou stránku v rámci zkoumaného souboru je 37 a nepřístupných 17. Celkem 7 webových stránek obstálo na 100 %, zatímco nejhorší zaznamenaná míra přístupnosti byla pouze 2 %. Celkem 2 případy dosáhly přístupnosti grafických prvků pod 10 %. Zajímavé jsou taktéž nalezené extrémy počtu výskytu celkového počtu grafický prvků v rámci jedné webové stránky, maximem přístupných prvků je 99, maximem prvků bez alternativního textu 20 a špatně vyplněných či prázdných alternativních textů bylo nalezeno v rámci jednoho případu 97. Což podtrhuje vysokou diverzitu, kterou taktéž znázorňuje níže zobrazený shrnující graf, pro jehož potřeby byly výsledky přístupnosti grafického obsahu rozděleny do 4 skupiny, kdy první je ideální 100% stav přístupnosti, následně skupina s uspokojivou přístupností v rozmezí 80 – 99%, následně skupina s výraznými nedostatky 50 – 79% a poslední skupiny pod 50%, kterou lze shledat za míru nedostatečné přístupnosti. Procentuální vyhodnocení je zvoleno z důvodu různorodého počtu získaného grafického obsahu na stránku, proto lze efektivně porovnat pouze míru přístupnosti získaných vzorků, které je však velice předmětná pro vyvozování výsledků.

*Graf 6 Vyhodnocení míry přístupnosti grafického obsahu*



Zdroj: vlastní zpracování

#### 4.3.5 Přístupnost souborů

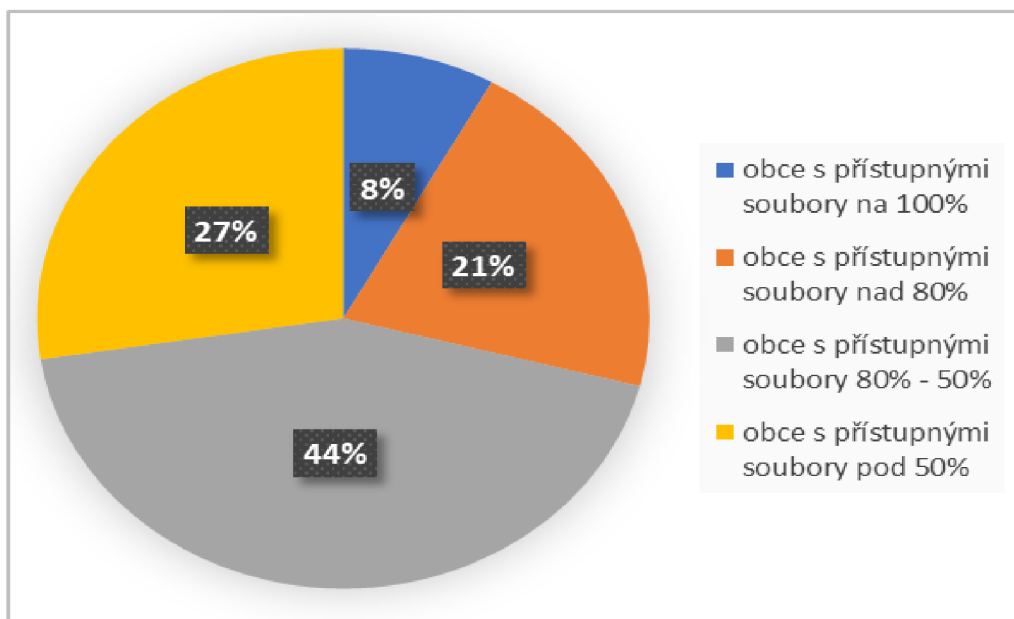
Analýza se věnuje i přístupnosti souborů, které jsou dostupné ke stažení v rámci webové stránky. Zajištění přístupností je předmětné hlavně s ohledem na množství publikovaného obsahu touto formou, kdy obce zveřejňují především různá oznámení v rámci úřední desky, vyhlášky, zákony, rozpočty a mnoho dalších informací spojených s působností obce, které mohou být důležité pro každého občana.

Jak již bylo zmíněno v předchozích kapitolách sběr dat a analýza je věnována dokumentům ve formátu PDF, který je by měl být pro publikaci na webových stránkách upřednostňovat a v praxi se jedná o nejčastěji využívaný formát. Při sběru došlo k analyzování a vyhodnocení, zda je soubor přístupný či nikoliv.

Skrze sběr dat bylo získáno a analyzováno celkem 905 souborů. Průměr počet souborů na jednu webovou stránku činí 15, maximálně jich bylo v rámci jedné webové stránky získáno 33 a minimálně 6. I přes odlišné extrémy, na rozdíl od problematiky přístupnosti grafického obsahu, nedochází k tak vysoké variabilitě získaného množství a z většiny webových stránek byl získán počet souborů blízký se průměru, v rozsahu 10–20 dokumentů se přesně jedná o 47 webových stránek. Nejvyšší, 100 % míry přístupnosti bylo dosaženo v 5 případech, naopak méně jak 10 % přístupnosti nebylo shledáno ani v jednom z případů, avšak v případě pod 20 % už se jednalo o 4 webové stránky, kdy nejnižší míra přístupnosti byla zjištěna 16 %. V rámci průměrů celého souboru bylo zjištěno 9 přístupných a 6 nepřístupných souborů na jednu webovou stránku. Nejvyšší počet dostupných bylo nalezeno v rámci jednoho případu 20 a počet nepřístupných 21.

Vzhledem k odlišnému počtu souboru získaných na jednu webovou stránku je hlavním ukazatelem procentuální zastoupení přístupných a nepřístupných souborů v rámci získaného vzorku. Celkové vyhodnocení znázorňuje graf níže, kdy byly po předchozím vzoru zvoleny 4 totožné procentuální skupiny. S ohledem na přístupnost grafického obsahu lze pozorovat podobné výsledky, u dokumentu je vidno drobné zhoršení v rámci plnění přístupnosti.

Graf 7 Vyhodnocení míry přístupnosti souborů



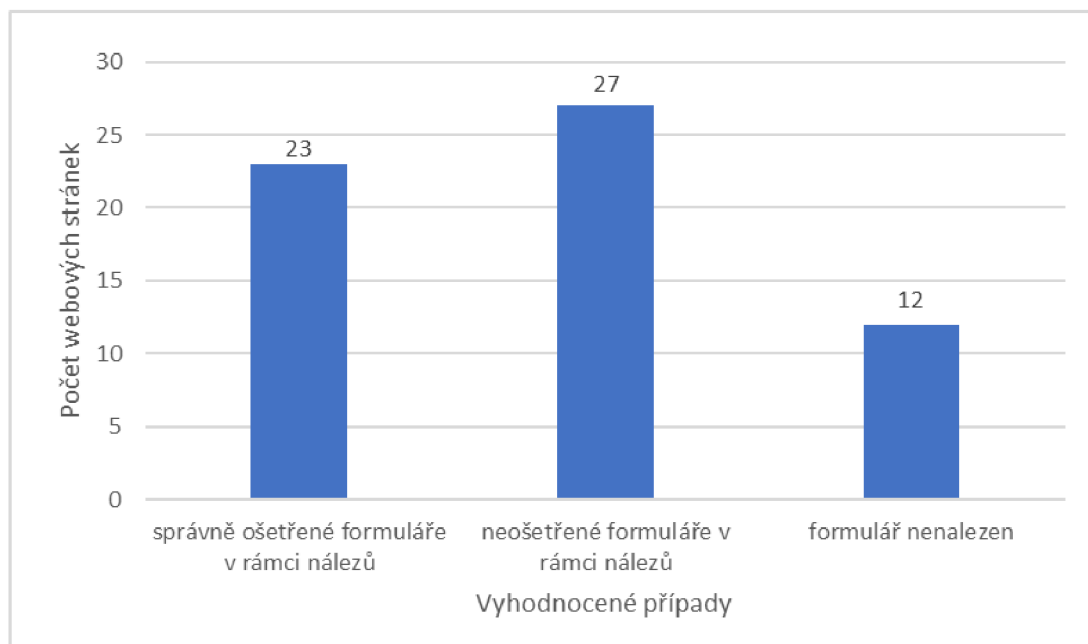
Zdroj: vlastní zpracování

#### 4.3.6 Formuláře

Vzhledem k povinnosti informování uživatele v okamžiku sběru na zásady o zpracování osobních údajů neboli GDP, byly analyzovány formuláře, které jsou častým nástrojem, který obce nabízejí návštěvníkům ke způsobu komunikace. Ze zákona takovýto formulář, musí být opatřen hláškou, odkazem, skrze který se mohou dostat uživatelé ke všem informacím, které by zpracované prohlášení mělo obsahovat.

V rámci sběru dat bylo nalezen, alespoň jeden formulář na 50 webových stránkách ze zkoumaných 62. Při sběru rovnou bylo vyhodnoceno, zda formulář obsahuje hlášku a odkaz. Nutno zmínit, že správnou variantou je i hláška, která je současně i odkazem, proto došlo ke sloučení variant, kdy nemusela být nalezena samotná hláška, ale za to byl nalezen odkaz. Ze získaných 50 formulářů bylo správně vybavených 23 a nesprávně ošetřených 27, kdy se jednalo především a totální absenci informací ve formuláři, který by měli vést na samostatné prohlášení. Graf níže prezentuje celkové výsledky v rámci analýzy formulářů.

Graf 8 Vyhodnocení ošetření formulářů v rámci GDPR



Zdroj: vlastní zpracování

#### 4.4 Uživatelské testování

Uživatelské testování je zaměřeno primárně na proces vkládání příspěvků na webové stránky a zjištění využitelnosti zákonem povinně zveřejňovaných informací. Předpokladem pro testování je, že obec disponuje webových stránkami, které jsou realizovány skrze redakční systém, který pověřený pracovník obce využívá pro vkládání obsahu. Lze předpokládat, že většina chyb pochází právě z této činnosti. Vzhledem k tomu, že pověřeným pracovníkem může být jakýkoliv občan z jakékoliv věkové kategorie s různým vzděláním a pracovním zaměřením, je tato různorodost zohledněna i při výběru participantů pro uživatelské testování.

V první části testování je pro účely uživatelského testování vytvořen prototyp, který simuluje nahrávání příspěvku v rámci redakčního systému. Vzhledem ke složitosti legislativy a požadavkům, které jsou na uživatele uplatňovány v rámci vkládání příspěvků je zvolen přístup moderovaného uživatelského testování, z důvodu dovysvětlení nejasností a lepšího pozorování uživatelů při testování. Cílem je ověření, zda opravdu dochází na straně uživatele k vyšší míře chybovosti.

Druhá část testování je zaměřena na využitelnost povinně umístěných informací a prohlášení stanovených zákonem. Pro tento účel bude vybrána reálná webová stránka, která splňuje všechny legislativní nároky. Cílem je zjistit, zda běžní uživatelé při hledání



informací skutečně využijí zákonem stanovené povinně zveřejňované informace nebo využijí k jejich dosažení jinou cestu. Dalším cílem je v rámci GDPR a Prohlášení o přístupnosti webových stránek zjistit, zda jsou uživatelé schopni informace vyhledat pro případné další využití.

#### **4.4.1 Dotazník na tvorbu obsahu**

Při uživatelském testování je nutné určit smysluplné oblasti testování. Problematika legislativních povinností je velmi rozsáhlá a v určitých aspektech vyžaduje specifické oborové znalosti z řad tvůrců webových stránek, právníků a dalších. Vzhledem k tomu byl vytvořen dotazník v rámci, kterého bylo náhodně osloveno celkem 132 obcí, které mají webové stránky od různorodých poskytovatelů, pro vyplnění dotazníku. Cílem dotazníku je ověření toho, kdo je autorem obsahu a kdo provádí úpravy v rámci dotazovaných sekcí. V potaz je brán pouze obsah, který zahrnuje především textové informace, nikoliv forma prezentace.

##### **4.4.1.1 Struktura dotazníku**

Dotazník je staven celkem ze 7 otázek, které se týkají 4 tematických okruhů, mezi nimiž jsou Prohlášení o přístupnosti webových stránek, Prohlášení GDPR, Povinně zveřejňované údaje a Rutinní úpravy. První 3 okruhy disponují 2 otázkami, na to, kdo je autem obsahem a dále kdo provádí následné úpravy. Poslední okruh je reprezentován pouze jednou otázkou, jejíž cílem je zodpovězení otázky, jakým způsobem obce provádějí rutinní přidávání obsahu na webové stránky.

Každá otázka disponuje stejnými 3 variantami odpovědí, které reprezentují 3 možné subjekty zapojené v procesu přidáváním obsahu. Jedná se o tvůrce webových stránek, který v rámci zakázky na tvorbu webových stránek vytváří obsah, případně jej v rámci poskytované podpory doplňuje. Dále je zpracování v rámci obce, kde může docházet ke samostatnému zpracování nebo upravování již stávajícího řešení. Poslední možností je zapojení 3. strany v rámci řešení potřebných problematik a zpracování obsahu. Tato možnost reprezentuje osobu nebo společnost, která do webových stránek zasahuje na popud obce, avšak nejedná se o nikoho z řad tvůrce webové stránky ani obce. Pro příklad lze uvést první společnost, která se může zabývat právě tvorbou obsahu, který je podmíněný zákonem.

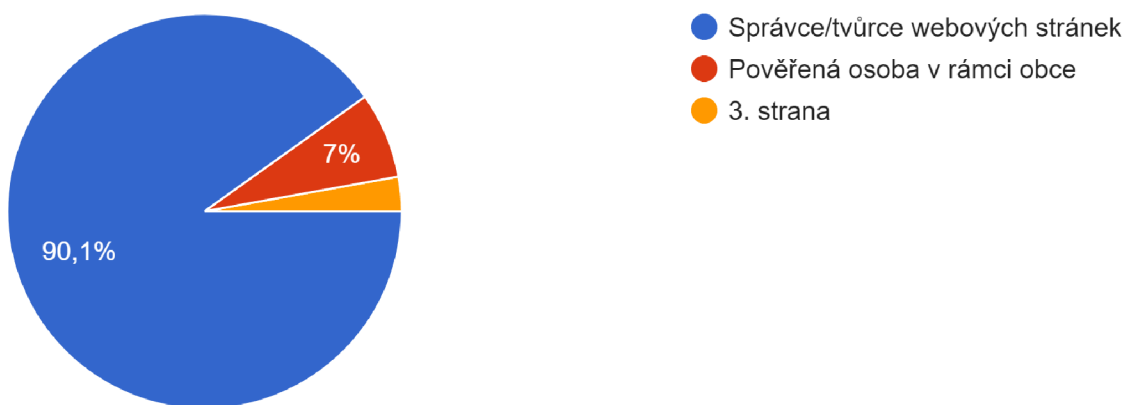
Za účelem eliminace nesprávných odpovědí vlivem špatného porozumění otázce, jsou v úvodu dotazník výše popsané informace respondentů vysvětleny, tj. zdůraznění toho, že otázky cílí pouze na obsah a vysvětlení významu odpovědí.

#### 4.4.1.2 Vyhodnocení dotazníku

K vytvoření dotazníku, sběru odpovědí a následnému vyhodnocení byl využit nástroj Google Forms, který nabízí přehledné prostředí, snadné zpracování a následné vyhodnocení dotazníku. Obdrženo bylo celkem 71 relevantních odpovědí, což činí návratnost dotazníků přibližně 54 % ze všech oslovených.

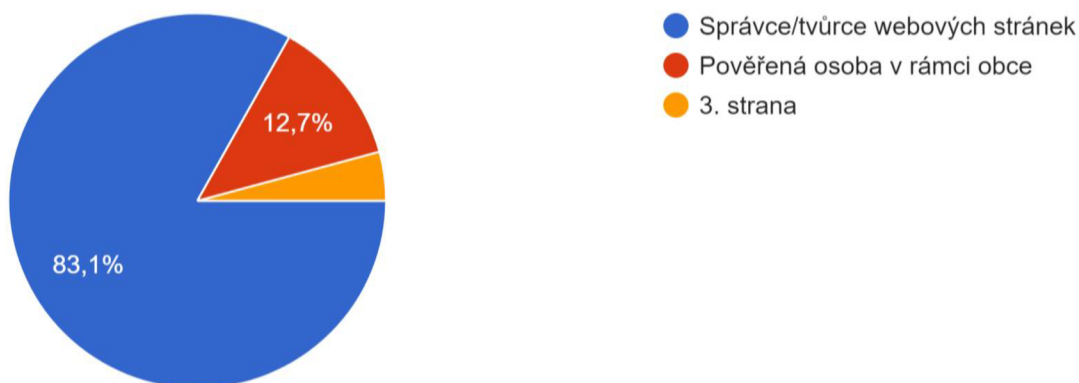
Na základě odpovědí lze stanovit, že tvorba prohlášení o přístupnosti webových stránek a jeho následné úpravy jsou prováděny v drtivé většině případů na straně tvůrce webových stránek. Lze to přisuzovat faktu, že toto prohlášení úzce souvisí nejen s obsahem webu, ale i se způsobem tvorby, která by měla být realizována v souladu s WCAG nebo jinou metodikou zaměřující se na tvorbu přístupných webových stránek. V oblasti úprav lze pozorovat zvýšené procentu na straně obce, které lze vysvětlit například změnou kontaktních údajů v rámci prohlášení, případně zpracování drobných změn, avšak režii zůstává především na straně tvůrce.

Graf 9 Autor prohlášení přístupnosti



Zdroj: vlastní zpracování

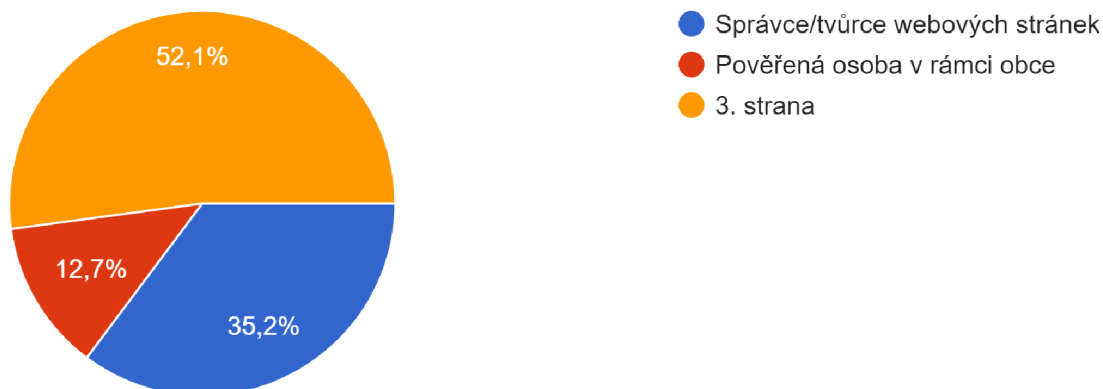
Graf 10 Editor prohlášení přístupnosti



Zdroj: vlastní zpracování

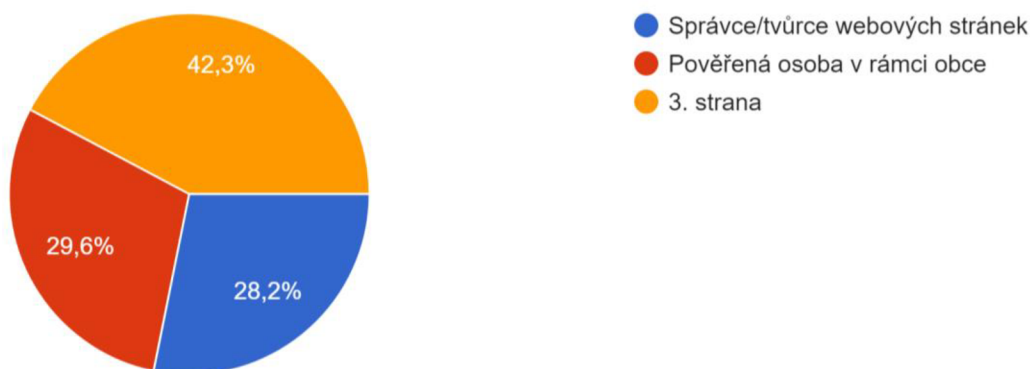
Další okruh je věnován prohlášení GDPR neboli zásad o ochraně osobních údajů. Z odpovědí je patrné, že se tvorbou ve více jak polovině případů zabývají subjekty 3 strany. Následuje zpracování tvůrcem webových stránek v rámci 35,2 % případů a ve 12,7 % realizují toto prohlášení samotné obce. Výsledek, lze přisoudit faktu, že samotné prohlášení v sobě obsahuje jak informace týkající se webových stránek, tak i informace týkající se vnitřních procesů v rámci obecního úřadu zaměřených na zpracovávání osobních dat, jejich ochranu, uložení atd. Tato problematika má velký přesah mimo prostor webových stránek a dotýká se vícero legislativních témat a právě proto, je předmětné oslovit pro zpracování osobu či společnost, která se touto problematikou zabývá převážně z právního hlediska. V rámci následných úprav lze opět pozorovat zvýšení zásahů ze strany obce, které je zde až 29,6 %, přibližně rovnoměrně na úkor ostatních zpracovatelů.

Graf 11 Autor obsahu prohlášení GDPR



Zdroj: vlastní zpracování

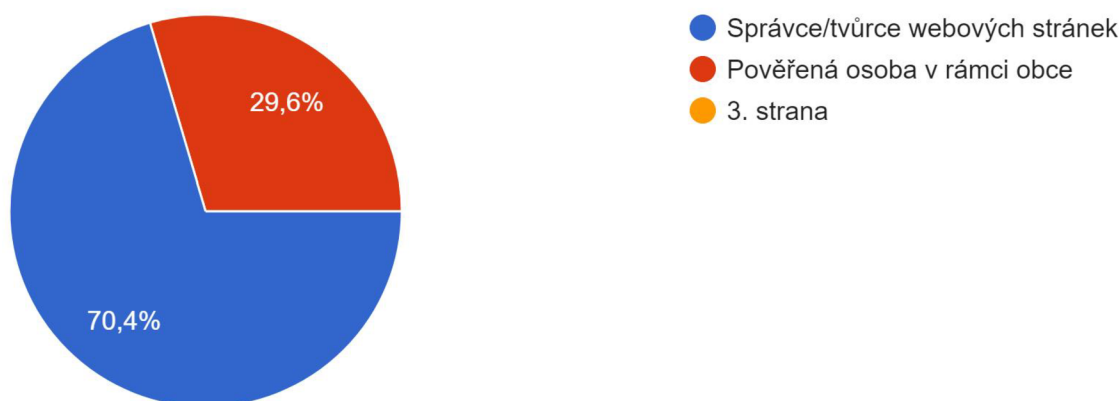
Graf 12 Editor obsahu prohlášení GDPR



Zdroj: vlastní zpracování

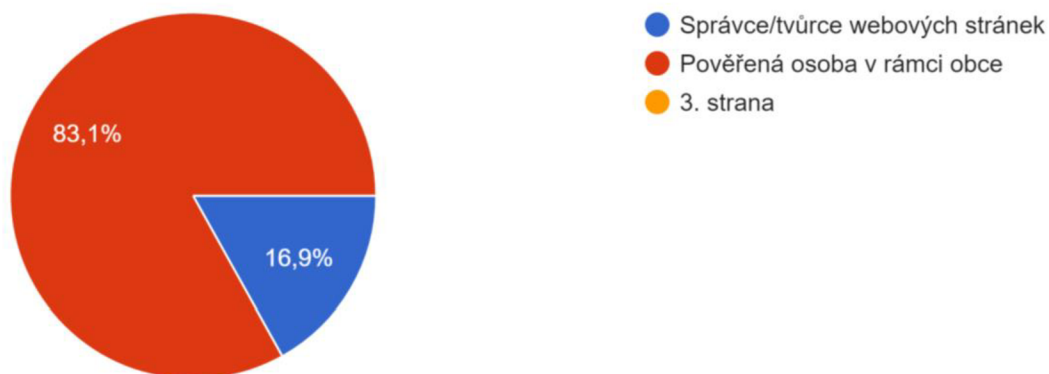
Následující okruh je zaměřen na zpracování povinně zveřejňovaných údajů. Zde dochází k zajímavému pozorování na rozdíl o předešlých okruzích, opět se opakuje situace, kdy tvorbu zajišťuje především tvůrce webových stránek a to v 70,4 % případů, avšak nutno zmínit již větší zapojení obce v rámci prvotní tvorby, přesně v 29,6 %. Odlišností je však samotné zpracování, kdy doslovně dochází k záměně preferovaného způsobu zpracování, to v 83,1 % přejímá obec a pouze 16,9 % provádí tvůrce webových stránek. Výsledek lze přisoudit faktu, že ze strany zákona je převážně udána struktura povinně zveřejňovaných údajů a dále specifika, jaké údaje by měly být zveřejněny. Tato problematika již na podmíněna tvorbou webových stránek ani dalšími procesy, které je zapotřebí dále posuzovat a zveřejňovat s ohledem na příslušné zákony. Zde se již jedná o běžný obsah spojený s působností obce, který může být často měněn a doplňován, například kontakty, vyhlášky, zákony a další informace, jež musí obec publikovat.

Graf 13 Autor obsahu povinně zveřejňovaných informací



Zdroj: vlastní zpracování

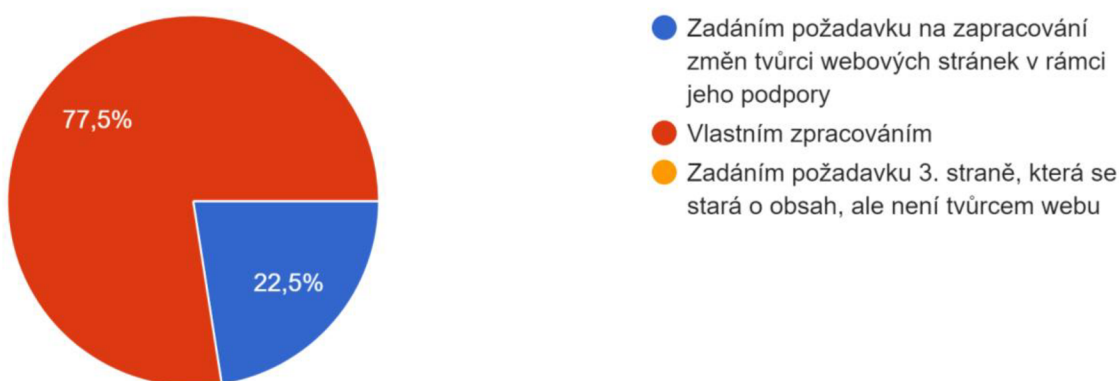
Graf 14 Editor obsahu povinně zveřejňovaných informací



Zdroj: vlastní zpracování

Poslední část je zaměřena pouze na rutinní úpravy obsahu webových stránek, pro demonstraci lze zmínit například publikování aktualit, přidávání nových oznámení a dokumentů na úřední desku, změnu kontaktních údajů, zveřejňování upozornění atd. Z výsledku je patrné, že tyto úpravy již obce ve většině realizují sami, přesně 77,5 % ze všech tázaných a v 22,5% obce preferují způsob zpracování skrze tvůrce webové stránky a jeho podporu. Výsledek může být odrazem již dnes standardního poskytování redakčního systému, který uživatelům umožňuje snadnou editaci obsahu.

Graf 15 Způsob přidávání/vytváření rutinního obsahu na webové stránky



Zdroj: vlastní zpracování

Následné uživatelské testování je zaměřeno přímo na pozici uživatele, tedy pověřeného pracovníka úřadu, který zpracovává a publikuje informace na webové stránky. Vzhledem k výsledkům dotazníků a náročnosti na znalosti ohledně dané problematiky je bezpředmětné zařadit zpracování prohlášení o přístupnosti webových stránek a GDPR, na

základě faktu, že pro tuto problematiku obce využívají především externích řešení, která poskytují odborné zpracování. Taktéž je bezpředmětné zařadit správnost zpracování povinně zveřejňovaných údajů, i přestože, že je zde větší zapojení obcí při potřebné tvorbě, tak je stále velká preference v rámci odborného zpracování na straně tvůrce. Vzhledem k výsledkům lze předpokládat, že tyto 3 okruhy by měly být v rámci předchozího algoritmického ověřování a následné analýzy na vysoké úrovni správnosti zpracování.

S ohledem na výsledky dotazníku se bude uživatelské testování soustředit především na přidávání rutinního obsahu v rámci webových stránek a následně použitelnost zákonem povinně zveřejňovaných informací, prohlášení atd.

#### **4.4.2 Testování vkládání příspěvků**

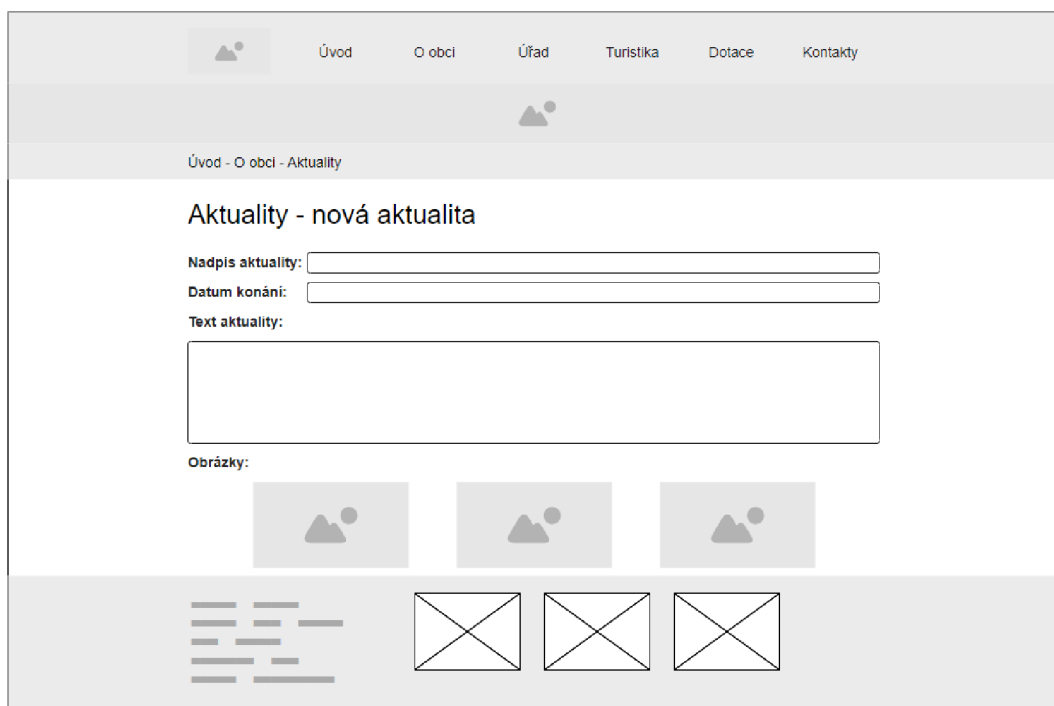
Vkládání nových příspěvků a různorodého informačního obsahu je nejčastější akcí, kterou provádí uživatelé pověřeni ze strany obce. Nejčastěji se jedná o textové informace doplněné o obrázky a dokumenty, které musí být vloženy v rámci povinného zveřejňování. Z hlediska legislativních povinností se jedná nejčastěji o požadavky na povinné zveřejňování a přístupnost obsahu webové stránky. Pro detailnější zjištění toho, co všeho v rámci legislativních požadavků obce zapracovávají byl vytvořen dotazník, který následně použit jako podklad při stanovení cílů a první fáze uživatelského testování, které je zaměřena na stranu editora.

#### **4.4.3 Prototyp**

Vzhledem k zajištění co nejautentičtějšího způsobu přidávání příspěvků bylo zapotřebí zajistit uživatelům rozhraní redakčního systému, který se nejvíce blíží v praxi často využívaným řešením, které jsou realizovány většinou v rámci uzavřených komerčního redakčních systému firem, jež nabízejí tvorbu webových stránek. Z tohoto důvodu byla vybrána platforma WIX, které nabízí nástroje pro tvorbu webových stránek včetně redakčního systému. V rámci nekomerčního využití lze využít základní verzi zdarma.

V rámci autentičnosti jsou vytvořeny 2 jednoduché stránky, které simulují přidání aktuality a oznámení v rámci úřední desky. Obě stránky obsahují přidání textových informací s možností úpravy barev a vzhledu, ale taktéž přidání obrázků a souborů. Pro účel vytvoření odpovídajícího prototypu byly využity wireframy.

Obrázek 2 Wireframe testovacího prototypu



Zdroj: vlastní zpracování

#### 4.4.4 Testování hledání obsahu

Druhá část testování se zaměřuje na vyhledávání konkrétního obsahu na reálné webové stránce. Pro potřeby testování je sestaven otevřený testovací scénář. Na uživatele jsou kladeny úkoly pro nalezení specifického obsahu v rámci zadaného úkolu. Uživatel není nikterak naváděn a nalezení obsahu je v jeho režii. Při následném nalezení pouze uvidí URL adresu, na které obsah našel. Nalezením obsahu je pouze vedlejším cílem. Hlavním cílem je jakým způsobem uživatel obsah našel a zda u toho uplatnil zákonem povinně zveřejňovaný obsah.

#### 4.4.5 Průběh testování a zvolené metody

Uživatelské testování probíhalo v rámci prvků moderovaného přístupu. Důvodem využití moderátora je případné pomoc s navigací v rámci první části testování, kdy mohou uživatelé tápat s nalezením funkcí pro přidávání a úpravu obsahu, avšak vzhledem k tomu, že cílem není otestování použitého redakčního systému pro potřeby testování. Nutno zdůraznit, že cílem není otestování použitelnosti reakčního systému, který je pro potřeby testování využit. Dalším důvodem využití moderátora je rozdělení uživatelů do 2 skupiny,

kdy jedna skupiny od moderátora obdrží veškeré potřebný znalostní základ v rámci přidávání obsahu na webové stránky s ohledem na legislativní nároky, zatím co druhá skupiny bude provádět tyto úkony bez těchto znalostí. Účelem zohlednění, že uživatele mohou být školeni či nikoliv.

Pro potřeby využit záznam, který byl důležité pro odhalení chování, respektive postupů, které uživatel využíval pro dosažení informací zejména v druhé části testování. Z výsledných záznamů byl stanoven hlavní způsob nalezení požadované informace.

#### **4.4.6 Analýza získaných dat**

V rámci testování byla shromážděna data v 3 základních typech. Prvním typem jsou upravené stránky v rámci přidávání obsahu. Zde došlo k rozboru obsahu, který uživatel na stránku přidal, především obrázků a souborů.

Dalším je doplněný scénář, který uživatelé obdrželi před testováním. Scénář sloužil zároveň i jako forma dotazníku, kde uživatele vyplňovali odkazy, kde našli požadované informace, případě doplnili, že požadované nenašli. Tento typ spíše plnil ověřovací funkci toho, co uživatelé splnili a co nikoliv.

Posledním typem dat je samotný záznam, který dále sloužil pro analýzu a ověření postupů, které uživatel v rámci plnění aplikoval. Záznam byl stěžejní pro odhalení postupů v druhé části testování za účelem ověření, zda uživatele skutečně naleznou povinný legislativní obsah a aplikují jej pro vyhledávání informací.

Všechny požadované vstupní informace byly dále zapsány do tabulky v rámci aplikace Microsoft EXCEL a vyhodnoceny v rámci vhodně zvolených ukazatelů. Mezi ty byl stanoven poměr uživatelů v rámci řešení. Procento využití legislativních údajů.

#### **4.4.7 Výsledky testování**

Výsledky první části testování jasně poukazují jasnou závislost na znalost uživatelů v rámci legislativních nároků na vkládaný obsah. První skupina uživatelů, která neobdržela žádné informace a v rámci nároku na vkládaný obsah vykazovala vysokou míru chybovosti. Ze 6 uživatelů vložil správný přístupný obrázek, který bylo zapotřebí doplnit v rámci redakčního systému o alternativní text a název pouze 1, zatímco ve druhé skupině byla situace opačná, avšak docházelo naopak k pouze částečnému vyplnění, výsledkem tedy bylo, že 1 uživatel obrázek vložil v nepřístupné podobě, 2 v částečně přístupné podobě, kdy doplnili pouze název a 3 byly přidány správně.



Podobné tomu bylo i u vkládání dokumentu v rámci přidání oznámení na úřední desku. Zde uživatelé měli na výběr ze 4 dokumentů, kdy 2 byly nevhodné obrázkové formáty, avšak jeden uložen v rámci formátu JPG a druhý ve formátu PDF. Dalším formátem byl již částečně vhodný DOCS, který však klade případnou bariéru v rámci nutnosti jeho stažení a otevření specializovanou aplikací. Nelze považovat vložení úplně za chybu, avšak zcela správné by mělo být vložení skrze čitelné PDF, které pro prohlížení neklade zbytečné další bariéry. I přesto, že se uživatelé nevyužili pro vložení formát JPG, tak v první skupině 2 figuroval nepřístupný PDF dokument a celkově první skupina opět vykazovala velké množství chyb. V souhrnu obou skupiny bylo využito 2 obrázků v formátu PDF, následně byl 4 využito formát DOCS a nejvíce byl využito čitelný soubor ve formátu PDF.

Pro druhou část testování nebyly ani jedné skupině poskytnuty žádné dodatečné informace, které by mohli ovlivnit způsob provedení úkolů. Z výsledků je jasně patrné, že většina uživatelů se uživatelé spoléhaly primárně na navigační prvky, převážně menu, než že by informace hledali na sjednocující stránce Povinně zveřejňované údaje, kterou obec musí být a jež většina informací byla obsažena zde. Například v prvním případě, kdy uživatelé měli najít úřední desku bylo procento využití Povinně zveřejňovaných údajů nulové. Celkově průměrné využití v rámci testování bylo zhruba 30 %. V rámci poslední dvou otázek měli uživatelé najít informace o zpracování jejich osobních údajů a informace o přístupnosti webových stránek. V prvním případě nebyl specifikován přímo název prohlášení nebo známá zkratka, ale pouze případné, kde se s tím mohou setkat z důvodu o toho, zda uživatelé dokážou tyto informace samostatně najít. Prohlášení GDPR, na které je povinně odkazováno z formuláře našlo 9 uživatelů a 3 jej nemohli najít. Prohlášení přístupnosti našlo uživatelů 7 uživatelů z 12.

## 5 Zhodnocení výsledků a doporučení

### 5.1 Výsledky analytické části

Z výsledků analytické části je jasně patrné, že v praxi nelze dosáhnout stavu 100% plnění zákonem daných povinností. Lze pozorovat jistou míru korelace úspěšnosti řešení v rámci odborného či neodborného zpracování a dále skrze míru absolutních požadavků ze strany legislativy. Tento vztah lze prezentovat na stránce Povinně zveřejňované údaje, která má zákonem jasně stanovenou strukturu a nelze se tak od ní volně odchylovat. Zároveň zde dochází k velké míře prvotního odborného zpracování ze strany tvůrce webových stránek, což bylo ověřeno skrze dotazník. Díky těmto aspektům zde dochází k nejnižší míře chybovosti, kdy 71 % zkoumaných stránek bylo vyhodnoceno bez chyby, dalších 8 % pouze s 1-2 nedostatky a dalších 10 % s 3-5 nedostatky v rámci 30 ověřovaných kritérií.

Zvyšující se chybovost přichází v případě přípustné volnosti formulace, který reprezentuje Prohlášení o přístupnosti webové stránky. Bezchybně v rámci analýzy prošlo 31 % webových stránek a s 1-2 chybami dalších 41 %. Dalších 13 % realizoval prohlášení s chybovostí 3-5 bodů z celkových 17. Nutno zmínit, že nejčastějším výrazem, který nebyl v rámci analýzy nalezen je „nepřístupný obsah“ a to přesně v 38 případech. Tomuto by měla být věnována samostatná sekce v rámci prohlášení, avšak většina webových stránek, které neprošly zvolila způsob řešení skrze drobného zmínění v rámci obecného textu. Toto řešení vzbuzuje diskusi o naplnění informační podstaty. Vzhledem k této povaze je tento stav považován za chybný z hlediska algoritmického vyhodnocení. Avšak z hlediska detailnější analýzy lze považovat pouze za částečný prohřešek.

Kde však značně klesá úspěšnost plnění legislativních podmínek jsou zejména v případech, dochází ke zpracování obsahu běžným uživatelem neboli pověřenou osobou obce. Jedná se zejména o problematiku přístupnosti, a to jak grafický prvků, tak souborů, které často obsahují důležité informace a měl by na ně být kladen velký důraz. Grafický obsah taktéž nelze nechat bez povšimnutí, protože v rámci analýzy byly často shledány případy, kdy byl nepřístupný obrázek ve formě plakátu vyvěšen skrze aktualitu, jejíž účelem bylo informování o nadcházející akci. Návštěvník odkázaný na asistivní technologie není schopen z takto přidaného obsahu získat jakoukoliv informaci. Nabízí se možnost využití alternativního textu, ale ten je většinu použit pouze obecný popis obrázků a neposkytuje konkrétní informace, které jsou v takovémto případě žádoucí. Z výsledků

analýzy je jasně patrné, že největší nedostatky bývají právě v rámci této problematiky. Co se týče grafického obsahu, tak pouze 32 % zkoumaných webových stránek disponuje více než 80 % mírou přístupnosti grafického obsahu v rámci získaného vzorku a u souboru je to dokonce 28 %.

Dále je nutné zmínit případy, kdy dochází k vysoké úrovni volnosti ve formě zpracování. Jedná se konkrétně o případ prohlášení o zásadách zpracování osobních údajů neboli GDPR. Legislativa vymezuje, jaké údaje by měly být zmíněny, avšak nestanovuje ani přibližnou rámcovou strukturu zpracování. Jedním z důvodů je fakt, že se rozsah zpracování a obsažené informace mohou značně u jednotlivých obcí lišit v souvislosti s nastavením vnitro obecních procesů, kterých se GDPR týká. Při sestavování vhodného algoritmu nebylo možné dosáhnout takové realizace, která by férově a alespoň uspokojivou mírou chybovosti ověřovala prohlášení nezávisle na formě zpracování. Další roli zde může hrát zjištění skrze dotazníkové šetření, a to sice vysoká míra zpracování obsahu 3. stranou, jde o 52,1 % případů. Tento fakt přináší do řešení ještě větší míru variability, která znesnadňuje možnost algoritmického ověření v rámci zvoleného způsobu řešení.

## **5.2 Výsledky uživatelského testování**

Úvodu testování je věnováno dotazníkové šetření, skrze které bylo osloveno náhodně vybraných 132 obcí, ze kterých 71 dotazník vyplnilo. Výsledky dotazníku poukazují na to, jak obce přistupují k řešení jednotlivých legislativních nároků, k čemuž využívají převážně tvůrce webových stránek, případně odbornou 3. stranu, a to především v rámci zpracování GDPR a Prohlášení o přístupnosti webových stránek. Obce samostatně spíše přistupují ke správě běžného obsahu, která se týká zejména problematika plnění přístupnosti.

Testování prokázalo jasnou souvislost s informační základnou, kterou uživatelé disponují v rámci problematiky legislativních povinností. 1. skupina, které nebyl poskytnut žádný informační podklad vykazala skokově větší míru chybovosti na rozdíl od druhé skupiny, které bylo poskytnuto simulované školení, především v rámci přístupnosti.

Druhá část testování se dodatečně zaměřila na použitelnost a povinně zveřejňovaného obsahu. Nutno zmínit, že většina obsahu byla obsažena na stránce Povinně zveřejňované údaje, kterou však uživatelé využili průměrně pouze ve 34 % případů při hledání tázaného obsahu.

### 5.3 Dedukované shrnutí výsledků

S odkazem na provedou analýzu, dotazníkové šetření a uživatelské testování lze určit hlavní rizikovým bodem pro plnění legislativních povinností běžného správce obsahu z řad obcí, který nedisponuje odbornými znalostmi. V případech, kdy roste míra vlastního zpracování obsahu v rámci obce, roste i míra chybovosti v plnění legislativních povinností skrze týkající se obsah. Tento fakt je patrný v případě přidávání běžného obsahu, na který jsou kladeny nároky v rámci zachování jeho přístupnosti. Výsledky dotazníku ukazují, že obce preferují vlastní zpracování toho obsahu. Zatímco výsledky analytické části jasně ukazují, že uspokojivé míry 80 % a vyššího plnění přístupnosti v rámci grafického obsahu dochází pouze ve 32 % případů a u souborů pouze u 28 % případů.

Jako další rizikový bod lze označit míru volnosti zpracování legislativně povinného obsahu. Z výsledků je patrné, že v případě jasně daných povinností a přístupných vzorů, dochází k menší míře porušování obsahových povinností či částečné až nevhodné interpretaci. Zde se lze odkázat na zpracování stránky Povinně zveřejňované údaje, kdy míra úspěšného plnění dosáhla 71 % v rámci testovaných kritérií, nelze opomenout, že ve většině ostatních případů se jednalo spíše o drobné chyby. Pokud se na problematiku zaměříme z opačné strany. Lze zmínit prohlášení ohledně GDPR, jehož variabilita zpracování dosahuje takové úrovně, kterou lze obtížně skrze algoritmické ověřování vyhodnotit.

Dalším rizikovým bodem je kvalitu zpracování na straně tvůrce webových stránek, který by měl poskytovat odborné řešení a poradenství v rámci všech povinností. Dále by měl poskytovat vhodně navržený redakční systém, který uživatelům z řad obcí umožňující snadnou správu obsahu a plnění legislativních povinností. Avšak tato práce není zaměřena na analýzu nabízených řešení v rámci tvorby obecních webových stránek. Nelze tak s určitostí říct, jak velký dopad mají současná řešení redakčních systémů, které obce reálně využívají na úspěšnost plnění legislativních nároků.

Skrze druhou část uživatelského testování, kdy průměrné využití shrnující stránky Povinně zveřejňované údaje bylo pouze 34 % lze uvažovat o smyslu povinného obsahu z hlediska běžného návštěvníka. Pokud se pozastavíme právě u povinně zveřejňovaných údajů, tak při větším povědomí mezi uživateli se může jednat o užitečnou stránku, na které najdou všechny potřebné údaje, případně je na ně nasměrují. Výsledky uživatelského testování mluví o opaku. Běžný návštěvník si použitým názvem nespojí obsah, který je na této stránce shrnut. Přínos této stránky lze hledat především pro správce obsahu, který

v rámci povinností musí zpracovat veškerý povinný obsah, který si v rámci přehledné struktury může zkontrolovat, zda je na webu obsažen, protože velká část obsahu je tvořena skrze odkazy vedoucí na další podstránky, které naplňují detailní informační podstatu.

V případě prohlášení lze opět mluvit hlavně o přínos, který vytváří tlak na tvůrce webových stránek a správce obsahu, aby byl obsah přidáván ve správném formátu. Při sběru uživatelských dat jiných činností bylo měl uživatel, poskytující data, přehled o tom jak je s nimi nakládáno a byl jasně informován jak v různých situacích může postupovat, což opět zvyšuje transparentnost a tlak na bezpečnosti v této oblasti.

## **5.4 Doporučené postupy pro eliminaci nedostatků**

### **5.4.1 Obecná doporučení**

Zásadním doporučením je poskytnutí zajištění relevantních a srozumitelných informací a školení pro pověřené osoby z řad obcí, které mají na starost správu obsahu na webových stránkách. Z hlediska zisku informací lze zvolit 3 přístupy, a to studiem metodiky vydaných příslušnými státními orgány. Výhodou je bezplatné obdržení informací, avšak metodiky jsou spíše cílené na tvůrce webového obsahu a skrze jejich složitost nemusí být všechny informace pro laika srozumitelné. Dalším způsobem je obrácení se na tvůrce webových stránek, který by měl všemi potřebnými znalostmi disponovat a měl by být schopen je předat spolu s informacemi ohledně ovládání redakčního systému formou školení. Nevýhodou toho řešení je cena, avšak velikou výhodou je osobitý přístup a možnost praktické ukázky. Dalším způsobem je kontaktování osob či společností, které samostatně problematikou zabývají, výhody a nevýhody jsou prakticky totožné s kontaktování tvůrce webových stránek.

Pokud obce nejsou schopny zajistit správnost přidávání povinné obsahu, lze doporučit využívání podpory ze strany tvůrce webové stránky, pokud ji nabízí. V současnosti je technická podpora standardem většiny firem, které tvorbu webových stránek nabízí. Technik na straně podpory, by měl disponovat všemi potřebnými znalostmi, které by měly zajistit správné přidání obsahu. Avšak obce by měly mít na paměti, že za přidaný obsah stále zodpovídají ony a nikoliv technik, který obsah přidal.

## 5.4.2 Konkrétní doporučení

Jako největší problém lze shledat nepřístupnost přidávaných souborů, které mohou plynout z případu, kdy obce takovýto soubor obdrží. Jako první lze doporučit uživatelsky snadnou kontrolu přístupnosti, a to pokusit se skrze kurzor a levé tlačítko myši vybrat jakoukoliv část textu v souboru a případně se jí pokusit zkopírovat. Tento metoda je sice velice primitivní a nelze se na ní úplně 100 % spolehnout a pro běžného uživatele je snadno pochopitelná a rychlá. Další metodou je využití aplikací, které slouží přímo pro ověření přístupnosti souborů, například lze uvést volně přístupnou aplikaci PAC 2021. Případně je možné v rámci souborů PDF využít algoritmus navržený v rámci této práce, avšak pro praktické využití je zapotřebí současné řešení doplnit o grafické rozhraní a výstup prezentovat skrze spustitelný soubor.

Ověřovací fáze je však jen polovina procesu, při zjištění nepřístupného souboru je zapotřebí soubor transformovat do přístupného formátu. V případech, které se týkají pouze krátkého textu, lze uvažovat o jeho přepsání, buď do nového souboru anebo rovnou do webové prezentace. Více žádoucí variantou je automatizované převedení souboru do čitelné podoby. K tomu je opět zapotřebí využít aplikace, které automatizovaně nepřístupnou verzi převedou na přístupnou. Aplikací lze najít nepřeberné množství, avšak pro potřeby obcí lze doporučit primárně řešení nabízené Portálem veřejné správy, které je dostupné na adrese „[anonymizace.gov.cz/crossroad](https://anonymizace.gov.cz/crossroad)“. Tento nástroj je určen pro orgány veřejné moci, mezi které se řadí i obce. Stačí se pouze přihlásit skrze datovou schránku a dle pokynů nahrát nepřístupný soubory. Webová stránka nabízí přehledně vysvětlený postup a případné nástroje pro případné upravení souboru.

U grafických prvků, nejčastěji fotografií lze doporučit osvojení postupu při nahrávání takového obsahu na webové stránky. Při analýze redakční systémů pro testování bylo ve všech případech umožněno při nahrávání souboru doplnit všechny potřebné údaje, tj. název, alternativní text atd. Lze předpokládat, že komerční redakční systémy, které jsou primárně využívány, tyto možnosti taktéž nabízejí s ohledem na skutečnost, že se firmy na taková řešení specializují. Téma přístupnosti je žádoucí v rámci každého webové stránky. Na základě těchto předpokladů lze stanovit unifikovaný postup, který je shrnut v následujících větách. Začínáme nahráním obrázků na webové stránky. Po nahrání vyhledáme pole Název a doplníme název fotografie, který by měl vyjadřovat zobrazovanou tematiku nebo objekt, který je dominantní v rámci fotografie. Následně vyhledáme pole

alternativní text a doplníme 1 větu, která jasně shrne, co lze na obrázku vidět. Pokud je v rámci přidávání fotografie nabízeno pouze 1 pole pro zadání textu, vždy upřednostněte doplnění popisné věty, před názvem.

V případě zaměření na jednotlivá prohlášení, lze doporučit postupovat zejména dle vydaných aktuálních metodický pokynů s odkazem přímo na související legislativu. Na místě je vyžití vzorových řešení vydaných státní správou. S ohledem na výsledky ověřování plnění prohlášení o přístupnosti webových stránek, lze doporučit členit obsah do souvisejících kapitol, a nikoliv jej soustředit do velkých shrnujících odstavců, které většinou plně nesplňují informační povinnost k daným bodům.

Dále lze doporučit tvůrcům webových stránek, aby každou novou legislativní povinnost či změnou ve stávajících konzultovali s právně zaměřenou osobou či advokátní kanceláří. Důvodem je zajištění odborného právního přístupu, který z hlediska pouhého technického zpracování nelze 100 % naplnit.

## **5.5 Zhodnocení použitých metod**

Při volbě metody zpracování byl hlavní předpokladem vysoká míra podobnosti v rámci zkoumaných bodů. Závěry analýzy legislativních požadavků taktéž naznačily, že by se v rámci povinností mělo jednat o velice podobná zpracování, které lze skrze zpracovat skrze jednoduché algoritmičké zpracování. Při tvorbě algoritmů a jejich optimalizování se následně ukázalo, že použitá metoda není úplně vhodná pro všechny potřebné skupiny zkoumání. Jedná se zejména o ověřování textového obsahu s vysokou variabilitou zpracování, konkrétně prohlášení GDPR, jehož vysoká variabilita zpracování neumožňovala efektivní ověření skrze zvolené řešení.

Výhodné se ukázalo být zejména zvoleného získávání dat, skrze web scraping, který se ukázal být velice účinný, a to zejména z důvodu snadného zisku dat, jejich následné filtrace, zpracování a uložení. V případě většího zaměření se na tuto problematiku, lze aplikovat i tzv. web crawler, který by zvedl efektivitu řešení, skrze automatizované sbírání dat napříč různými webovými stránkami. Vytvořený algoritmus sice řeší úlohy skrze zadaný seznam konkrétních webových stránek, ale nejedná se web scrawler v pravém slova smyslu.

Následná zpracování a analýza skrze web scraping se ukázala být vhodná především u homogenních prvků, především v případě testování přístupnosti grafický prvků a souborů, kdy současné řešení poměrně efektivně vyhodnocuje, zde jsou prvky přístupné či nikoliv.

Dále u ověřování přítomnosti povinně zveřejňovaných stránek a prohlášení. U ověřování ošetření formulářů. Uspokojivou míru využití lze sledovat u stránek Povinně zveřejňované údaje a Prohlášení přístupnosti.

Obecně lze shrnout, že web scraping je vhodnou metodou pro získání obsahu webové stránky, filtrace a uložení dat. Z pohledu úloh, které jsou založené na ověření přítomnosti různých atributů ve zdrojovém kódu anebo přesně daných frázích v rámci textového zpracování viz. povinně zveřejňované údaje, je zvolené algoritmické zpracování taktéž vhodné.

Pokud je zapotřebí zkoumat, porovnávat a ověřovat textový obsah s vysokou mírou variability, tak tato metoda začíná ztrácet na účinnosti. V takovém případě se již nabízí například zpracování skrze textové modely umělé inteligence, které by při kvalitním zpracováním a optimální míře naučení mohly efektivně vyhodnocovat správnost variabilního obsahu s určitou pravděpodobností.

## 6 Závěr

Teoretická část práce byla v první části zaměřena legislativní požadavky na webové stránky obcí, které jsou zde popsány s ohledem na podstatné informace. Rozsah zpracování je brán s ohledem na stanové cíle této práce, které jsou zaměřeny převážně komplexní pohled v rámci zkoumané problematiky. Zpracované tematické okruhy se vzájemně odlišují rozsahem a jejich složitostí, například problematiky přístupnosti a GDPR jsou velice obsáhlými a často jsou zpracovány jinými autory skrze detailní zaměření. Cílem zpracování bylo zjištění požadavků, které z jednotlivých témat legislativy vyplývají, tento cíl lze považovat za naplněný, skrze stanovená kritéria pro ověřování plnění v rámci praktické části. V teoretické části byly dále rozebrány metody zpracování dat a uživatelského testování s ohledem na stanové cíle. Tyto metody byly dále aplikovány v teoretické části.

Praktická část navázala na teoretickou stanovením kritérií, která byla využita pro proces ověřování plnění legislativních nároků. Následovalo navržení algoritmů, které byly implementovány skrze metodu web scrapingu. Takto byl zajištěn poměrně efektivní sběr dat jejich následné filtrace a uložení do formátu CSV pro následné vyhodnocení. V navržených algoritmech byly implementovány analytické části, které měly za cíl ověření plnění legislativní požadavků s ohledem na stanové cíle práce. Při optimalizaci algoritmů a



ověřovacích kritérií, vyšlo najevo, že zvolená metoda zpracování není optimální pro všechny případy zkoumání. V případech, kdy bylo ověřování závislé na zpracování atributů zdrojového kódu nebo se odvíjelo od jasně daných povinností, od který se nebylo možné odchýlit, se zvolená metoda ukázala jako poměrně efektivní a došlo k naplnění cíle práce na realizaci analýzy plnění skrze navržený algoritmus. Toto nelze říci v případech kdy je řešení aplikováno na složitější textové struktury, které mohou být zpracovávány s velkou mírou volnosti co se týče formátu zpracování. Zejména se jedná o prohlášení GDPR, kdy nebylo možné sestavit testovací kritéria spolu s algoritmem na takové úrovni, aby mohlo dojít k efektivnímu vyhodnocování napříč různými webovými stránkami. V případě automatizovaného přístupu se nabízí alternativa zpracování skrze textové modely umělé inteligence, které by při správné realizace mohly takto variabilní obsah vyhodnocovat s uspokojivou úspěšností. Z těchto důvodů lze cíl práce na algoritmičticky provedou analýzu považovat pouze za částečně naplněný.

Druhá část praktické části se věnovala uživatelskému testování, kterému předcházelo dotazníkové šetření. Účelem bylo zjistit, jakým způsobem dochází ke zpracování povinného legislativně podmíněného obsahu. Výsledky sloužily nejen k definici smysluplných cílů testování, ale taktéž jako podklad pro celkové výsledky práce. Z výsledků uživatelského testování bylo jasně patrné, že hlavním rizikovým faktorem při nedodržování legislativních nároků je zejména běžný uživatel v roli správce, který nedisponuje potřebnými znalostmi. Tento zjištěný fakt pouze podtrhuje výsledky předchozí analytické části, kdy se největší míra chybovosti vyskytuje právě v případech běžně přidávaného rutinního obsahu. Jde především o otázku přístupnosti v rámci přidávaných souborů a grafického obsahu. S ohledem na výsledky analytické části, dotazníkového šetření a uživatelského testování byly v závěru praktické části stanovena obecná a konkrétní doporučení, která cílí především na správce obsahu ze strany obcí. Byly stanoveny postupy pro přidávání obsahu a navrhnuté způsoby, jak docílit plnění legislativních požadavků, a to zejména s ohledem na zajištění přístupnosti a zlepšení obsahové části jednotlivých prohlášení.

Při zpracování se ukázalo, že jednotlivé problematiky legislativních požadavků jsou velice obsáhle a stálo by za to se jim zabývat jednotlivě v rámci detailních zpracování. U zpracování textového obsahu, který má vysokou míru variability ve formě zpracování, lze uvažovat o zapojení textových modelů umělé inteligence pro efektivní analýzu obsahu povinných informací, protože cílem zvolená realizace se ukázala být v tomto případě

neefektivní. Dalším zjištěním, na které je zapotřebí poukázat je i měnící se legislativa s odkazem na strukturu povinně zveřejňovaných údajů, která má v současné chvíli 14 bodů, avšak před poslední úpravou v průběhu roku 2021 měla bodů 17. Lze uvést i nové povinnosti v rámci zákona o ochraně oznamovatelů, které však vyjedou v platnost až v průběhu prosince 2023 a proto jej nebylo možné zahrnout do zpracované analýzy. Způsob a interval měnící se legislativy může taktéž ovlivňovat současný stav plnění legislativních požadavků s ohledem na různou reakční dobu obcí pro zavedení změn.

## 7 Seznam použitých zdrojů

Apify. Web scraping & Automation: API scraping. Apify Documentation [online]. [cit. 2023-09-16]. Dostupné z: <https://docs.apify.com/academy/api-scraping>

Banu, Afshan a Pedamkar Priya. 2023. What is Data Processing?. EDUCBA [online]. 13. říjen 2023 [cit. 2023-09-25]. Dostupné z: <https://www.educba.com/what-is-data-processing/>

Carr, Steven. 2022. Find out how remote usability testing can work for you. CARR, Steven. User Testing [online]. 14. únor 2022. [cit. 2023-10-01]. Dostupné z: <https://www.usertesting.com/blog/what-is-remote-usability-testing>.

Casado Eddie. 2020. Usability testing methods: Best 7 ways to test your users. CASADO, Eddie. Mouseflow [online]. 18. listopad 2020. [cit. 2023-10-04]. Dostupné z: <https://mouseflow.com/blog/usability-testing-methods/>

CZ.NIC. 2014. Autorský zákon na Internetu. Jak na Internet [online]. 2014. [cit. 2023-09-18]. Dostupné z: <https://www.jaknainternet.cz/page/1771/autorsky-zakon-na-internetu/>

Čech, Pavel. 2022. Databáze a data. Jejich ochrana a možnosti použití. SEDLAKOVA LEGAL [online]. 2. března 2022. [cit. 2023-09-18]. Dostupné z: <https://www.sedlakovalegal.cz/cs/databaze-a-data-jejich-ochrana-a-moznosti-pouziti>

Dobrý web, s.r.o. Česká pravidla přístupnosti. Přístupnost.cz [online]. [cit. 2023-06-11]. Dostupné z: <http://www.pristupnost.cz/jak-tvorit-pristupny-web/pravidla-pristupnosti/ceska-pravidla-pristupnosti/>

Duggal, Nikita. 2023. What Is Data Processing: Cycle, Types, Methods, Steps and Examples. Simplilerarn [online]. 24. červenec 2023 [cit. 2023-09-25]. Dostupné z: <https://www.simplilearn.com/what-is-data-processing-article>

FAPI Business s.r.o. 2021. Autorské právo na internetu: Kdy váš web porušuje zákon a co když někdo zneužije váš obsah? FAPI [online]. 30. srpen 2021. [cit. 2023-09-18].

Dostupné z: <https://fapi.cz/blog/autorske-pravo-na-internetu-kdy-vas-web-porusuje-zakon-a-co-kdyz-nekdo-zneuzi-vas-obsah/>

Frankenfield, Jake. 2023. Data Analytics: What It Is, How It's Used, and 4 Basic Techniques. Investopedia [online]. 9. srpen. 2023 [cit. 2023-09-27]. Dostupné z: <https://www.investopedia.com/terms/d/data-analytics.asp>

Harkiran78. 2023. What is Web Scraping and How to Use It? Geeks for geeks [online]. 23. březen 2023 [cit. 2023-09-15]. Dostupné z: <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>

Hotjar Ltd. 2022. The different types of usability testing methods for your projects. Hotjar [online]. 2. únor 2022. [cit. 2023-10-01]. Dostupné z: <https://www.hotjar.com/usability-testing/methods/>

Hotjar Ltd. 2023. A complete guide to user testing. Hotjar: Website Heatmaps & Behavior Analytics Tool [online]. 16. červen. 2023. [cit. 2023-10-01]. 2023. Dostupné z: <https://www.hotjar.com/user-testing/>

Chvátalová, Daniela. 2023. Ochrana oznamovatelů v České republice. Právní prostor [online]. 6. červen. 2023 [cit. 2023-07-04]. Dostupné z: <https://www.pravniprostor.cz/clanky/ostatni-pravo/ochrana-oznamovatel-u-v-ceske-republice>

Karatas, Gulbahar. 2023. A Comprehensive Guide to Web Scraping Techniques in 2023. AIMultiple [online]. 8. září 2023 [cit. 2023-09-15]. Dostupné z: <https://research.aimultiple.com/scraping-techniques/>

Kelley, Karin. 2023. What is Data Analysis?: Process, Types, Methods, and Techniques. Simplilerarn [online]. 4. srpen 2023 [cit. 2023-09-27]. Dostupné z: <https://www.simplilearn.com/data-analysis-methods-process-types-article>

Khandelwal, Anuj. 2023. What Is Data Processing? Definition, Stages and Methods.. Indeed [online]. 4. květen 2023. [cit. 2023-09-25]. Dostupné z: <https://www.indeed.com/career-advice/career-development/what-is-data-processing>

Kouba Tomáš. 2021. Jak na GDPR na webu – praktický návod. Net magnet [online]. 21. prosince 2021. [cit. 2023-06-28]. Dostupné z: <https://www.netmagnet.cz/blog/gdpr/>

Ministerstvo vnitra České republiky. 2023. Metodické materiály ODK. [online]. [cit. 2023-06-16]. Dostupné z: <https://www.mvcr.cz/odk2/clanek/metodicke-materialy-k-zakonnym-zmocnenim.aspx?q=Y2hudW09Mw%3D%3D>

Odbor hlavního architekta eGovernmentu. Otevřená data: Popis otevřených dat. Národní architektonický plán: Architektura eGovernmentu ČR [online]. [cit. 2023-06-16]. Dostupné z: [https://archi.gov.cz/nap:otvarena\\_data](https://archi.gov.cz/nap:otvarena_data)

Omniconvert. 2023. What is User testing. [online]. 11. květen 2023 [cit. 2023-10-01] Dostupné z: <https://www.omniconvert.com/what-is/user-testing/>.

Pavlíček, Radek. 2009. Přístupný web a jak se vyvarovat chyb. Web Ministerstva vnitra České republiky [online]. 2009 [cit. 2023-06-02]. Dostupné z: <https://www.mvcr.cz/mvcren/article/pristupny-web-a-jak-se-vyvarovat-chyb.aspx>

Pavlíček, Radek. 2019. Web Content Accessibility Guidelines (WCAG): seznamte se, prosím. Zdroják [online]. 22. února 2019. [cit. 2023-06-11]. Dostupné z: <https://zdrojak.cz/clanky/web-content-accessibility-guidelines-wcag-seznamte-se-prosim/>

Pracovní skupina WP29. 2017. Stručný výklad GDPR – kdy, proč a jak. Asociace za lepší ICT řešení [online]. 26. června 2017. [cit. 2023-06-28]. Dostupné z: <https://lepsi-reseni.cz/ochrana-osobnich-udaju-gdpr/vyklad-gdpr-proc-jak/>

Preuss, Ondřej. 2023. Jaký je význam úřední desky pro občany a jak ji efektivně využít? Dostupný advokát [online]. 10. února 2023. [cit. 2023-06-17]. Dostupné z: <https://dostupnyadvokat.cz/blog/uredni-deska>

Schade, Amy. 2013. Remote Usability Tests: Moderated and Unmoderated. Nielsen Norman Group [online]. 12. říjen 2013. [cit. 2023-10-01]. Dostupné z: <https://www.nngroup.com/articles/remote-usability-tests/>

Schade, Amy. 2013. Task Analysis: Support Users in Achieving Their Goals. ROSALA, Maria. Nielsen Norman Group [online]. 12. říjen 2013 [cit. 2023-10-04]. Dostupné z: <https://www.nngroup.com/articles/remote-usability-tests/>

Škop, Michal. 2022. Nová éra úředních desek. Otevřená data [online]. 11. února 2022. [cit. 2023-06-17]. Dostupné z: <https://data.gov.cz/články/nová-éra-úředních-desek>

Tromm Torsten. User Testing: Definition, How To, Tools and Methods. Userpeek [online] [cit. 2023-10-01] Dostupné z: <https://userpeek.com/blog/user-experience-testing/>

Úřad pro ochranu osobních údajů. Základní příručka k ochraně údajů. [online]. [cit. 2023-06-28]. Dostupné z: <https://old.uoou.cz/obecne-narizeni-gdpr-strucne/ds-4843/archiv=0&p1=3938>

User Interviews Inc. Task Analysis. User interviews [online]. [cit. 2023-10-04]. Dostupné z: <https://www.userinterviews.com/ux-research-field-guide-chapter/task-analysis>

Vranková Michaela. 2023. Přístupné webové stránky: 10. Zákon o přístupnosti. IPK

Metodické příručky pro knihovny [online]. [cit. 2023-06-02]. Dostupné z: <https://pristupne-stranky.cz/zakon-a-standardy/>

Vranková Michaela. 2023. Vzor prohlášení o přístupnosti. Přístupné-stranky.cz [online]. [cit. 2023-06-11]. Dostupné z: <https://pristupne-stranky.cz/vzor-prohlaseni-o-pristupnosti/>

Vyhláška č. 515/2020 Sb.: Vyhláška o struktuře informací zveřejňovaných o povinném subjektu a o osnově popisu úkonů vykonávaných v rámci agendy. 2021. Zákon pro lidi [online]. 1. srpen 2021. [cit. 2023-06-16]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2020-515>

W3C konsorcium. 2018. Web Content Accessibility Guidelines (WCAG) 2.0 [online]. 5. července 2018. [cit. 2023-06-11]. Dostupné z: <https://www.w3.org/TR/WCAG20/>

Zákon 171/2023 Sb.: Zákon o ochraně oznamovatelů. 2023. ASPI, Walters Kluwer ČR [online]. 2. červen 2023. [cit. 2023-07-04]. Dostupné z: [https://www.aspi.cz/products/lawText/1/100662/1/2#c\\_4277](https://www.aspi.cz/products/lawText/1/100662/1/2#c_4277)

Zákon č. 106/1999 Sb.: Zákon o svobodném přístupu k informacím. 2023. Zákony pro lidi [online]. 1. duben 2023. [cit. 2023-06-16]. Dostupné z: <https://www.zakonyprolidi.cz/cs/1999-106>

Zákon č. 121/2000 Sb.: Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). 2023. Zákony pro lidi [online]. 5. ledna 2023. [cit. 2023-09-18]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2000-121>

Zákon č. 500/2004 Sb.: Zákon správní řád. 2021. Zákon pro lidi [online]. 1. leden 2021. [cit. 2023-06-17]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2004-500>

Zákon č. 99/2019 Sb. Zákony pro lidi [online]. [cit. 2023-06-02]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2019-99>

ZenRows, Inc. 2022. What Is a Headless Browser and Best Ones for Web Scraping. ZenRows [online]. 23. listopad 2022 [cit. 2023-09-16]. Dostupné z: <https://www.zenrows.com/blog/headless-browser-scraping#headless-browser-scraping>

## **8 Přílohy**

Příloha 1 - diagram předpokládaného řešení aktuality

Příloha 2 - Diagram předpokládaného řešení přidání nového oznámení

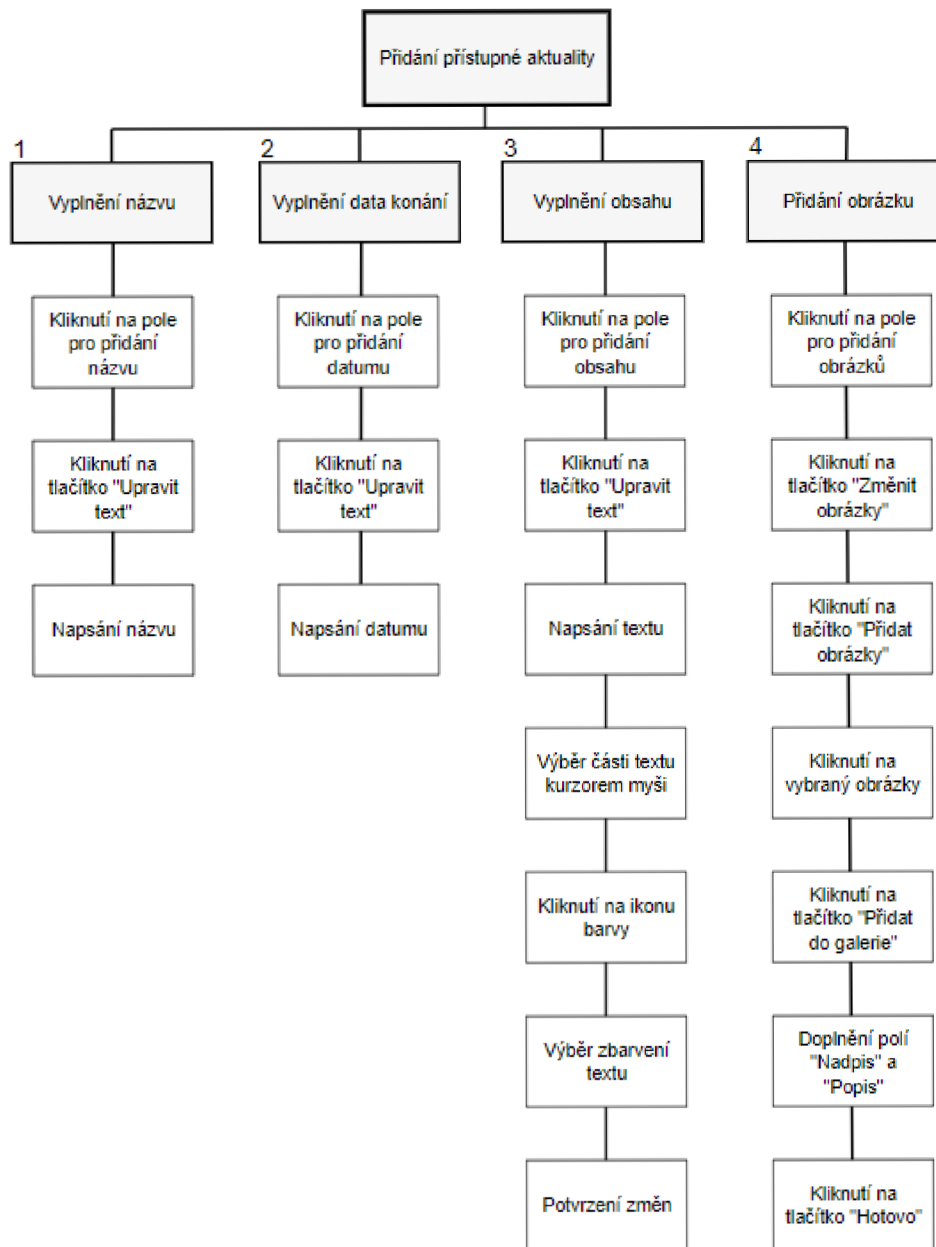
Příloha 3 - Testovací scénář

Příloha 4a – Dotazník

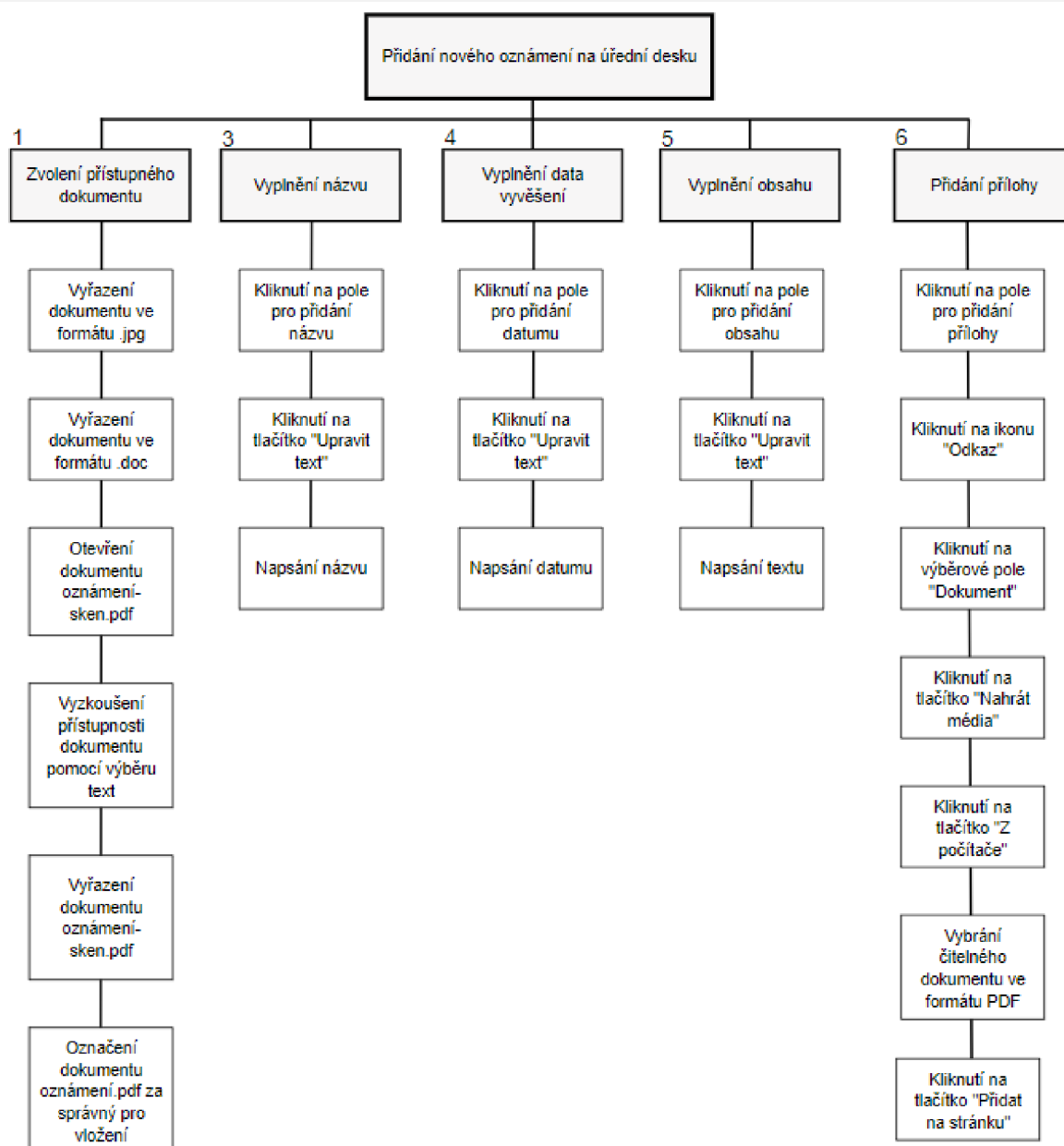
Příloha 4b – Dotazník



## 8.1 Příloha 1 - diagram předpokládaného řešení aktuality



## 8.2 Příloha 2 - Diagram předpokládaného řešení přidání nového oznámení



## **8.3 Příloha 3 - Testovací scénář**

### **Protokol testování - 1.část**

První část je zaměřena na přidávání příspěvků v rámci redakčního systému. V případě, že si nebudete vědět rady s určitou úpravou, zeptej se moderátora.

#### **Přidejte novou aktualitu.**

Volnost řešení je zcela na Vás.

Povinné je přidání názvu aktuality, data konání a obrázku, avšak aktualitu musíte vyplnit tak, aby splnila svůj informativní účel, tj. po jejím zobrazení se návštěvník dozví všechny potřebné informace.

#### **Přidejte nové oznámení na úřední desku.**

Volnost řešení je zcela na Vás.

Povinné je pouze vyplnění názvu, data vyvěšení a přidání dokumentu.

V rámci dokumentu máte na výběr z 4 dokumentů, vyberte pouze 1, který je podle Vás pro vložení nejvhodnější.

### **Protokol testování - 2.část**

Webové stránky: <https://www.kuncicepo.cz/>

V rámci dotazů vždy doplňte odkaz na stránku, kde jste informace našli.

Informace hledejte kdekoliv rámci webových stránek, avšak nevyužívejte vyhledávač. Při nalezení informace s vždy vraťte na úvodní stránky a opět začněte hledání odtamtud hledání.

**Kde naleznete úřední desku?**

**Kde naleznete elektronikou podatelnu?**

**Kde naleznete ID datové schránky?**

**Kde naleznete dokumenty a informace o rozpočtu?**

**Kde naleznete sazebník úhrad v rámci poskytování informací?**

**Kde naleznete vyhlášky a zákony?**

**Kde naleznete výroční zprávu?**

**Kde zjistíte, jakým způsobem je nakládáno s osobními údaji, které můžete poskytnout například při vyplnění formuláře či jinou formou poskytnutí?**

**Kde naleznete informace týkající se přístupnosti webových stránek, respektive o jejich současném stavu vůči přístupnosti?**

## 8.4 Příloha 4a – Dotazník

### Správa zákonných částí webových stránek

Dobrý den,

tento formulář slouží k analýze toho, kdo je hlavním tvůrcem a následným upravovatelem zákonem povinného obsahu, který musí být na webových stránkách zveřejněn, jedná se především o povinná prohlášení.

Odpovědi jsou vždy stejného typu:

- 1) Správce/tvůrce webových stránek = společnost či osoba, která Vám webové stránky vytvořila a poskytuje Vám případný servis a podporu
- 2) Pověřená osoba v rámci obce = myšlena je pověřená osoba z vedení obce, úřadu, atd., která má na starosti obsah webových stránek
- 3) 3. strana = myšlena je osoba či společnost, která Vám na míru vytvořila daný obsah dle zakázky a v případě Vašich požadavků Vám jej i upravuje, ale nejedná se o tvůrce Vašich webových stránek

Poznámka:

Dotazy se vždy týkají přímo obsahu, tj. textů nebo souborů, které jsou umístěny na konkrétní podstránce na Vašich webových stránkách.

Kdo je tvůrcem obsahu "Prohlášení o přístupnosti"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

Kdo provádí změny v obsahu "Prohlášení o přístupnosti"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

## 8.5 Příloha 4b – Dotazník

Kdo je tvůrcem obsahu prohlášení "GDPR" neboli "zásad o ochraně osobních údajů"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

Kdo provádí změny v obsahu prohlášení "GDPR" neboli "zásad o ochraně osobních údajů"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

Kdo je tvůrcem obsahu "Povinně zveřejňovaných údajů"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

Kdo provádí změny v obsahu "Povinně zveřejňovaných údajů"?

- Správce/tvůrce webových stránek
- Pověřená osoba v rámci obce
- 3. strana

Jakým způsobem většinou provádíte rutinní změny obsahu na webových stránkách? (Např. změna běžného textu, přidání dokumentů, přidání aktuality, přidání oznámení v rámci úřední desky atd.)

- Zadáním požadavku na zapracování změn tvůrci webových stránek v rámci jeho podpory
- Vlastním zpracováním
- Zadáním požadavku 3. straně, která se stará o obsah, ale není tvůrcem webu