

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní program: Informační studia a knihovnictví

Studijní obor: INSK

DIPLOMOVÁ PRÁCE

Obsahová metadata vytvářená uživateli a informačními profesionály:
srovnávací analýza

Content metadata created by users and information professionals: a
comparative analysis

Hana Světelská

Považuji za svou milou povinnost poděkovat vedoucí práce paní PhDr. Heleně Kučerové, Ph.D. za odborné vedení a cenné rady při zpracování této práce.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 5. května 2020

Hana Světelská

Abstrakt:

Diplomová práce pojednává o profesionálních a uživatelských metadatech, zejména se zaměřuje na metadata obsahová. Jsou zde popsána metadata obecně, jejich důležité charakteristiky, funkce a typy. Detailněji jsou popsána profesionální obsahová metadata a uživatelská metadata, definovány jsou také další termíny, které jsou v práci využívány. Cílem analytické části bylo porovnání profesionálních a uživatelských obsahových metadat. Probíhalo za pomoci souboru obsahových metadatových výroků popisujících beletristické knihy a sebraných z několika desítek databází. Zároveň bylo cílem analýzy určit, zda existují některá uživatelská metadata, která přinášejí nové informace oproti metadatům profesionálním.

Klíčová slova:

profesionální metadata, uživatelská metadata, obsahová metadata, čtenářské portály, tagy

Abstract:

The master thesis is focused on professional and user-generated metadata with emphasis on content metadata. It describes metadata in general, their important characteristics, functions, and types, and defines other related terms. The professional content metadata and user-generated metadata are described in more detail. The analytical part aims to compare professional and user-generated content metadata. For this purpose, a dataset of metadata statements on fiction books collected from dozens of databases was used. It also aims to determine whether user-generated metadata bring additional information to metadata created by professionals.

Keywords:

professional metadata, user-generated metadata, content metadata, reader's platforms, tags

Obsah

Seznam zkratk.....	8
Předmluva.....	10
1. Úvod	11
2. Teoretický úvod.....	12
2.1 Metadata	12
2.2 Obsahová metadata.....	18
2.3 Profesionální obsahová metadata	19
2.4 Uživatelská metadata	22
3. Přehled dosavadních výzkumů	33
3.1 Obsahová metadata tvořená uživateli obecně.....	35
3.2 Obsahová metadata – řešení v knihovnách a vědeckých institucích	38
3.3 Obsahová metadata – řešení v komerčním prostředí.....	41
3.4 Srovnání více způsobů tvorby obsahových metadat.....	42
4. Cíle analýzy a výzkumné otázky	46
5. Metodika analýzy	48
5.1 Definice důležitých pojmů.....	48
5.1.1 Databáze.....	48
5.1.2 Thema.....	49
5.1.3 SKOS	50
5.1.4 Profesionální a uživatelská metadata	51
5.1.5 Objektivní a subjektivní metadatové výroky	51
5.2 Výběr knih	52
5.3 Výběr databází	55
5.3.1 Knihovní katalogy české.....	56
5.3.2 Katalogy knihkupectví české	57
5.3.3 Knihovní katalogy anglicky mluvících zemí	58

5.3.4 Katalogy knihkupectví anglicky mluvících zemí.....	59
5.4 Výběr analyzovaných metadatových elementů	59
5.5 Sběr dat	60
5.5.1 Výběr záznamu a zobrazení	61
5.5.2 Přehled typů metadat.....	61
5.5.3 Speciální případ: pole Subject.....	71
5.5.4 Další pravidla zápisu dat	72
5.6 Analýza dat	73
5.6.1 Objektivita/subjektivita uživatelských metadatových výroků	73
5.6.2 Shoda uživatelských metadat s profesionálními	74
5.6.3 Počet metadatových výroků pro jednotlivé knihy.....	76
5.6.4 Mapování na Thema.....	76
6. Vyhodnocení analýzy	79
6.1 Objektivní uživatelská metadata.....	79
6.1.1 Objektivní uživatelské metadatové výroky, které odpovídají profesionálním	80
6.1.2 Objektivní uživatelské metadatové výroky, které neodpovídají profesionálním	81
6.2 Porovnání profesionálních a uživatelských metadat podle typu metadatových výroků.....	83
6.3 Metadata a Thema.....	86
6.3.1 Hodnoty profesionálních metadatových výroků	86
6.3.2 Hodnoty uživatelských metadatových výroků.....	89
6.3.3 Porovnání profesionálních a uživatelských metadatových výroků.....	93
7. Odpovědi na výzkumné otázky	95
7.1 Které typy obsahových metadatových výroků existují? Liší se to mezi profesionálními a uživatelskými metadaty?	95
7.2 Kolik metadatových výroků kterého typu databáze obsahují?	99

7.3 Z jaké části jsou uživatelská metadata objektivní? Liší se to mezi jejich formami?.....	101
7.4 Z jaké části se profesionální a uživatelská metadata shodují?.....	101
7.5 Existují uživatelská metadata, která jsou potenciálně užitečná a mohla by doplnit profesionální metadata?.....	102
7.6 Kdo při popisu informačních zdrojů tvoří složitější koncepty, profesionálové nebo uživatelé?	103
7.7 Jaká je míra sémantické interoperability uživatelských a profesionálních metadat?.....	104
8. Diskuse	106
8.1 Porovnání s dosavadními výzkumy	106
8.2 Poznatky mimo výzkumné otázky.....	107
8.3 Zpětné hodnocení výzkumu.....	108
8.4 Možnosti dalšího zkoumání.....	110
9. Závěr.....	111
Použitá literatura.....	112
Seznamy tabulek a grafů	129
Seznam tabulek.....	129
Seznam grafů	130
Seznam příloh.....	131

Seznam zkratk

AACR2 – Anglo-American Cataloguing Rules, second edition

BIC – Book Industry Communication

CEJSH – Central European Journal of Social Sciences and Humanities

DDC – Dewey Decimal Classification

DDT – Deweyho desetinné třídění

dnf - did not finish

DOAB – Directory of Open Access Books

DOAJ – Directory of Open Access Journals

EAD – Encoded Archival Description

EBSCO – Elton Bryson Stephens Company

ERIC – Education Resources Information Center

FAST – Faceted Application of Subject Terminology

FRBR – Functional Requirements for Bibliographic Records

IBZ – Internationale Bibliographie der Zeitschriftenliteratur

IFLA – International Federation of Library Associations and Institutions

IMLS – Institute of Museum and Library Services

ISO – International Organization for Standardization

JSTOR – Journal Storage

KBART – Knowledge Bases And Related Tools

KIV – Knihovnictví a informační věda

LCSH – Library of Congress Subject Headings

LISS – Library & Information Science Source

LISTA – Library, Information Science and Technology Abstracts

MARC – MACHine-Readable Cataloging

MDS – Melvil Decimal System

MDT – Mezinárodní desetinné třídění

MODS – Metadata Object Description Schema

NISO – National Information Standards Organization

OAPEN – Open Access Publishing in European Networks

OPAC – Online Public Access Catalogue

PAC – Publicly Available Content Database

RDA – Resource Description and Access

SKOS – Simple Knowledge Organization System

TEI – Text Encoding Initiative

UNESCO – United Nations Educational, Scientific and Cultural Organization

URL – Uniform Resource Locator

VRA – Visual Resources Association

Předmluva

Dlouhodobě mě zajímá téma obsahových metadat u knih a také uživatelského tagování. Proto jsem si tuto problematiku vybrala jako téma své diplomové práce s tím, že mě bude zajímat právě porovnání profesionálních a uživatelských metadat.

Zároveň mě na tomto tématu láká i to, že sama jsem uživatelem českého čtenářského portálu Databáze knih a pokud vytvářím štítky ke knihám, tak je to vždy s ohledem na jejich objektivitu a podle mého nejlepšího vědomí a svědomí. Zajímá mě, jestli to tak mají i ostatní uživatelé.

Průběh výzkumu vnímám pozitivně, narazila jsem na několik drobných problémů, to se ale u práce tohoto rozsahu dalo předpokládat. Poněkud jsem se potýkala s terminologickými otázkami, jelikož problematika uživatelských metadat je poměrně nová, a ne všechny pojmy mají svou ustálenou verzi v češtině.

Tato práce od předmluvy po závěr čítá 172 105 znaků.

1. Úvod

Se zvyšujícím se vlivem informačních technologií na životní styl naší společnosti se zvyšují možnosti publikování obsahu na internetu pro všechny lidi bez rozdílu. Díky tomu na počátku 21. století vznikly první platformy, které umožňují uživatelské tagování informačních zdrojů, ať už jde o fotografie, hudbu nebo texty. Tvorba metadat, která byla do té doby vyhrazena jen těm nejpopulárnějším – knihovníkům a dalším informačním pracovníkům – je v tu chvíli umožněna každému.

Cílem této diplomové práce je porovnat profesionální a uživatelská obsahová metadata za využití souboru beletristických knih a metadatových výroků sebraných z několika desítek databází. Zároveň bylo cílem analýzy určit, zda existují některá uživatelská metadata, která přinášejí nové informace oproti metadatům profesionálním.

Druhá kapitola představuje teoretický úvod do problematiky metadat a dělí se na čtyři podkapitoly. První se zabývá metadaty obecně, jejich důležitými charakteristikami, funkcemi a typy. Druhá definuje obsahová metadata a detailněji jsou ve třetí podkapitole popsána profesionální obsahová metadata. Čtvrtá podkapitola se pak zabývá uživatelskými metadaty. Kapitola třetí přináší přehled dosavadních výzkumů v této oblasti.

Čtvrtá kapitola stanovuje cíle analytické části a výzkumné otázky, v páté kapitole je zdokumentován přesný postup výzkumu, díky němuž by měl být proveditelný opakovaně. Šestá kapitola pak přináší vyhodnocení jednotlivých zkoumaných parametrů, které by mělo přinést celkový obraz problematiky, sedmá pak konkrétně zodpovídá výzkumné otázky a odpovědi opírá o konkrétní čísla, která byla v rámci výzkumu zjištěna. Osmou kapitolou je diskuse, kde je zejména uvedeno zpětné hodnocení výzkumu a náměty na další zkoumání. Poslední, devátá, kapitola shrnuje zjištěné poznatky a podává stručné zamyšlení nad problematikou uživatelských metadat do budoucnosti.

2. Teoretický úvod

2.1 Metadata

Ačkoli je pojem metadata starý jen pár desetiletí (v angličtině se tento termín poprvé objevil v roce 1968), myšlenka metadat je mnohem starší. Objevila se v souvislosti s prvními knihovny: za první knihovní katalog je považován Pinakes, který byl vytvořen 245 př. n. l. Kallimachem pro Alexandrijskou knihovnu. Do dnešní doby se dochovaly pouze zlomky, ale i tak je patrné, že dílo obsahovalo žánr, název, jméno autora spolu s biografickými informacemi, shrnutí díla a celkový počet řádků díla. (Pomerantz, 2015, s. 5-6) Pokud považujeme knihy a svitky za informační objekty, můžeme takovýto katalog považovat za sbírku metadat (Haynes, 2018, s. 4).

Podle etymologie slova jsou metadata „data o datech“, nicméně tato definice nezahrnuje všechny aspekty metadat, které jsou považovány za důležité, a zároveň je příliš široká. O přesné definici metadat nepanuje mezi odborníky shoda, je zde několik oblastí, kde se jejich názory rozcházejí. Podle některých se pojem metadata vztahuje pouze na digitální informace, podle jiných i na „papírové“, jiné hledisko je např. zda musí být metadata vytvořena na základě nějakého schématu nebo standardizované struktury či nikoli. (Sicilia, 2014, s. 2)

Prefix „meta-“ pochází z řečtiny, kde znamená „po“ (Μετά, 2020). V definici „data o datech“ nejde tedy o přesný překlad, nicméně je konzistentní s běžným užíváním slova „meta“ v dnešním jazyce – indikuje vyšší míru abstrakce (Pomerantz, 2015, s.6).

Metadata jsou zjednodušenou reprezentací jiného objektu, takovým jiným objektem může být v zásadě cokoli, avšak nejčastěji se používá označení informační zdroj. Metadata popisují různé atributy takového zdroje a jejich cílem je učinit ho užitečným. (Pomerantz, 2015, s.11, s. 26; Sicilia, 2014, s. 4-5, s. 13; Chowdhury, 2007, s. 140)

Za klíčové vlastnosti metadat jsou různými autory považovány:

- zjednodušení oproti realitě (Pomerantz, 2015, s.11),

- referenční povaha (říkají něco o něčem jiném) (Sicilia, 2014, s. 5),
- strojová čitelnost (Sicilia, 2014, s. 10),
- objektivita (Pomerantz, 2015, s. 16),
- strukturovanost (Sicilia, 2014, s. 4),
- užitečnost pro uživatele (Dempsey, 1998, citováno podle Chowdhury, 2007, s. 140),
- interoperabilita (Sicilia, 2014, s. v),
- elektronická forma (Chan, 2016, s. 113-114).

Metadata jsou využívána různými uživateli. Jednak lidmi i počítačovými programy, a jednak poskytovateli informačních služeb i jejich uživateli. (Chowdhury, 2007, s. 141)

Role metadat v dnešní společnosti vyvolává otázky ohledně jejich etického použití, které se týkají zejména soukromí, bezpečnosti, vlastnictví a kontroly. Důvodem pro obavy je to, že potenciálně umožňují přístup k informacím široké veřejnosti. Osobní data, zejména ta, která prezentují názory a postoje, mohou být snadno zneužita k nevyžádané reklamě, obtěžování nebo krádeži identity, pokud nejsou přenášena bezpečně. To se týká také některých typů metadat. Za další etický aspekt metadat může být považována informační nerovnost (také nazývaná digitální propast). Existují skupiny, které jsou vyloučené z digitální společnosti z hlediska majetku, věku nebo vzdělání. Metadata mají možnost tuto propast překonat, např. u lidí znevýhodněných tělesným postižením. (Haynes, 2018, s. 221-223)

Jelikož terminologie v této oblasti je nejednotná, musíme zde přesně stanovit, jak budeme označovat entity týkající se metadat v rámci této práce.

Pod termínem metadatový výrok rozumíme jednu trojici složenou z následujících částí: (1) informační zdroj, (2) metadatový element a (3) hodnota. Metadatový element je atribut nebo vlastnost informačních zdrojů, např. autor. V rámci jednoho metadatového výroku má přiřazenou konkrétní hodnotu, což je řetězec znaků přiřazený metadatovému elementu a reprezentující daný informační zdroj. (Pomerantz, 2015, s. 27; Vocabulary Encoding Schemes, 2020) Na příkladu můžeme metadatový výrok znázornit takto:

Hamlet

Autor: Shakespeare, William

Pro dosažení systematickosti metadat jsou používána metadatová schémata. Metadatové schéma je soustava pravidel, která určuje, jaké trojice se tvoří a jakým způsobem. Metadatový záznam je souhrn všech metadatových výroků o jednom informačním zdroji. Také platí pravidlo, které říká, že by měl existovat pouze jeden záznam daného subjektu podle jednoho metadatového schématu. (Pomerantz, 2015, s. 27-29, s. 54-55)

Metadatová schémata mohou být dvou druhů: sada metadatových elementů (anglicky metadata element set) definuje metadatové elementy a slovník hodnot (anglicky value vocabulary) definuje hodnoty. Případně se pro něj může používat i termín schéma pro kódování slovní zásoby (anglicky vocabulary encoding scheme), který klade důraz na strukturovanost slovníku a řízené hodnoty. (Vocabulary Encoding Schemes, 2020) Nejznámějším příkladem sady metadatových elementů je Dublin Core, obsahuje např. elementy tvůrce, datum, identifikátor nebo jazyk (Dublin Core™ Metadata Element Set, Version 1.1, 2012).

Je velmi obtížné metadata hodnotit. Jeden způsob je založen na hodnocení konzistence metadatových výroků použitých pro určité informační zdroje. Avšak v případě, kde je cílem pluralita pohledů, by takovéto měření vhodné nebylo. (Mai, 2011, s. 116)

Podle Sicilii (2014, s. 5-6) se u metadat hodnotí minimálně tyto tři dimenze:

1. Kvalita: Většinou je hodnocení zaměřeno na úplnost metadatových záznamů nebo na použití řízených slovníků.
2. Bohatost: Bohatost chápeme jako „*množství užitečných informací nebo možností k propojení metadatových kolekcí nebo systémů*“ (Sicilia, 2014, s. 5) a lze ji chápat na dvou úrovních: na úrovni metadatových schémat a na úrovni metadatových záznamů. Zatím neexistuje žádná metrika pro hodnocení metadatových schémat, která by pracovala s možnostmi zachycení podrobnějších informací. Na úrovni metadatových záznamů je hodnocení bohatosti ještě obtížnější, protože závisí na použitém schématu,

na úplnosti záznamů a na dalších parametrech, které se mohou lišit v závislosti na oboru.

3. Interoperabilita: První úroveň interoperability je syntax. Při běžném použití metadat může dojít k tomu, že je preferována jednoduchost na úkor bohatosti, což se projeví např. při výběru metadatového schématu. Ten má na interoperabilitu samozřejmě velký vliv. Druhá – a důležitější – úroveň interoperability je sémantika. Pro ni je základem použití některého systému organizace znalostí, tyto systémy jsou však vyvíjeny a uchovávány mimo metadatová schémata. A protože se v čase mění, je důležité, jak je systém připraven na přejímání změn a aktualizaci použitých termínů.

U hlediska kvality metadat se ještě na chvíli zastavíme. Sicilia ve své monografii zmiňuje důležitost hodnocení metadat v rámci systémů, přestože přiznává, že žádné zavedené způsoby hodnocení zatím neexistují (Sicilia, 2014, s. 84). Při určování kvality metadat vychází ze studie Funkční požadavky na bibliografické záznamy a jejich čtyř uživatelských úloh (najít, určit, vybrat, získat). Přidává k nim dvě další úlohy: klastrování metadatových elementů a vylepšování efektivity systému. Kvalita metadat pak přímo souvisí s tím, jak zajišťují těchto šest funkcí. Přesto při hodnocení mohou nastat různé typy problémů (Sicilia, 2014, s. 64-66):

- Existuje více vlastností metadat, které ovlivňují jejich kvalitu, ale zároveň jsou na sobě nezávislé. Např. úplnost metadatového záznamu i počet gramatických chyb ovlivňují kvalitu záznamu, avšak z různých úhlů pohledu.
- Kvalita metadat závisí na uživatelské komunitě a na používaném systému. Jakékoli měření odpovídající jednomu prostředí nemusí být v jiném případě vhodné.
- Kvalita se v čase mění: záznam může časem přestat odpovídat realitě (např. při popisu webových stránek), do kolekce mohou být přidány další záznamy nebo může dojít ke změně využívání metadat. To vše ovlivňuje kvalitu metadat.

Protože metadata popisují měnící se skutečnost, jakákoli jejich klasifikace bude neúplná a rychle zastarávající. Stejně tak metadatová schémata existují pouze tak

dlouho, dokud je využívá dostatečná uživatelská komunita. (Sicilia, 2014, s. 9-10)
Na rozdíl od většiny ostatních autorů, Sicilia člení metadata více různými způsoby, aby co nejlépe postihl jejich vícekriteriální podstatu (Sicilia, 2014, s. 12-15)¹:

1. Členění podle způsobu vytvoření:
 - a. metadata vytvořená ručně (nebo metadata vytvořená lidmi),
 - b. metadata vytvořená automatizovaným způsobem (nebo metadata vytvořená stroji).
 - c. metadata vytvořená profesionály,
 - d. metadata vytvořená laiky.
2. Členění podle času vytvoření:
 - a. metadata vytvořená ve stejnou dobu jako informační zdroj,
 - b. metadata vytvořená později než vznik či digitalizace informačního zdroje.
3. Členění podle způsobu uložení:
 - a. interní metadata uložená přímo v informačním zdroji, který popisují,
 - b. externí metadata uložená nezávisle na informačním zdroji, který popisují.
4. Členění podle úrovně struktury (toto členění můžeme aplikovat jak na úrovni metadatového schématu, tak na úrovni metadatového záznamu):
 - a. nestrukturovaná metadata,
 - b. částečně strukturovaná metadata,
 - c. strukturovaná metadata.
5. Členění podle účelu (na úrovni metadatového schématu):
 - a. metadata pro obecné účely (např. Dublin Core),
 - b. metadata pro specifické účely, např. metadata pro geoprostorové informace.

¹ Přestože zde používáme termín metadata, v jednom metadatovém záznamu mohou být zastoupeny metadatové výroky více typů. Pokud bychom mluvili konkrétně, měli bychom použít termín metadatový výrok.

Členění podle účelu (na úrovni metadatového elementu):

- a. metadata nezávislá na oboru, která popisují obecné vlastnosti informačních zdrojů,
 - b. metadata závislá na oboru, která popisují informace a znalosti z oblasti, do které daný informační zdroj patří.
6. Členění podle úrovně standardizace (na úrovni metadatového schématu):
- a. metadatové standardy de iure, které prošly procesem standardizace a jsou většinou uznávány mezinárodně (např. standardy ISO),
 - b. metadatové standardy de facto, které mají dominantní pozici v rámci své oborové nebo geografické oblasti, ale nejsou standardy de iure.
7. Členění podle využití (na úrovni metadatového schématu, záznamu nebo elementu): Toto členění je nejdůležitější, protože cílem metadat je učinit data užitečnými, tedy aplikovatelnými. Jde také o členění mezi autory nejčastější, přestože se ne vždy shodují na konkrétních typech. Proto mu věnujeme nejvíce prostoru a pokusíme se shrnout jejich poznatky do jedné souvislé klasifikace (Sicilia, 2014, s. 12-15, s. 27-28; Ahmed, 2001, citováno podle Sicilia, 2014, s. 12-15; Chan, 2016, s. 114; Gilliland-Swetland, 1998, citováno podle Chowdhury, 2007, s. 143-144; Pomerantz, 2015, s. 17-18; Gilliland, 2008, citováno podle Sicilia, 2014, s. 28-29):
- a. Popisná metadata jsou typem metadat, který je založen čistě na informačním zdroji, popisuje ho. Zajišťují identifikaci a vyhledatelnost daného informačního zdroje, také platí, že čím lepší popisná metadata, tím lepší vyhledatelnost informačního zdroje (v kontextu dané uživatelské komunity a systému). Příklady popisných metadat jsou např. autor, název, rok vydání, klíčová slova a předmětová klasifikace.
 - b. Metadata pro předmětové rejstříky (nebo obecněji systémy organizace znalostí) jsou zvláštní podskupinou metadat popisných. Informačním zdrojem, který popisují, jsou předměty nebo předmětové kategorie. Kladou důraz na vztahy mezi jednotlivými předměty.

- c. Administrativní metadata zajišťují správu a uložení informačních zdrojů. Patří tam tyto údaje: datum a způsob vzniku, datum a způsob akvizice, typ souboru, kdo k němu má přístup a další. Mohou také zahrnovat „meta-metadata“, čímž rozumíme informace o metadatach (např. kdo je vytvořil).
- d. Strukturální metadata se někdy považují za podskupinu metadat administrativních, jelikož jsou nezbytná k zajištění správy informačního zdroje. Poskytují informace o interní struktuře informačního zdroje a případně o jeho vztazích se souvisejícími objekty. Zajišťují správné uložení, navigaci a prezentaci zdrojů, příkladem mohou být tyto údaje: číslování stránek a kapitol, obsahy, rejstříky a správné pořadí všech částí za sebou.
- e. Archivační metadata jsou také někdy řazena pod administrativní metadata. Slouží k zajištění procesu archivace informačních zdrojů nebo celých kolekcí. Mezi používané údaje patří např. fyzický stav zdroje, akce a předpisy zajišťující archivaci zdroje, okolnosti digitalizace či zda je nutné pro přístup k souboru emulovat určité prostředí.
- f. Technická metadata popisují fungování systému a chování metadat. Řadí se mezi ně informace o hardwaru a softwaru, formátech, kompresních poměrech apod. a také autentizační a bezpečnostní data.
- g. Metadata o použití specifikují úroveň a typ užívání informačních zdrojů nebo jejich kolekcí. Mezi nejčastější metadata o použití patří počet stažení, data, kdy k nim došlo, a kteří uživatelé si zdroj stáhli, informace o verzích nebo logy vyhledávání ve zdroji.

Sicilia typy z bodu 7 řadí do dvou základních kategorií: metadata pro vyhledání zdroje a metadata o využití zdroje (Sicilia, 2014, s. 14).

2.2 Obsahová metadata

Pokud chceme mluvit o obsahu informačních zdrojů, nemůžeme opomenout pojem aboutness. Jedná se o abstraktní pojem používaný v knihovnictví a informační vědě,

filozofii jazyka a filozofii mysli, z knihovnického hlediska pak jde o základní pojem předmětové katalogizace. (Hjørland, 2017, s. 61; Broughton, 2015, s. 57)

V Oxfordském slovníku najdeme takovouto definici aboutness: „*vlastnost nebo fakt vztahu k něčemu nebo bytí o něčem*“ (Aboutness, 2009), nicméně odpovídající český termín zatím nebyl stanoven, proto používáme anglický originál. Aboutness můžeme definovat jako hlavní téma textu (Mertesová používá termín „metatéma“) a zároveň další dílčí témata a podtémata informačního zdroje, která jsou zachycena a reprezentována obsahovou analýzou při tvorbě předmětového indexu. Jelikož je aboutness výtvořem individuálního lidského myšlení, můžeme konstatovat, že je vždy do určité míry subjektivní. Zároveň je posuzována vzhledem k relevanci a informačním potřebám popisovatele a/nebo jiného subjektu. (Mertes, 2017, s. 77-78; Kučerová, 2014, s. 38, s. 46)

Obsahová metadata definujeme jako popisná metadata, která reprezentují aboutness informačního zdroje.

Hodnocení kvality obsahových metadat je – podobně jako u metadat obecně – nelehká záležitost. Podle Hjørlanda může být kvalita určena pouze z teoretického hlediska, když se vezme v úvahu, jak má správná předmětová reprezentace vypadat. Nicméně zde záleží na funkci metadat. Pokud je účelem odpovídat na uživatelské dotazy, obsahová metadata by měla být konzistentní ve vztahu k určitým předpokládaným dotazům nebo ke stanoveným vnitřním pravidlům. Objektivní obsahová metadata nemusí být vždy ta nejlepší. (Hjørland, 2017, s. 59)

2.3 Profesionální obsahová metadata

Z hlediska profesionálního obsahového popisu existují dva typy předmětových přístupových bodů a je třeba zkoumat oba (Hjørland, 2017, s. 55; Chan, 2016, s. 467-468):

1. Prvním typem jsou termíny, které nejsou přímo profesionálními obsahovými metadatovými výroky, ale přesto mohou potenciálně obsahovat informace o obsahu informačního zdroje. Jedná se o slova z názvů, abstrakty, poznámky, plné texty nebo bibliografické reference, případně sem můžeme zařadit i uživatelská metadata. Způsob

vyhledávání, který se zaměřuje na tyto oblasti, nazýváme hledání pomocí klíčových slov. Výhodou je aktuálnost terminologie (vysoká přesnost), nevýhodou naopak nižší úplnost vyhledávání. Pokud by uživatel chtěl najít všechny informační zdroje na dané téma, musel by při vyhledávání použít různá synonyma stejného termínu.

2. Druhým typem jsou termíny, které byly profesionálem určeny jako obsahové metadatové výroky – jde o řízenou slovní zásobu. V tomto případě je vybrán jeden preferovaný termín pro jedno téma, uživatel tedy vyhledá všechny informační zdroje na dané téma pouze při zadání jednoho termínu. Naopak ale vývoj řízených slovníků může být pomalejší, což by neumožňovalo dostatečnou aktualizaci oborové terminologie, čímž může být snížena přesnost vyhledávání.

Nyní se zaměříme na druhý případ, a to na profesionální předmětový popis informačních zdrojů. Nejprve stručně shrneme obecná doporučení pro tvůrce obsahových metadat (zpravidla se jedná o katalogizátory) (Broughton, 2015, s. 58-64):

- Zdroje informací, ze kterých je čerpáno:
 - název informačního zdroje,
 - podnázev,
 - přebal a/nebo obálka,
 - obsah,
 - reference,
 - jiné katalogizační záznamy.
- Nejprve je třeba rozhodnout o obecné předmětové oblasti, do které informační zdroj spadá, zpravidla první úroveň klasifikace.
- V další fázi dochází k identifikaci důležitých pojmů v informačním zdroji obsažených. Zaměření je na ty pojmy, které nejlépe vystihují obsah informačního zdroje, a ty, které bude uživatel pravděpodobně při hledání používat a které mu mohou pomoci.
- Pojmy, které se v aboutness často objevují a které je třeba správně identifikovat (některé z nich jsou vyjádřeny vlastní notací v klasifikačních systémech):
 - místo a časové období,

- forma, ve které jsou informace ve zdroji prezentovány (např. bibliografie, učebnice apod.),
- jazyk,
- osoby a jejich pohlaví, věk, národnost a další charakteristiky,
- uživatelské určení.

Důležitou součástí profesionálního předmětového popisu je řízená slovní zásoba. V lingvistice tento pojem označuje „*organizaci slov do skupin a úrovní, obzvláště jako výstup frekvenční analýzy ve formě seznamu slov, který je určen jako pomoc při psaní, čtení a učení jazyků*“ (Vocabulary Control, 2018). V oblasti knihovnictví je termínem řízená slovní zásoba označována podmnožina přirozeného jazyka, jehož slovní zásoba, syntax, sémantika i pragmatika jsou omezeny (Wellisch, 1996, citováno podle McClung, 2009). Jejím cílem je udržet konzistenci ve tvorbě (obsahových) metadat a navádět uživatele k požadované informaci (Hedden, 2008, s. 33, citováno podle McClung, 2009). Řízená slovní zásoba je prostředkem pro vylepšování vyhledatelnosti informačních zdrojů, při její konzistentní aplikaci uživatel zdroje nalezne při hledání nebo prohlížení podle předmětu efektivněji. (Hanrath, 2017, s. 360)

Nástroj řízené slovní zásoby – řízený slovník nebo také slovník hodnot – je organizovaný seznam termínů a frází, které jsou používány při tvorbě obsahových metadatových výroků (předmětové indexaci informačního zdroje) a v některých případech při vyhledávání informačních zdrojů v kolekci. Obsahují abecedně seřazené termíny s odkazy a poznámkovým aparátem, pro každý termín pak existuje autoritní záznam. Propojením s centrálními autoritními záznamy dostanou lokální bibliografické záznamy přidanou hodnotu: autoritní záznamy poskytují mnoho užitečných informací, které v bibliografických záznamech nemohou být, a díky jejich propojení je možné automaticky generovat navigaci pro uživatele. (O’Neill, 2014, s. 9; Chowdhury, 2007, s. 111-113; Chan, 2016, s. 467-468)

Chowdhury a Chowdhuryová definují dva základní druhy řízených slovníků (Chowdhury, 2007, s. 111-113):

1. Předmětový heslář je prvním typem řízeného slovníku, má tedy všechny jeho vlastnosti. Pod každým heslem najdeme seznam sémanticky souvisejících hesel. Předmětové hesláře byly navrženy, aby doplňovaly

bibliografické klasifikační systémy, jako např. Mezinárodní desetinné třídění. Využívají princip prekoordinace – termíny mezi sebou mají definovaný syntaktický vztah. Nejznámějším příkladem předmětového hesláře je Library of Congress Subject Headings, který spravuje Kongresová knihovna.

2. Tezaurus je druhým hlavním typem řízeného slovníku, který byl vyvinut pro řízení oborových terminologií. Jeho hlavním přínosem je řešení problémů přirozeného jazyka (zejména synonymie, homonymie a různých variant zápisu) a mapování kontextu termínů. Definuje několik typů vztahů mezi jednotlivými termíny: viz (odkazuje na deskriptor), viz též (odkazuje na nedeskriptor), užší a širší termín, související termín. Příkladem tezauru je UNESCO thesaurus.

Broughtonová definuje dva rozměry předmětového popisu (Broughton, 2015, s. 77-79):

1. Přesnost předmětové indexace se zabývá tím, jak přesně tvůrce metadat zařadil předmět informačního zdroje do hierarchie (např. do klasifikačního schématu).
2. Při úplnosti je hodnoceno, zda byly pokryty všechny části a aspekty aboutness informačního zdroje.

Budoucnost využití principů řízené slovní zásoby jako přístupu k tvorbě obsahových metadat není zatím jasná. Podle O'Neilla se budou muset stát dvě věci, aby si řízené slovníky mohly zachovat svou klíčovou roli: náklady na jejich tvorbu, správu a využívání se budou muset znatelně snížit a budou muset být vytvořeny nové přístupy k vyhledávání, aby byly informace z autoritních záznamů lépe zpracovány, aniž by došlo ke zvyšování komplexity systému. (O'Neill, 2014, s. 8)

2.4 Uživatelská metadata

Profesionální tvorba metadat (a organizace znalostí obecně) byla až donedávna jediným způsobem, jak vytvořit z chaosu systém. Zabývalo se tím množství různých profesí, tito lidé byli experty na třídění informací a znalostí, na vytváření systémů organizace znalostí a také na jejich využívání. Obecně se považovalo za platné, že jsou v nejlepší pozici k jakékoli organizaci znalostí nebo informačních

zdrojů. V nedávné minulosti se však objevil druhý způsob: organizovat můžou všichni lidé i bez profesionálních znalostí. (Mai, 2011, s. 114-115, s. 117)

Vznik uživatelských metadat a folksonomií byl zčásti podmíněn myšlenkou, že mnoho existujících systémů organizace znalostí reprezentuje neaktuální pohled na svět. A částečně ho také podpořila idea, že neexistuje jedna celistvá aktivita, která by sjednocovala celé univerzum znalostí, a že v nejlepší pozici pro její vytvoření jsou přímo lidé, kteří se konkrétními znalostmi zabývají. Vzniklo tak hnutí, které nemělo žádný teoretický rámec ani strategii a jeho jediným cílem bylo vytvořit něco, co bude funkční pro jeho členy. Prakticky se toto hnutí zrodilo, když lidé začali používat funkce softwarů, které jim umožňovaly klasifikovat svůj obsah. K prvním platformám, které takové funkce představily, patří Delicious, Flickr, 43things, Furl a Technorati. Nyní lidé mohou označovat – tagovat – dokumenty, události, místa, data, informace i jiné objekty a myšlenky, se kterými se setkají. (Mai, 2011, s. 115-116)

Pokud chceme definovat tagování a folksonomie, musíme se nejprve zaměřit na obsah tvořený uživateli (z anglického user-generated content). Tím rozumíme jakýkoli obsah vložený uživatelem sociálních médií na online platformu daného média. K vložení uživatel nepotřebuje znalosti programování. Na rozdíl od vyhledávačů, pro které je nejmenší jednotkou webová stránka (URL), toto dělení je pro oblast obsahu tvořeného uživatelem nedostatečné. Za jeho nejmenší jednotku považujeme jeden jednotlivý příspěvek jednoho uživatele v jednom čase. (Wyrwoll, 2014, s. 15-16)

Pojem sociální tagování (nebo tagování) se používá od roku 2003, kdy začala fungovat služba Delicious. Rozumíme pod ním činnost, kdy uživatelé na internetu v sociálním sdíleném prostředí vytvářejí klíčová slova – nazýváme je tagy –, kterými popisují, kategorizují nebo komentují digitální informační zdroje. Uživatelé při tagování nejsou omezováni, co se týče množství a typů tagů a neřídí se žádnými pevnými pravidly. Cílem tagování je zachytit individualitu jednotlivých reakcí a/nebo rozdílné pochopení informačního zdroje; uživatelé poskytují zdrojům význam a prostředky pro propojování zdrojů tak, jak je sami chápou. Většinou se jedná o formu trojic uživatel-informační zdroj-zvolený tag. Informační zdroje musí být odlišitelné od ostatních informačních zdrojů např. pomocí URL nebo jiného

identifikátoru. (Rafferty, 2018, s. 501-503; Peters, 2009, s. 153-157; Vander Wal, 2007, citováno podle Peters, 2009, s. 154)

Furner (2010, citováno podle Rafferty, 2018, s. 501-503) dále zdůrazňuje některé vlastnosti tagování:

- manuální proces,
- přirozený jazyk (jako protiklad řízené slovní zásoby),
- demokratický proces,
- provádějí ho tvůrci nebo uživatelé obsahu,
- nízká odbornost v indexaci,
- vysoká odbornost v daném oboru,
- různorodá motivace,
- aplikovatelnost i na informační zdroje, které neobsahují text.

Folksonomii definujeme jako souhrn tagů na určité sociální platformě, výsledek tagování. Je to určitý systém organizace informačních zdrojů, který obsahuje pouze jednu úroveň (není hierarchický) a není řízený. To se může zdát v rozporu s původem slova folksonomie: „folk“ je výraz pro obyčejné lidi a „taxonomie“ označuje řízený systém organizace znalostí, který obvykle mívá více úrovní. Označení folksonomie tedy původně znamená „taxonomie vytvořená obyčejnými lidmi“. (Rafferty, 2018, s. 501; Vander Wal, 2005, citováno podle Rafferty, 2018, s. 501; Benz, 2007, citováno podle Rafferty, 2018, s. 501; Peters, 2009, s. 153-154)

V počátcích folksonomií si jejich příznivci zakládali na myšlence, že inteligence skupiny lidí bude přesnější než inteligence jednotlivce, i když jde o experta. Nazývali je moudrost davu. Nicméně o terminologii v oblasti tvorby uživatelských metadat se stále vedou odborné diskuse, někteří autoři navrhnou místo pojmů tagování a folksonomie používat jiné termíny. Také existuje větší množství pojmů, které jsou tematicky příbuzné, zde si pro zajímavost uvedeme některé z nich (Rafferty, 2018, s. 501, s. 504; Peters, 2009, s. 153-4):

- kolaborativní tagování,
- distribuovaná klasifikace,
- demokratická indexace,
- sociální klasifikace,

- metadata tvořená uživateli;
- tagosféra.

Nejobecněji přijímanými termíny jsou ale v dnešní době tagy, tagování a folksonomie, které budeme pro svoji práci využívat i my. V této práci budeme také pracovat s jedním dalším pojmem, který se vztahuje k oblasti uživatelských metadat, a to je polička. Její následující definice je vytvořena autorkou pro účely této práce. Termín má napodobovat řazení knih v domácí knihovně na jednotlivé poličky, používáme český překlad anglického slova „shelf“, nicméně terminologie v této oblasti není odborníky zavedena. Pokud tagy jsou klíčová slova přiřazená informačnímu zdroji, tak polička je klíčové slovo, *kterému* jsou přiřazeny informační zdroje. Důraz je oproti tagům otočený, základem je „umístění“, kam uživatelé přiřazují informační zdroje. Můžeme použít analogii s knihovnickým pojmem kategorie, který chápeme jako „*skupinu ekvivalentních objektů*“ (Kučerová, 2017), tedy nejprve máme připravené poličky a těm pak přiřazujeme informační zdroje. Často má také polička svůj popis, kde může být uvedeno např. proč vznikla nebo jaké informační zdroje obsahuje.

Řazení do poliček může mít podobně jako folksonomie dvě formy: úzkou a širokou. Úzká forma řazení do poliček umožňuje existenci jedné poličky v systému pouze jednou, pak jde o seznam, na který může jeden nebo více uživatelů přidávat konkrétní informační zdroje. V jedné poličce je sdruženo více informačních zdrojů a je patrné, který byl přidán kterým uživatelem. Tuto formu využívá např. systém LibraryThing, příkladem poličky je pak „Tour of Alabama“, která obsahuje knihy odehrávající se na území tohoto amerického státu. V rámci široké formy řazení do poliček může mít každý uživatel svou poličku se stejným názvem. Pak je možné sdružovat uživatele, kteří přiřadili stejný informační zdroj na stejnou poličku. Tuto formu využívá např. česká služba Databáze knih, jedna z využívaných poliček je „Chystám se číst“.

Teď už se ale vrátíme k tagům a folksonomiím. V minulosti byly folksonomie odborníky považovány spíše za zdroj pobavení a příklady hovorových výrazů než za systém organizace informačních zdrojů. Za přelom v jejich přijímání můžeme považovat rok 2006, kdy Golder a Huberman vydali studii folksonomických dat ze služby Delicious. Dokázali, že se tagy dají rozdělit do koherentních skupin a že

popisují základní vlastnosti informačních zdrojů (odpovídají na otázky kdo, co, kde, kdy, proč a jak). Z toho bylo zřejmé, že mohou sloužit jako doplněk k formálním bibliografickým záznamům. (Manzo, 2015)

I přesto, že už dnes tagy a folksonomie nejsou brány na lehkou váhu, považujeme za důležité zdůraznit vliv různých úhlů pohledu, které se při jejich tvorbě uplatňují. Popis jakýchkoli entit obecně je komplikovaná záležitost, různí lidé je mohou interpretovat, pojmenovat a popsat různě. Proto profesionální tvorba metadat může být aplikována pouze v určitém kontextu a prostředí. U folksonomií takový kontext určený není, každý uživatel taguje informační zdroje ze svého úhlu pohledu. Z tohoto důvodu mohou být folksonomie vnímány jako chaotické, ale pokud chaos dokážeme kontrolovat, stane se z něj výhoda folksonomií. Mai k tomu konstatuje, že lidé jsou dobří v rozpoznávání, o čem informační zdroje jsou, že mají mnoho nápadů, jak jejich obsah vyjádřit, a že tagy se mění v závislosti na kontextu, cíli a času. (Mai, 2011, s. 116-117)

Mai také poznamenává, že se rozmazává hranice mezi profesionálem a uživatelem. Profesionála lze odlišit pouze tím, že reprezentace či organizace informačních zdrojů pro druhé je jeho placenou prací. Avšak uživatelé mohou být profesionály v tom smyslu, že tagují informační zdroje z oboru, ve kterém jsou vzdělaní, přestože nemají vzdělání knihovnického a/nebo informačního směru. (Mai, 2011, s. 117)

Za zmínku také stojí pojem relevance. Relevance (a také důležitost) tagu je určena frekvencí jeho použití, jelikož každý tag může být každým uživatelem k jednomu informačnímu zdroji přidán pouze jednou. Skrze tagování vlastně uživatelé „hlasují“ o obsahu zdroje, nejčastěji použité tagy mu odpovídají nejpřesněji. S relevancí tagů také velmi často souvisí relevance informačního zdroje, čím vícekrát má informační zdroj přiřazený určitý tag, tím relevantnější danému tématu je. Tento princip může fungovat jako doporučení tagů (ve fázi tagování) i informačních zdrojů (ve fázi vyhledávání). (Peters, 2009, s. 161)

S tím souvisí také výzkum spolehlivosti tagů. Přestože žádný z uživatelů netvoří tagy s myšlenkou, aby se shodoval s ostatními uživateli, celkový konsenzus je poměrně stabilní. Menšinové názory tak mohou existovat vedle populárních, aniž

by docházelo ke znehodnocení relevance a spolehlivosti tagů. (Manzo, 2015; Smith, 2008, citováno podle Peters, 2009, s. 157)

Obecně jsou odlišovány dva základní typy folksonomií² (Rafferty, 2018, s. 501; Peters, 2009, s. 164-165):

1. Široká folksonomie umožňuje, aby více uživatelů mohlo přidat ke stejnému informačnímu zdroji stejný tag. Každý uživatel může zdroj tagovat ze svého úhlu pohledu bez ohledu na již existující tagy. To umožňuje měření frekvence tagů, jak jsme o něm mluvili výše, a zároveň je předpokladem pro vytvoření tzv. long tail efektu³. Měření frekvence tagů a jeho využití pro kolektivní inteligenci je klíčovou výhodou širokých folksonomií. Příkladem takové folksonomie je služba Delicious.
2. Úzká folksonomie umožňuje každý tag k danému informačnímu zdroji přidat pouze jednou, mohou tedy být přidávány jedině tagy odlišné. Měření frekvence a long tail efekt jsou tímto znemožněny, díky tomu uživatel, který v systému vyhledává, nedokáže rozlišit relevanci jednotlivých tagů ani informačních zdrojů. Přidávání tagů také může být omezeno pouze na tvůrce zdroje. Většinou se zde vyskytuje menší počet tagů než v širokých folksonomiích. Příklady tohoto typu folksonomie jsou Flickr, Technorati a YouTube.

Na rozdíl od typů folksonomií, na kterých panuje obecná shoda, existuje mnoho názorů, jak dělit tagy. Je to však proces umělý a probíhající až zpětně u existujících tagů, jelikož uživatelé při jejich přiřazování informačním zdrojům žádné kategorie netvoří. Na následujících řádcích se pokusíme sjednotit různorodé typologie a kategorizace tagů (Gupta, 2011, citováno podle Rafferty, 2018, s. 503; Golder, 2006, citováno podle Manzo, 2015; Dubinko, 2006, citováno podle Peters, 2009, s. 196; Winget, 2006, citováno podle Peters, 2009, s. 196; Al-Khalifa, 2007, citováno podle Peters, 2009, s. 196-197; Peters, 2009, s. 197):

² Myšlenkovým otcem tohoto rozdělení (stejně jako folksonomií obecně) je americký informační architekt Thomas Vander Wal.

³ Tento termín se vztahuje k rozložení pojmů a v tomto případě označuje skutečnost, že přibližně 80 % tagů je tvořeno tagy velmi málo využívanými, zatímco tagy velmi využívané tvoří pouze 20 % celkového množství. (Dlouhý chvost, 2019)

- Faktické tagy identifikují faktické informace o informačním zdroji, objektivně ho popisují. Můžeme identifikovat tagy, které se věnují různým faktorům informačního zdroje:
 - Prvním faktorem, kterému se faktické tagy mohou věnovat, je kontext informačního zdroje. Takové tagy specifikují okolnosti vytvoření nebo nahrání informačního zdroje.
 - Mezi kontextové tagy řadíme tagy geografické. Mohou určovat místo vzniku informačního zdroje (např. fotografie) nebo např. místo, o kterém se mluví v textu.
 - Časové určení také řadíme mezi kontextové tagy. Nemusí být nutně ve tvaru data, může jít i o textovou specifikaci času, např. Samovy narozeniny.
 - Dalším typem faktických tagů jsou tagy určující obsah informačního zdroje. Můžeme je rozdělit na několik skupin podle toho, jakému obsahu se věnují:
 - lidé, příp. známé osobnosti,
 - věci,
 - události nebo svátky apod.
 - Další vlastnosti informačního zdroje mohou být také specifikovány v rámci faktických tagů. Nemusí pocházet přímo z informačního zdroje, jako např. informace o autorovi.
 - Zvláště můžeme vyčlenit informaci o vlastníkovvi informačního zdroje.
 - Řadíme sem také informace, které propojují informační zdroj s dalšími souvisejícími objekty.
- Osobní tagy se týkají uživatele, který je tvoří, souvisí se vztahem mezi uživatelem a daným informačním zdrojem. Často jsou pro ostatní uživatele nedůležité a mohou se měnit v čase. Opět je můžeme dělit na několik dalších typů:
 - Jako subjektivní tagy označujeme názory a emoce uživatele, mohou také obsahovat doporučení informačních zdrojů.

- Dalším typem jsou tagy vztahující se k osobě tvůrce, řadíme sem tagy, ve kterých uživatel zmiňuje sám sebe. Většinou obsahují slova „můj“ nebo „moje“, např. „moje dokumenty“.
- Organizace činností je posledním typem osobních tagů. Jejich cílem je připomenout uživateli činnosti, které chce s daným informačním zdrojem provést. Jako příklad tohoto typu tagů můžeme uvést „k přečtení“ nebo „poslat Báře“.
- Individuální tagy jsou samostatně vytvořené za nějakým konkrétním účelem, mohou být zcela unikátní a ostatním uživatelům nemusí dávat smysl. Mohou být vytvořeny jednotlivými uživateli (např. „úkol na zeměpis 21.5.2018“ nebo „xx5136“) nebo komunitami (např. „Faces_in_holes“).

Stejně jako typy tagů, může být rozličná i motivace uživatelů k tagování (Gupta, 2011, citováno podle Rafferty, 2018, s. 509):

- budoucí vyhledání informačních zdrojů,
- sdílení informačních zdrojů,
- přitahování pozornosti,
- hra a/nebo soutěžení,
- prezentace svojí osoby (platí hlavně pro tagy vztahující se k osobě tvůrce),
- vyjadřování názoru,
- organizace činností,
- vydělání peněz,
- jednoduchost tagování.

U některých individuálních tagů může být ale zároveň důležitou motivací udržení soukromí. Uživatelé mohou záměrně vytvářet tagy, které jsou srozumitelné jen pro ně nebo jejich blízké známé. (Ames, 2007, s. 978, citováno podle Peters, 2009, s. 198)

Přesto ale tagování není tak jednoduché, jak by se mohlo na první pohled zdát. Není to technologická stránka procesu, která vykazuje vyšší míru náročnosti, ale kognitivní. Müller-Prove ověřil, že s množstvím informačních zdrojů a tagů přímo úměrně stoupá i kognitivní náročnost procesu tagování. (Müller-Prove, 2008, citováno podle Peters, 2009, s. 163)

A pokud bychom hodnotili kvalitu indexace, je zřejmé, že je závislá na uživatelích, kteří tagy vytváří, a také na kvalitě vybraných tagů. A zároveň stejně jako jakýkoli jiný systém organizace znalostí tagování může privilegovat některé pohledy na svět na úkor jiných. Následně můžeme také říci, že úspěšnost sociálního systému sloužícího k tagování závisí na kvalitě a aktivním zapojování uživatelů. Existují herní přístupy, které slouží k povzbuzování uživatelů a tím k vylepšování tagování. (Rafferty, 2018, s. 502, s. 510)

Nyní se zaměříme na shrnutí výhod tagování a folksonomií (Peters, 2007):

- reprezentují autentické používání jazyka,
- umožňují více interpretací,
- rozšiřují přístup k informačním zdrojům,
- jde o levnou metodu indexace (díky tomu, že je rozdělena mezi velké množství lidí),
- jsou jedinou možností, jak tvořit metadata pro velké objemy informací na internetu,
- slouží jako zdroje pro vývoj systémů organizace znalostí a nástrojů řízené slovní zásoby,
- umožňují vyhledávání i prohlížení, a to i velmi specifické,
- umožňují použití neologismů,
- pomáhají identifikovat komunity,
- jsou zdrojem pro kolaborativní doporučovací systémy,
- vylepšují se s narůstajícím počtem aktivních uživatelů,
- umožňují kontrolu kvality informačních zdrojů přímo uživateli,
- rozšiřují povědomí o existenci a potřebnosti indexace.

A stejně tak bychom zde rádi shrnuli i nevýhody tagování a folksonomií (Peters, 2009, s. 227; Peters, 2007, citováno podle Rafferty, 2018, s. 505; Kroski, 2005, citováno podle Rafferty, 2018, s. 504):

- absence řízené slovní zásoby,
- absence přesnosti a hierarchie,
- neexistuje kontext indexace,
- různé úrovně popisu (jsou používány zároveň široké i úzké termíny),
- jsou používány různé jazyky,

- nejsou prozkoumané skryté paradigmatické vztahy,
- formální a předmětový popis se nedá oddělit,
- existence individuálních tagů, tagů použitých jako spam a dalších nejasných slov,
- informační zdroje jsou popisovány pouze jako celek,
- náchylnost k neetickému chování uživatelů.

Budoucnost tagování se v tuto chvíli nedá přesně určit, zejména co se týče jeho vztahu k profesionální tvorbě metadat. Na toto téma však existují různé názory a případné využití tagů v knihovních systémech je námětem mnoha odborných diskusí. Vaidya a Harinarayana ve své studii uvádí, že silnou stránkou folksonomií je spolupráce více lidí na indexaci. Za jejich slabou stránku naopak považují nedostatečnou přesnost při vyhledávání. Tvrdí, že tagy mohou doplnit profesionální metadatové výroky, nicméně není pravděpodobné, že by je zcela nahradily. (Vaidya, 2016, citováno podle Rafferty, 2018, s. 502)

Bogers a Petrasová uvádí, že obsah tvořený uživateli může doplňovat a obohacovat klasické vyhledávací systémy, přestože odborná literatura naznačuje, že si při vyhledávání vedou hůře než řízená slovní zásoba. Raffertyová doplňuje, že tagování umožňuje lépe zachytit aboutness a emoce spojené s informačním zdrojem, zejména pokud jde o zdroje jiné než textové povahy. (Bogers, 2015, citováno podle Rafferty, 2018, s. 510; Rafferty, 2018, s. 510)

Manzoová et al. se přidávají, když říkají, že tyto dva systémy se nemusí vzájemně vylučovat, přestože se to tak mnohým odborníkům zdá. Folksonomie i řízená slovní zásoba mají své výhody i nevýhody. Systémy profesionálních metadat mají vyšší spolehlivost a strukturovanost, ale na druhou stranu je časově i finančně náročná jejich tvorba a údržba. Mohou také být nepříliš uživatelsky přívětivé, naopak tagování je přívětivé a uživateli hojně využívané, jak dokazuje jeho četnost na sociálních sítích. A zároveň jde o ekonomicky i časově výhodnější variantu. Zároveň tento autor uvádí, že pokud se profesionální a uživatelská metadata použijí společně, dojde ke zlepšení vyhledatelnosti informačních zdrojů, k prevenci neúspěšných vyhledávacích dotazů a ke zvýšení přesnosti systému. Také to podle něj vypadá, že knihovníci a uživatelé vytvářejí velmi podobné rozložení užitečných

metadat. Tyto a další podobné výsledky mohou vést ke změně pohledu na organizaci informačních zdrojů. (Manzo, 2015)

Pokud bychom uznali, že použití klasických profesionálních metadat a tagů společně může přinést systému přidanou hodnotu, pořád stojíme před problémem, že nemáme k dispozici vhodný systém. Výzvou pro vývojáře informačních systémů je vytvoření takového systému, který bude vhodně propojovat tyto dva světy metadat tak, aby nedošlo k potlačení výhod ani jednoho z nich a k dominanci určitých pohledů na svět nad jinými. (Rafferty, 2018, s. 511)

Možnost veřejně zpřístupňovat jakékoli informační zdroje na internetu, to, že prakticky kdokoli může předávat světu svůj obsah, je velkým úspěchem lidstva jako celku. Ale kromě nových možností to přináší také nové výzvy, kterým budeme do budoucna čelit. (Wyrwoll, 2014, s. 25)

3. Přehled dosavadních výzkumů

V této kapitole jsme se pokusili vybrat nejdůležitější výzkumy z oblasti tvorby obsahových metadat a předložit jejich závěry. Pro tyto účely bylo vybráno 27 zdrojů z posledních deseti let. Pro přehlednost jsme kapitolu rozdělili do čtyř částí: 3.1 obsahová metadata tvořená uživateli obecně, 3.2 obsahová metadata – řešení v knihovnách a vědeckých institucích, 3.3 obsahová metadata – řešení v komerčním prostředí a 3.4 srovnání více způsobů tvorby obsahových metadat.

Významným fenoménem, který se v odborné literatuře začíná objevovat, jsou obsahová metadata generovaná za pomoci umělé inteligence a strojového učení. Budoucnost těchto metod i to, zda budou schopné napodobit obsah vytvářený člověkem, jsou v současnosti nejisté a v této diplomové práci se jimi nebudeme zabývat.

Byla provedena rešerše, která sestávala z replikace předem stanovených vyhledávacích dotazů v množství odborných fulltextových nebo bibliografických databází oboru informační věda a knihovnictví a některých příbuzných oborů. Vyhledávací dotazy byly vytvořeny tak, aby pokryly celou šíři tématu, které nás v tomto případě zajímá, tedy tvorbu obsahových metadat různými skupinami profesionálů a uživateli (tedy laiky).

Vyhledávací dotazy, které byly pro rešerši použity (anglická mutace):

- („content metadata“ OR „subject metadata“) AND („tagging“ OR folksonom*);
- („content metadata“ OR „subject metadata“) AND (professional* OR librarian* OR catalogiza*);
- („content metadata“ OR „subject metadata“) AND („reader* platform“ OR „library catalog*“ OR OPAC OR bookshop OR publish*).

Vyhledávací dotazy, které byly pro rešerši použity (česká mutace):

- („obsahová metadata“) AND (tag* OR folksonom*);
- („obsahová metadata“) AND (profesionál* OR knihovn* OR katalogiz*);
- („obsahová metadata“) AND („čtenářsk* portál*“ OR katalog* OR OPAC OR knihkupectví OR nakladatelství OR vydavatelství).

V některých případech bylo zjištěno, že databáze neumožňuje vyhledávat pomocí uvozovek. V takovém případě byly termíny v uvozovkách nahrazeny booleovským operátorem AND, např. místo „obsahová metadata“ bylo použito (obsahová AND metadata). V některých případech bylo také zjištěno, že vyhledávání v databázi nepodporuje uvozovky ani logické operátory. V takovém případě byla databáze z rešerše vynechána z důvodu přílišné obtížnosti nalezení relevantních dokumentů. V následujícím seznamu nejsou tyto databáze zahrnuty.

Následuje abecedně řazený seznam názvů databází, ve kterých byla rešerše provedena⁴:

1. Academic Search Ultimate;
2. CEJSH;
3. Classiques des sciences sociales;
4. Classiques Garnier Numérique;
5. Directory of Open Access Books (DOAB);
6. Directory of Open Access Journals (DOAJ);
7. EBSCO eBooks;
8. ERIC - Education Resources Information Center;
9. Humanities Source Ultimate;
10. IBZ Online;
11. JSTOR;
12. Library, Information Science and Technology Abstracts (LISTA);
13. Library & Information Science Source (LISS);
14. OAPEN Library;
15. Oborová brána Knihovnictví a informační věda (KIV);
16. Open Library;
17. Oxford Bibliographies Online;
18. Oxford Journals Online Full Collection;
19. Oxford Scholarship Online;
20. Publicly Available Content Database (PAC);

⁴ Názvy databází byly převzaty z Portálu elektronických zdrojů Univerzity Karlovy, viz <https://ezdroje.cuni.cz/>. Odtud je také možný k těmto databázím přístup pro příslušníky Univerzity Karlovy.

21. SAGE Journals;
22. Taylor & Francis Online.

Během rešerše byly vybrány proběhlé výzkumy, resp. články o nich, podle následujících kritérií: (1) musely být relevantní k vyhledávanému tématu, (2) nesměly být starší deseti let, a zároveň (3) muselo se jednat o výzkum, nikoli o monografii. Hranice deseti let byla stanovena proto, abychom nebyli zahlceni množstvím výsledků, které vzhledem k vývoji problematiky už nejsou dále relevantní. Následně byly zdroje rozděleny do čtyř kategorií podle toho, čemu přesně se věnují a byly stručně parafrázovány do češtiny tak, aby byla podána souhrnná informace o výsledcích výzkumu.

3.1 Obsahová metadata tvořená uživateli obecně

V mnoha člancích publikovaných na téma tagování je patrná určitá změna paradigmatu, která naznačuje, že ne vždy jsou metadata tvořená odborníky dostačující, a že tagy a další obsah tvořený uživateli může nabídnout možnosti, které dříve neexistovaly. A také naznačuje, že uživatelé mají o tyto možnosti zájem – a to jak z hlediska vytváření obsahových metadat, tak z hlediska jejich využívání – a že se tagování bude i nadále rozvíjet.

Cílem Grzegorze je popsat fenomén sociální organizace informací v síťovém prostředí a související nástroje a aplikace. To zahrnuje pojem sociální katalogizace, metadata tvořená uživateli a crowdsourcovaná metadata v knihovních projektech. Výzkum je založen na zkoumání literatury (literature review), kritické analýze a funkční analýze softwaru a/nebo webových služeb, kde to bylo možné. V posledních letech se odehrála velká změna: málo oblíbené stránky často mizí a jsou nahrazeny novými, uživatelsky přívětivými nástroji. Existují aplikace, které uživatelům pomáhají sbírat, spravovat a sdílet obsah, funkcionalita některých z nich je založena na otevřenosti, webových komunitách a myšlence sémantického webu. (Grzegorz, 2016)

Podle výzkumu Benoita a Munsonové se mezi lety 2010 a 2016 značně zvýšila zkušenost účastníků výzkumu s tagováním současně s tím, jak hodně tuto metodu využívají, zároveň ale došlo také k posílení obav týkajících se jeho kontroly, konzistence a potenciálního zneužití. (Benoit, 2018)

Pablo a Goicoecheaová se zabývají elektronickou literaturou a tím, jak je popisována a jak k ní lze přistupovat. Podle nich jde tagování elektronické literatury za hranici klasifikace a je základem pro získávání zdrojů z databází. Dále popisují protichůdné tendence tagování elektronické vědecké literatury, které lze v různých sbírkách v praxi pozorovat: taxonomie tagů versus rovnocenné horizontální tagy nebo uzavřené versus otevřené sady tagů. Přílišná uzavřenost a hierarchizace i přílišná otevřenost a horizontalita mohou způsobit, že některá díla mohou být nedosažitelná – klíčem je najít střední cestu. Autoři doporučují databázovou strukturu, která je dostatečně flexibilní. (Pablo, 2014)

Joorabchi, English a Mahdi se snažili nalézt řešení na problémy, které vznikají nekontrolovanou povahou uživatelského tagování. Různé variace zápisu, synonyma nebo akronyma mohou způsobit např. fenomén exploze tagů⁵. Navrhují mapovat tagy na odpovídající články na Wikipedii, jakožto dobře utvořené pojmy. Má to mnoho výhod: normalizace nesrovnalostí, eliminace osobních tagů a zlepšení možností sdílení existujících metadat mezi systémy. Autoři článku navrhnou metodu založenou na strojovém učení, která je schopná automaticky přiřazovat tagům odpovídající pojmy na Wikipedii. Byla vytvořena aplikace a otestována na 20 milionech příspěvcích na platformě StackOverflow.com. (Joorabchi, 2015)

Dennenová, Bagdyová a Catesová se zaměřily na tagování výukových zdrojů studenty, konkrétně porovnání tagování s pomocí slovníku, volného tagování a kombinace obou možností (hybridní tagování). Analyzovaly přesnost, počet tagů a popularitu jednotlivých typů tagů. Jejich výzkum ukázal, že při tagování pouze pomocí slovníku studenti používali méně tagů a zároveň méně tagů použitých pouze jednou. Používali také užitečné typy tagů jako typ obsahu, které studenti, kteří tagovali volně, nezahrnuli. Na druhou stranu volné tagování přineslo také své výhody: studenti měli možnost použít tagy, které odrážely jejich zájmy, a také tagy se specifickými tématy, které nebyly zahrnuty ve slovníku. Největší potenciál má při opatrné výuce a aplikaci hybridní přístup. Výzkum také potvrdil, že většina

⁵ Anglicky tag explosion. Je obdobou pro termín information explosion, česky překládáno jako informační exploze.

studentů se naučí kompetentně tagovat i s relativně málo instrukcemi. (Dennen, 2018)

Podobným tématem se zabývá také článek Johnsona/ové, který porovnává výhody a nevýhody tří stejných přístupů (tagování se slovníkem, volné tagování a hybridní tagování) u tagování fanfikce, tj. „*literární tvorba, při níž autor používá postavy a prostředí z již existujícího díla, jehož je fanouškem*“ (Auzká, 2016). Potřeby čtenářů fanfikce se ale poněkud liší od čtenářů beletrie obecně: potřebují být schopni rychle a efektivně nalézt příběhy se specifickými kritérii a také předem vědět, jak se ostatním čtenářům daný text líbil. To může být v rozporu s potřebami autorů fanfikce, kteří chtějí zůstat nezávislí a bez vnější kontroly jejich děl. Různé přístupy k tagování reprezentují právě tyto odlišné potřeby fanfiction komunity ohledně organizace informací a vyhledatelnosti. (Johnson, 2014)

Další výzkum, který se zabývá uživatelským obsahovým popisem beletrie, publikovala Šauperlová. Zaměřuje se v něm na analýzu uživatelských tagů a recenzí románů. Kromě názvů děl a jmen jejich tvůrců nejčastěji tagy/recenze obsahují: jména postav, geografické názvy, časové období, ve kterém se kniha odehrává, a literární žánr. Také zmiňují ocenění, které kniha získala, svůj názor na román a propojují je s ostatními díly v sérii nebo s jinak příbuznými knihami nebo filmy. Autorka navrhuje, aby se časové období, žánr a ocenění staly součástí funkčních požadavků na bibliografické záznamy. (Šauperl, 2012)

Výzkum Alsarhana se zaměřuje na vliv cílové skupiny na tagování, konkrétně na to, zda předpokládaná sdílená znalost mezi tvůrcem tagů a jejich cílovou skupinou ovlivňuje volbu a přiřazování tagů. Účastníci výzkumu měli tagovat stejné zdroje pro jednu z cílových skupin: pro sebe, pro kolegu a pro neznámého člověka. Autor se zaměřil především na počet tagů, počet pojmů reprezentovaných tagy a specifickou tagů v souvislosti s jednotlivými cílovými skupinami. Podle závěrů výzkumu se ve větší míře v závislosti na cílové skupině lišil pouze počet přiřazených tagů. (Alsarhan, 2013)

Choiová, Lee, Hu a Downie se ve svém výzkumu zabývají obsahovým přístupem k hudbě. Snaha o vytvoření automaticky generovaných obsahových klasifikací písní se zakládala na derivaci obsahu ze dvou zdrojů, které obsahují informace o obsahu písní: text písně a uživatelské interpretace. Autorka zjistila, že větší

přesnost klasifikace měly vždy informace pocházející z uživatelských interpretací oproti informacím pocházejících z textu písní. To naznačuje, že uživatelské interpretace lépe vyjadřují význam písně, protože text bývá poetický, tzn. sémanticky nejednoznačný. (Choi, 2016)

3.2 Obsahová metadata – řešení v knihovnách a vědeckých institucích

Když pomíneme automatizované metody, knihovny a vědecké instituce používají různé způsoby tvorby obsahových metadat. Jejich přijetí uživateli se může lišit.

Základní přehled současné situace nabízí ve svém článku Golubová. Konstatuje, že přestože předmětový přístup k informačním zdrojům je důležitým úkolem informačních institucí, tak je v dnešních digitálních knihovnách a integrovaných vyhledávacích rozhraních značně omezený. Často není postupováno systematicky a v integrovaných vyhledávacích rozhraních není zajištěno mapování mezi použitými systémy organizace znalostí. Tyto nedostatky pramení pravděpodobně z nedostatku prostředků, ale podle autorky současný stav nespĺňuje ani základní předpoklady předmětového přístupu. Za nejdůležitější možnosti nápravy předmětového přístupu v knihovnách autorka považuje automatickou indexaci a sociální tagování, ale tato oblast je velmi málo zkoumaná, což je třeba napravit. (Golub, 2016)

Studie Parkové je zaměřena na aktuální praktiky při tvorbě obsahových metadat napříč digitálními repozitáři a kolekcemi. Výzkum byl proveden pomocí dotazníku pro katalogizátory a metadatové profesionály v celých Spojených státech amerických. Z výsledků bylo zjištěno, že nejčastěji používanými metadatovými schémata jsou MARC, AACR2 a LCSH, dále Dublin Core, EAD, MODS, VRA Core a TEI. Při výběru metadatového schématu hrají roli specifika popisované sbírky, jako jsou typ zdrojů, povaha kolekce nebo potřeby uživatelské komunity. Důležité jsou také existující technologické zajištění a zkušenosti zaměstnanců. Interoperabilita obsahových metadat stále zůstává výzvou pro odbornou komunitu kvůli nedostatku sdílených mechanismů pro lokálně přidané metadatové elementy. (Park, 2010)

Podobným tématem se zabývá i Walsh, který věnuje svůj článek použití LCSH v digitálních sbírkách. Popisuje výhody a nevýhody použití tohoto řízeného

slovníku, diskutuje možnosti vylepšení předmětového přístupu u systémů, které LCSH využívají, a využití jiných klasifikací pro předmětový přístup. LCSH je v současnosti nejvyužívanějším řízeným slovníkem pro obsahová metadata, i přes problémy s jeho užíváním, nízkou uživatelskou přívětivostí, a tím i se špatnou přesností a úplností. Způsobem, jak využít LCSH v digitálních knihovnách vylepšit, je podle autora jeho použití ve fasetovém vyhledávání a automatické rozpoznávání vzorců za použití ontologie založené na LCSH. (Walsh, 2011)

Tosakaův výzkum se zaměřuje na obsahově obohacené bibliografické záznamy a jejich vliv na cirkulaci dokumentů. Byla zkoumána data ze středně velké akademické knihovny, a to jak souvislost obohacených metadat s mírou půjčovnosti, tak také uživatelské dotazy v OPACu za účelem zjištění, jak jsou obohacená metadata používána. Z výzkumu vyplývá, že obohacené záznamy jsou opravdu asociovány s vyšší mírou cirkulace a že obsahová metadata hrají klíčovou roli v nalezitelnosti dokumentů. (Tosaka, 2011)

S tím souvisí i výzkum Zavalinaové, který se zabývá tématem předmětového přístupu k dokumentům a tím, jak k jeho zlepšení přispívají katalogizační pravidla RDA. Autorka jako analytický rámec použila model FRBR a zkoumala uživatelské dotazy v large-scale digital library Opening History⁶. Data od 1. ledna do 31. prosince 2010 byla analyzována metodou analýzy transakčního logu. Data transakčního logu byla získána pomocí aplikace Google Analytics. Výzkum ukázal, že uživatelské dotazy zahrnovaly entity ze všech tří skupin modelu FRBR: pojem, objekt, akce, místo, osoba, korporace, dílo. Dále byly také identifikovány dvě nové entity: třída osob a etnická skupina. Výzkum prokázal důležitost vztahů mezi obsahovými entitami. (Zavalina, 2014)

Studie Angelové se zabývá porovnáním popisného tagování profesionálů v knihovnách, archivech a muzeích. Klade si za cíl určit míru podobnosti mezi skupinami při přiřazování tagů objektům, které se mohou vyskytovat ve všech třech

⁶ Digitální knihovna Opening History byla největším agregátorem obsahu zabývajícím se historií USA. V roce 2012 se stala součástí mateřské IMLS Digital Collections and Content (Zavalina, 2014). Institute of Museum and Library Services (IMLS) je organizace, jejímž cílem je podporovat a posilovat americké knihovny a muzea pomocí grantů, výzkumu a vývoje politik (Institute of Museum and Library Services, 2019).

typech institucí, a dále objektům, které se typicky vyskytují v jednom z typů institucí. Podle výsledků výzkumu autorka určila, že mezi skupinami informačních pracovníků se vyskytovaly pouze malé rozdíly v hloubce indexace jednotlivých typů objektů. Na různé typy objektů však odborníci aplikovali různé úrovně popisu. Na úroveň popisu mělo vliv několik faktorů: trojrozměrné nebo fyzické médium, digitální zástupce, aboutness objektu, technika a materiály použité k výrobě fyzického objektu a psaný text. Datová analýza také indikuje významný rozdíl mezi celkovým počtem popisných tagů a počtem přesně se shodujících primárních tagů u jednotlivých typů objektů. (Angel, 2012)

I některé knihovny a vědecké instituce implementovaly do svých služeb možnost uživatelského tagování. Kombinovaný přístup popisují ve svém článku Hypénová a Mäkelä. Pojednávají o aplikaci pro nový způsob práce s obsahovými metadaty beletrie – projekt si klade za cíl spustit nástroj s názvem Kirjasampo, který bude založen na datovém modelu sémantického webu a bude umožňovat flexibilní linkování metadat získaných z různých zdrojů. Aspekty, které se liší od dřívějších katalogů, jsou zejména: odkazy na jiná díla, vnímání díla čtenáři a kritiky, klasifikace, obsahový popis a kulturní kontext, ve kterém dílo vzniklo a který reflektuje. Hodnocení a doporučování děl bude převzato z předchozí aplikace Kirjasto. Jedná se o první finskou platformu, která shromažďuje a ukládá data od knihoven i uživatelů a jejímž cílem je kumulovat a distribuovat literární znalosti a zkušenosti. (Hypén, 2011)

Bundzaová popisuje platformu pro institucionální repozitáře a způsob, jakým se vloženým dokumentům přiřazují obsahové metadatové výroky. Platforma The Digital Commons nabízí třívrstvou taxonomii akademických disciplín, kterou uživatelé používají při vkládání dokumentů do repozitáře. Tento způsob přiřazování obsahových metadatových výroků umožňuje jednotnou organizaci materiálů napříč repozitáři, které tuto platformu využívají, a zlepšuje viditelnost výzkumu pro vyhledávače. (Bundza, 2014)

Studie Yangové poskytuje náhled na současné tagovací praktiky v knihovních systémech a dokazuje, že tagy jsou užitečnými obsahovými metadaty. I přes to ale pouze 47 % discovery systémů a 5 % knihovních systémů mají možnost tagování. Dokonce při zkoumání knihoven, které implementovaly knihovní systém Koha,

autorka zjistila, že pouze 49 % knihoven tagování umožnilo a 51 % ho zakázalo. Je třeba další výzkum příčin tohoto odmítání, protože může existovat více různých vysvětlení. Z výzkumu je zřejmé, že při uživatelském tagování mohou vznikat různé nedostatky, autorka navrhuje dvě možnosti jejich řešení: zkombinovat uživatelské tagy s klíčovými slovy extrahovanými z MARC záznamů nebo navázat uživatelské tagy na LCSH. Knihovny a poskytovatelé by měli spolupracovat na co nejlepším využívání uživatelských tagů. Výzkum také ukázal, že uživatelé veřejných knihoven tagují častěji než uživatelé akademických knihoven. (Yang, 2012)

3.3 Obsahová metadata – řešení v komerčním prostředí

Rešerše ukázala, že odborná literatura na téma zpracování metadat v komerčním prostředí téměř neexistuje a bylo by třeba dalšího výzkumu, abychom zjistili, proč tomu tak je. Jednou z pravděpodobných příčin může být to, že se z konkurenčních důvodů nechtějí o své postupy a metody dělit, což není z vědeckého hlediska ideální. Proto přinášíme alespoň dvě zajímavá témata, která s oblastí našeho zkoumání souvisí.

Hutchens zmiňuje důležitost obsahových metadat v kontextu open access informačních zdrojů. Vzhledem k různým definicím open access je pro knihovny a potenciální uživatele velmi obtížné zjistit, které zdroje jsou open access a které nikoliv, situace se ještě zhoršila s nástupem hybridních časopisů. Na vylepšení situace pracuje NISO KBART workgroup, NISO's Open Discovery Initiative a NISO's Open Access Initiative. (Hutchens, 2013)

Beardsleyův článek diskutuje strategické spojení mezi vydavateli a technologickými firmami. To by podle autora mohlo spočívat ve výměně produktů a služeb, které by přinášely zvýšení hodnoty a sdílení rizik oběma partnery. Autor uvádí příklad na středním vydavatelství, které by díky rozhodnutí přenechat kompozici stránek specializované firmě mohlo ušetřit až milion dolarů za rok. Ke zlepšení dochází také v celkovém čase a kvalitě procesů, nicméně je třeba předem dobře zvážit, pro které procesy se outsourcing vyplatí. Tato strategická spojení se bezpochyby dají uplatnit i v oblasti tvorby obsahových metadat. (Beardsley, 2016)

3.4 Srovnání více způsobů tvorby obsahových metadat

Přehled literatury zakončíme částí, kde budeme srovnávat různé způsoby tvorby obsahových metadat mezi sebou, zejména tvorbu profesionálů s tvorbou uživatelů. Výzkumy obecně dokazují vzrůstající tendenci k užití tagování jako právoplatné metody tvorby obsahových metadat a ukazuje se, že kombinace tradičních obsahových metadat s uživatelskými tagy v praxi reálně zvyšuje úspěšnost vyhledávání. Autoři se shodují na tom – a často na to kladou důraz –, že spolupráce profesionálů s uživateli je možná a žádoucí.

Lawsonová provedla výzkum, ve kterém porovnávala předmětová hesla LCSH s tagy na platformách Amazon.com a LibraryThing u stejných titulů. Zkoumáno bylo 155 knih ze 31 předmětových divizí katalogu WorldCat, klíčovým parametrem výběru bylo to, aby knihy byly ve fondu mnoha knihoven různých typů. Z výsledků výzkumu bylo zjištěno, že přestože uživatelské tagy se skládají z určité části z tagů subjektivních, existuje zde i dostatek objektivních tagů, pro které existují odpovídající hesla v LCSH, ale které v profesionálním popisu použity nebyly. Z toho je jasné, že uživatelské tagy mohou být použity k obohacení bibliografických záznamů nebo mohou katalogizátorovi navrhnout nové přístupové body. (Lawson, 2009)

Zervas, Sampson a Asteridouová se ve své studii soustředili na vliv různých stylů tagování (popis a kategorizace) na metadata digitálních elektronických zdrojů. Výzkum byl proveden na datasetu existujícího Learning Object Repository⁷ – OpenScienceResources Repository. Bylo zjištěno, že uživatelské tagy popisné se značně liší od formálních metadatových výroků, metadatovým výrokům tvořeným profesionály odpovídá z větší části pouze kategorizační část tagování. Autoři dále zkoumají motivaci k tagování a zjišťují, že má přímý vliv na rozšíření metadat digitálních elektronických zdrojů a také na výslednou folksonomii, pokud ji porovnáme s formálními strukturovanými slovníky, které používají profesionálové. (Zervas, 2012)

⁷ Do češtiny můžeme přeložit jako úložiště výukových objektů, nicméně terminologie není ustálená.

Whiteová se ve své práci zabývá repozitáři a rozdílem ve tvorbě metadat mezi profesionály a vědci, tedy uživateli. Porovnávala, jak obě skupiny vytváří popisná a obsahová metadata vědeckých dokumentů, konkrétně byl výzkum proveden na datasetech ptáci a myslivost. Výzkum ukazuje, že profesionálové používali řízené slovníky pro popisná metadata častěji než vědci, nejčastěji používaným slovníkem pro vědecké názvy byl Integrated Taxonomic Information System a pro předmětová hesla Library of Congress Subject Headings. Přibližně čtvrtina metadatových výroků byla vložena stejně jak profesionály, tak vědci – rozdíly v tvorbě metadat mezi těmito dvěma skupinami jsou tedy velké. Autorka navrhuje, aby pro multioborové repozitáře byla používána metadata profesionálů a pro oborové repozitáře metadata vědců. (White, 2012)

Stewart se ve svém článku věnuje předmětovému přístupu k obrazovým zdrojům, konkrétně se zaměřuje na historické fotografie, u kterých je správná indexace základem. První část článku komentuje výsledky studie, jejímž účelem bylo zjistit, jak v současnosti profesionálové indexují historické fotografie. Autor při tom bral v potaz Shatford/Panofsky matici pro klasifikaci předmětové úrovně, která obsahuje čtyři základní otázky, které by měla indexace zodpovědět (kdo, co, kde a kdy) a na každou z nich je pohlíženo ze tří pohledů podle předmětové úrovně (specifika, obecnosti, abstrakce). Bylo zjištěno, že profesionálové používají spíše tradiční přístup založený na textu a indexují pouze omezený počet předmětových úrovní. Nejméně indexovanou úrovní je úroveň abstraktní. Autor se domnívá, že dříve byl tento přístup dostačující, ale protože jsou dnes sbírky dostupné online, jsou uživatelé závislí pouze na správné indexaci historických fotografií, která je však podle něj zatím nedostatečná. V druhé části autor porovnává profesionální indexaci s tagováním uživatelů. Bylo zjištěno, že tagování reprezentuje tři čtvrtiny zmíněné matice a že uživatelé tagují i vyšší úrovně abstrakce než profesionálové. Výzkum také říká, že pokud uživatelé projdou fází tréninkové studie, jejich reprezentace se ještě rozšíří na celou matici. Autor tvrdí, že je možné využít spolupráce profesionálů a uživatelů k vylepšení kvality obsahových metadat historických fotografií. (Stewart, 2015)

Stanwicksův výzkum byl zaměřen na tagování písní a mapování tagů na entity FRBR. Byly náhodně vybrány nejlépe hodnocené písně z hitparády Billboard Hot 100 z rozsahu let 1958 až 2013 a extrahovány související tagy z webu last.fm.

Následně byl analyzován vztah tagů k entitám a atributům FRBR. Bylo zjištěno, že většina tagů koresponduje s atributy FRBR a že jsou rozloženy do všech tří skupin entit. Nejčastějšími tagy byly předmět, název a tvůrce, pomocí předmětových tagů uživatelé nejčastěji vyjadřují aboutness písně, žánr a související témata. Tagy odpovídaly téměř všem aplikovatelným atributům, nebyly nalezeny tagy pouze pro čtyři atributy: rychlost přehrávání, konfigurace pásku, způsob ozvučení a speciální přehrávací charakteristika. Metadata vzniklá tagováním byla v obecné rovině následně srovnána s metadaty v knihovních katalozích. Knihovny popisují agregační díla, např. celá alba, a mohou také v poznámce uvádět názvy jednotlivých písní, ale neposkytují úroveň popisu pro jednotlivé skladby. Uživatelé tagování umožňuje hlubší úroveň popisu a výhodou je také kombinace užších a širších předmětových a žánrových pojmů, které uživatelé vkládají. (Stanwicks, 2017)

Linová se ve své práci soustředila na porovnání profesionálních metadat a tagů z hlediska jejich použití uživateli při vyhledávání. Autorka zjišťovala, zda uživatelé naleznou požadované obrazové zdroje účinněji a efektivněji za pomoci předmětových hesel nebo tagů, k čemuž bylo využito speciální uživatelské rozhraní, které uchovávalo informační stopu vyhledávání. Bylo empiricky ověřeno, že nejlepšími výsledky uživatelé dosahovali, pokud k vyhledávání používali předmětová hesla v kombinaci s tagy. Výsledky studie mohou být použity k designu uživatelského rozhraní, a přestože je výzkum zaměřen pouze na obrazové zdroje, autorka uvádí, že se uvedené výsledky dají aplikovat na ostatní typy zdrojů. Uživatelé tagování může být využito jako ekonomicky efektivní obohacení profesionálního popisu. (Lin, 2013)

Článek Manzoové je zaměřen na podobné téma. Navazuje na předchozí studie, které zkoumaly přesnost tagování a jejich použitelnost pro vyhledávání. V této studii byly porovnávány dva způsoby vyhledávání: pouze metadata z řízeného slovníku a metadata z řízeného slovníku v kombinaci s uživatelskými tagy. Bylo zjištěno, že pomocí kombinované formy vyhledávali účastníci výzkumu průměrně šestkrát rychleji a celkem úspěšně získali více položek než účastníci, kteří vyhledávali pouze pomocí metadat z řízeného slovníku. Následně účastníci výzkumu sami opatřovali objekty tagy a bylo zjištěno, že 88 % takto získaných tagů bylo posouzeno jako přesných. Zároveň se míra přesnosti výrazně nelišila mezi

profesionálními knihovníky a laiky, což ilustruje možnost použití folksonomií pro zlepšení vyhledatelnosti dokumentů. (Manzo, 2015)

4. Cíle analýzy a výzkumné otázky

Cílem analýzy bylo porovnat obsahová metadata, která jsou tvořena profesionály (v knihovních katalozích a katalozích knihkupectví) s obsahovými metadaty tvořenými uživateli (ve čtenářských portálech). Sebraná data jsou kvalitativní sondou do této problematiky, ale zároveň byla sebrána v takovém množství, aby se dala kvantitativně vyhodnotit (aby byl omezen vliv náhody) a tím alespoň naznačit obecné zákonitosti.

Autorka si pro tuto práci stanovila několik výzkumných otázek:

1. Které typy obsahových metadatových výroků existují? Liší se to mezi profesionálními a uživatelskými metadaty?
2. Kolik metadatových výroků kterého typu databáze obsahují?
3. Z jaké části jsou uživatelská metadata objektivní? Liší se to mezi jejich formami?
4. Z jaké části se profesionální a uživatelská metadata shodují?
5. Existují uživatelská metadata, která jsou potenciálně užitečná a mohla by doplnit profesionální metadata?
6. Kdo při popisu informačních zdrojů tvoří složitější koncepty, profesionálové nebo uživatelé?
7. Jaká je míra sémantické interoperability uživatelských a profesionálních metadat?

Autorka při zodpovídání těchto otázek vycházela z modelu IFLA Library Reference Model a definovaných funkčních požadavků (Riva, 2017). Výzkumné otázky se dotýkají některých funkčních požadavků, a to v těchto případech (české názvy Kučerová, 2018): (1) Požadavku najít se dotýká otázka 3. Z části se zabývá metadatovými výroky ve formě poliček, které pomáhají uživatelům nacházet související informační zdroje. (2) Požadavku určit se dotýká více oblastí tohoto výzkumu, zejména ty, které se zabývají interpretací metadat. Otázka 1 definuje typy metadatových výroků, z toho u metadatových výroků, které reprezentují řízené slovníky, může dojít k nejasné interpretaci ze strany uživatele. Zároveň u uživatelských metadatových výroků může dojít ke zkreslení informace v komunikaci mezi uživatelem, který metadatový výrok vložil, a uživatelem, který

jej čte. (3) Požadavku vybrat se nedotýká přímo výzkumná otázka, ale byl zohledněn už při sběru dat. Jak je zmíněno v kapitole 2.4, u uživatelských metadat se dá relevance rozlišovat v rámci širokých folksonomií. Pro výzkum byla vybrána pouze relevantní uživatelská metadata, konkrétně je to definováno v kapitole 5.5.2. (4) Požadavek získat nebudeme v rámci tohoto výzkumu sledovat. (5) Požadavku prozkoumat se dotýká zejména otázka 1, v jejímž rámci jsou popsány některé typy metadat, které pomáhají nalézt související informační zdroje. Jedná se především o související dílo a související film. Tomuto požadavku bezesporu napomáhají také hypertextové odkazy uživatelských metadat, ale tyto související informační zdroje nejsou v této práci zkoumány.

5. Metodika analýzy

V této kapitole budou co nejpřesněji popsány jednotlivé části výzkumu, aby bylo možné výzkum v případě zájmu zopakovat.

5.1 Definice důležitých pojmů

5.1.1 Databáze

Databáze je systém, který slouží k modelování objektů reálného světa a jejich vztahů. Základními prvky jsou obsah databáze, tedy množina dat, která musí být jednotně strukturovaná a uložena tak, aby se s daty dalo snadno manipulovat, a program pro přístup k datům. (Kučerová, 2003)

Pro potřeby tohoto výzkumu však budeme pojem databáze používat v mírně upraveném významu. Databázi definujeme jako systém, který slouží jako zdroj metadatových výroků. Byly zváženy i další termíny, které by se mohly v tomto významu využívat, např. dataset nebo metadatový repozitář, ale nakonec bylo z důvodu pohodlnějšího psaní a porozumění rozhodnuto právě pro pojem databáze. V rámci výzkumu rozlišujeme tři typy databází:

(1) Knihovní katalog je databáze záznamů dokumentů, které daná knihovna uchovává ve svém fondu a/nebo zpřístupňuje. Jednotlivé záznamy jsou vytvářeny podle předem stanovených pravidel. (Balíková, 2003) Tato pravidla jsou z určité části pro všechny zúčastněné knihovny stejná, protože jsou stanovena na celosvětové a české knihovnické úrovni. V této práci se zabýváme pouze katalogy elektronickými a jejich prezentací ve webovém prostředí.

(2) Katalog knihkupectví je také databáze záznamů dokumentů, které dané knihkupectví má v nabídce k prodeji. Jednotlivé záznamy jsou i zde vytvářeny podle předem daných pravidel, nicméně od knihovního katalogu se liší tím, že nad jednotlivými knihkupectvími nestojí žádná vyšší autorita, která by určovala jednotná pravidla pro katalogy, tudíž má každé knihkupectví vlastní pravidla pro tvorbu katalogových záznamů. Také má vlastní kategorie, do kterých jednotlivé dokumenty řadí, jak uvidíme později v průběhu výzkumu.

Oba tyto typy databáze mají společné to, že jejich tvůrci dělají tuto činnost profesionálně, tj. jsou za ni placeni.

(3) Třetím typem databáze je čtenářský portál. Jelikož na rozdíl od dvou předchozích typů není tento termín v knihovnictví běžně používán, věnujeme mu větší pozornost. Pod pojmem čtenářský portál v rámci tohoto výzkumu rozumíme databázi záznamů dokumentů (převážně knih), která je tvořena čtenáři pro čtenáře. Tato databáze je na rozdíl od předchozích dvou typů vytvářena uživateli v jejich volném čase, nejsou za tuto činnost placeni. Většinou se jedná o laiky, tzn. uživatele, kteří nemají vzdělání v knihovnickém ani informačním oboru. Jejím hlavním cílem je poskytovat uživatelům možnost:

- komentovat, organizovat dokumenty v databázi a reagovat na ně prostřednictvím různých funkcí, např. tagů nebo políček,
- vybírat si knihy k četbě,
- diskutovat a sdílet informace a názory s ostatními uživateli,
- účastnit se soutěží nebo výzev a další.

Vedlejším cílem čtenářských portálů může být vytvořit úplný výčet dokumentů určitého typu, např. knih vydaných v češtině, nezávisle na tom, jak se dají získat. K přispívání na čtenářský portál je potřeba vytvořit si účet s vlastními přihlašovacími údaji, u všech změn je pak zpravidla veřejně zpřístupněné uživatelské jméno uživatele, který změnu provedl. Také mohou být určeni uživatelé s většími právy, kteří buď schvalují nebo kontrolují příspěvky ostatních. Díky tomu je zajištěna ochrana před úmyslným (nebo i neúmyslným) přidáváním chybných informací a spamu nebo dalšími nežádoucími vlivy.

5.1.2 Thema

Thema je schéma pro předmětovou kategorizaci pro mezinárodní použití v rámci knižního trhu. Bylo vytvořeno pro použití ve všech sektorech globálního knižního trhu, tedy pro vydavatelství, velkoobchod i všechna knihkupectví, a pro veřejné knihovny. Schéma je spravováno společností EDItEUR a nejnovější verze Thema 1.3 byla vydána v dubnu 2018. (Thema v1.4, 2009)

Cílem pro vytvoření tohoto standardu bylo získat nástroj, který propojuje jednotlivé jazyky a kulturní oblasti, a tím zabránit duplicitní práci při využívání více schémat nebo složitému procesu mapování jednoho schématu na druhé. (Saynor, 2019) Proto byl také vybrán pro účely této práce. V našem výzkumu bude Thema sloužit ke srovnávací analýze profesionálních a uživatelských obsahových metadat jakožto nástroj, na který se budou jednotlivé metadatové výroky, resp. jejich hodnoty, mapovat.

5.1.3 SKOS

S tím souvisí i termín SKOS, tedy Simple Knowledge Organization System. Jedná se o model pro vyjadřování struktury a obsahu pojmových schémat, umožňuje jejich publikování na webu, propojování s dalšími daty a integraci do jiných pojmových schémat (Isaac, 2009).

Nás v této práci zajímají tzv. mapovací vlastnosti, které jsou definované ve SKOS Simple Knowledge Organization System Reference (v originále mapping properties), které budou využity pro mapování obsahových metadatových výroků na Thema. Jsou zde definovány tyto vlastnosti:

- skos:closeMatch,
- skos:exactMatch,
- skos:broadMatch,
- skos:narrowMatch,
- skos:relatedMatch (Miles, 2009).

Bylo rozhodnuto, že pro účely této práce postačí vybrat pouze dvě mapovací vlastnosti z toho důvodu, že by jinak došlo k přílišnému rozmělnění výsledků a ke snížení jejich výpovědní hodnoty. Užity budou skos:closeMatch (dále jen closeMatch) a skos:exactMatch (dále jen exactMatch), avšak ne nutně vždy ve svých přesných významech, jak jsou definovány v SKOS Simple Knowledge Organization System Reference. Ve výzkumu nám jde o rozlišení toho, zda analyzované obsahové metadatové výroky odpovídají třídíku Thema přesně, nebo zda byl pro mapování zvolen pojem dostatečně blízký, ale nikoli zcela přesně odpovídající. Zde následuje naše definice těchto dvou mapovacích vlastností:

- exactMatch: Tato vlastnost byla zvolena, pokud analyzované obsahové metadatové výroky mají zcela shodnou podstatu (ideu) jako odpovídající třídění Thema. Není při tom rozhodující, zda odpovídají řetězce znaků, kterými jsou metadatové výroky a třídění vyjádřeny.
- closeMatch: Tato vlastnost byla zvolena, pokud analyzované obsahové metadatové výroky odpovídají třídění Thema, ovšem nikoli zcela přesně, nebo pokud je jejich vztah pro tento výzkum jinak důležitý.

Této problematice se ještě budeme dále věnovat v kapitole 5.6.4.

5.1.4 Profesionální a uživatelská metadata

Jako profesionální metadata označujeme taková metadata, která jsou vytvořena profesionálními tvůrci, tedy lidmi, kteří jsou za tuto činnost placeni. Předpokládá se, že mají předem jasně daná pravidla, jak při tvorbě metadat postupovat, a/nebo jsou před začátkem činnosti vyškoleni. V rámci této práce termínem profesionální metadata označujeme metadata nacházející se v knihovních katalozích a v katalozích knihkupectví.

Uživatelská metadata jsou tvořena uživateli, kteří nejsou za tuto činnost placeni, nemají žádná nebo velmi volná pravidla tvorby metadat, a zároveň nedostali žádné oficiální školení. Do této kategorie řadíme metadata v čtenářských portálech.

5.1.5 Objektivní a subjektivní metadatové výroky

Pro potřeby této práce vytvoříme ještě jedno členění metadat, a to na objektivní a subjektivní. Subjektivní metadatový výrok definujeme jako takový metadatový výrok, kde se alespoň jedna jeho část vztahuje k osobě, která obsahová metadata vytváří, případně k její afilované instituci. Typickým příkladem subjektivního metadatového výroku je označení „moje oblíbené detektivky“. Zde vidíme, že část „detektivky“ je objektivní, vztahuje se ke knize, ale část „moje oblíbené“ se vztahuje k člověku, který knihu takto označil. Jiný člověk by stejnou knihu mohl snadno označit např. jako „nedočetl jsem“.

Za objektivní metadatový výrok považujeme takový metadatový výrok, který se vztahuje pouze k vlastnostem daného dokumentu nebo přidružených objektů (např. série, do které dokument patří) či subjektů (např. autor). Může být složen

z více částí, ale žádná se nesmí vztahovat k osobě vytvářející metadata, ani k její instituci. Objektivní metadatové výroky u stejného dokumentu platí vždy, nejsou závislé na dalších podmínkách (např. formát, ve kterém je dokument k dispozici). Příkladem tohoto typu metadatového výroku tak může být i označení „detektivky vydané v roce 2016“ nebo „série Harry Hole“, jelikož se vztahují ke knize a budou platit vždy nezávisle na popisující osobě.

Z této definice je jasné, že za objektivní metadatové výroky považujeme i všechny metadatové výroky týkající se tématu knihy, např. „intriky“ nebo „nehoda“. Ve výjimečných případech může nastat situace, že se objektivita/subjektivita metadatového výroku nedá zcela jasně určit podle výše řečených pravidel, pak záleží na úsudku hodnotitele.

5.2 Výběr knih

Cílem bylo vybrat beletristické knihy, které jsou známé širokému publiku, aby byl zajištěn dostatek dat pro analýzu. Knihy také měly být různorodé, mít odlišné žánry a témata, a tím pádem i odlišnou čtenářskou základnu, aby byla zajištěna i dostatečná variabilita dat. Pro výběr knih byly určeny tři zdroje, z nichž byly vybrány nejprodávanější knihy v letech 2015-2017:

(1) Z žebříčku Publishers Weekly byly vybrány všechny knihy s počtem prodaných výtisků vyšším než 1 milion, nebo první tři knihy v kategorii adult fiction ze tří článků popisujících roky 2015-2017 (Bestselling Books of 2015, 2016; The Bestselling Books of 2016, 2017; The Bestsellers of 2017, 2018).

(2) Z žebříčku The New York Times byly vybrány knihy, které se alespoň dvakrát umístily v týdenním žebříčku New York Times Bestseller v kategorii fiction v daném roce (a to i v případě, že jedno umístění bylo v jednom roce a druhé umístění v roce druhém). Z důvodu usnadnění získání těchto dat nebyla jako zdroj využita stránka www.nytimes.com, ale články z Wikipedie, kde jsou umístěné knihy přehledně uvedeny (The New York Times Fiction Best Sellers of 2015, 2019; The New York Times Fiction Best Sellers of 2016, 2018; The New York Times Fiction Best Sellers of 2017, 2020).

(3) Z žebříčku Český Bestseller byly vybrány všechny knihy v kategorii překladová literatura pro dospělé z článků popisujících rok 2015-2017 (Historie za rok 2015, 2015; Historie za rok 2016, 2016; Historie za rok 2017, 2017).

Každá kniha mohla být vybrána pouze jednou. Zároveň knihy musely být vydány v českém i v anglickém jazyce nejpozději v roce 2018, tato hranice byla stanovena z toho důvodu, aby byl poskytnut dostatečný čas pro vytvoření všech potřebných obsahových metadat v databázích. Zejména jde o čtenářské portály, u kterých se předpokládá největší nárůst počtu uživatelských obsahových metadat v prvním roce po publikaci knihy v daném jazykovém prostředí.

Níže se nachází seznam dvaceti sedmi knih, které byly do analýzy zařazeny. Číslo knihy je číslo, které bylo dané knize přiřazeno pro použití v tomto výzkumu, nemá jiný význam. Rok vydání česky byl převzat ze Souborného katalogu ČR⁸, rok vydání anglicky byl převzat z katalogu WorldCat⁹. Při tom bylo podle potřeby užito těchto funkcí katalogů: vyhledávání frází pomocí uvozovek, vyhledávání podle názvu, výběr formátu (tištěná kniha), filtrování podle autora, filtrování podle jazyka, řazení podle data od nejstaršího a zobrazení všech vydání a formátů dané knihy. V případě nejasností byl vybrán záznam s textem „1. vydání“, „první vydání“ nebo „first edition“.

Tabulka 1: Seznam analyzovaných knih

Číslo knihy	Autor	Název knihy česky	Rok vydání česky	Název knihy anglicky	Rok vydání anglicky
1	Lee, Harper	Postav hlídku	2016	Go Set a Watchman	2015
2	James, E. L.	Grey: padesát odstínů šedi pohledem Christiana Greye	2016	Grey	2015

⁸

https://aleph.nkp.cz/F/4Q46JAC6CEMRS53P42S8TP2F6MVHENBYKCPBHTUEL1BH2KR58E-38723?func=file&file_name=find-b&local_base=SKC

⁹ <https://www.worldcat.org/>

Číslo knihy	Autor	Název knihy česky	Rok vydání česky	Název knihy anglicky	Rok vydání anglicky
3	Hawkins, Paula	Dívka ve vlaku	2015	The Girl on the Train	2015
4	Doerr, Anthony	Jsou světla, která nevidíme	2015	All the Light We Cannot See	2014
5	Weir, Andy	Mart'an	2015	The Martian	2014
6	Grisham, John	Osamělý střelec	2016	Rogue Lawyer	2015
7	Lagercrantz, David	Dívka v pavoučí síti	2015	The Girl in the Spider's Web	2015
8	Nesbo, Jo	Policie	2015	The Police	2013
9	Nesbo, Jo	Syn	2014	The Son	2014
10	Galbraith, Robert	Hedvábník	2015	The Silkworm	2014
11	Backman, Fredrik	Muž jménem Ove	2014	A Man Called Ove	2014
12	Grisham, John	Píšťalka	2017	The Whistler	2016
13	Moyes, Jojo	Než jsem tě poznala	2013	Me Before You	2012
14	Moyes, Jojo	Život po tobě	2016	After You	2015
15	Jonasson, Jonas	Zabiják Anders a jeho přátelé (a sem tam nepřítel)	2016	Hitman Anders and the Meaning of It All	2016
16	Galbraith, Robert	Ve službách zla	2016	Career of Evil	2015
17	Towles, Amor	Gentleman v Moskvě	2018	A Gentleman in Moscow	2016
18	Wingate, Lisa	Než jsme byly tvoje	2017	Before We Were Yours	2017

Číslo knihy	Autor	Název knihy česky	Rok vydání česky	Název knihy anglicky	Rok vydání anglicky
19	Connelly, Michael	Odvrácená strana konce	2017	The Wrong Side of Goodbye	2016
20	Cameron, W. Bruce	Psí poslání	2013	A Dog's Purpose	2010
21	Young, William P.	Chatrč	2009	The Shack	2007
22	Hawkins, Paula	Do vody	2017	Into the Water	2017
23	Grisham, John	Obchodník	2018	Camino Island	2017
24	Brown, Dan	Počátek	2018	Origin	2017
25	Cole, Daniel	Hadrový panák	2017	Ragdoll	2017
26	Lagercrantz, David	Muž, který hledal svůj stín	2017	The Girl Who Takes an Eye for an Eye	2017
27	Nesbo, Jo	Žízeň	2017	The Thirst	2017

5.3 Výběr databází

Jak již bylo řečeno výše, v tomto výzkumu se zabýváme třemi typy databází: knihovními katalogy, katalogy knihkupectví a čtenářskými portály. Zároveň byly z každého typu vybrány databáze ve dvou jazykových variantách: v češtině a angličtině. Důležité je upozornit, že nás zajímají pouze databáze institucí ze zemí anglicky mluvícího světa, konkrétně bylo toto kritérium vymezeno na těchto šest zemí:

- Spojené státy americké,
- Velká Británie,
- Kanada,
- Austrálie,
- Nový Zéland,

- Irsko.

Důvodem pro toto omezení je, že obsah (význam) analyzovaných metadatových výroků je spjatý s jazykem, ve kterém metadata jsou. V jiných jazycích než v češtině a angličtině by nebylo možné zaručit správnost interpretace metadatových výroků. Namátkou bylo zjištěno, že nemohou být použity ani katalogy institucí z jiných než anglicky mluvících zemí, které mají uživatelské rozhraní v angličtině, protože metadatové výroky, které obsahují, mohou být (a často také jsou) v národním jazyce instituce, nikoli v angličtině.

Vybrané databáze musí dále splňovat tato kritéria:

- musí se v nich nacházet alespoň 55 % analyzovaných knih;
- nesmí se jednat o souborné katalogy;
- musí mít na své webové prezentaci dostupný katalog online;
- v katalogu musí být v záznamech jednotlivých knih dostupná strukturovaná obsahová metadata (jaké metadatové výroky byly sbírány, bude podrobněji definováno v kapitole 5.5.2).

Výběr čtenářských portálů nepotřeboval vlastní metodiku, pro českou i anglickou jazykovou oblast byly nalezeny právě dvě databáze, které splňovaly definici čtenářského portálu.

Seznam sedmdesáti dvou databází, ze kterých bylo čerpáno pro sběr dat, se nachází v příloze 1. Číslo databáze je číslo, které bylo dané databázi přiřazeno pro použití v tomto výzkumu, nemá jiný význam. Z důvodu přehlednosti jsou databáze rozděleny podle jejich typů (viz kapitola 5.1.1) a podle jazyka. Dále uvádíme postup výběru databází pro zbylé čtyři skupiny, které vznikly kombinací typu databáze a jazykové oblasti.

5.3.1 Knihovní katalogy české

V tomto případě jsme postupovali podle systému knihoven, jak je uveden v knihovním zákoně z roku 2001:

„(1) Systém knihoven tvoří

a) Národní knihovna České republiky (dále jen "Národní knihovna"), Knihovna a tiskárna pro nevidomé K. E. Macana, Moravská zemská knihovna v Brně, zřízené Ministerstvem kultury (dále jen "ministerstvo"),

b) krajské knihovny, zřízené příslušným orgánem kraje,

c) základní knihovny, zřízené příslušným orgánem obce,

d) specializované knihovny.“ (Česko, 2001)

Knihovny byly vybrány podle této metodiky:

- všechny knihovny zřízené Ministerstvem kultury;
- všechny krajské knihovny podle adresáře na webových stránkách Souborného katalogu České republiky (Adresáře knihoven v ČR, 2016);
- k tomu byla přidána Městská knihovna Praha, jelikož také obsluhuje kraj;
- ze základních knihoven byly vybrány ty, které vyhrály soutěž Městská knihovna roku mezi lety 2010-2019, kterou vyhlašuje Svaz knihovníků a informačních pracovníků (Městská knihovna roku, 2010);
- specializované knihovny byly pominuty, jelikož bylo namátkově ověřeno, že nemají ve svém fondu analyzované knihy.

5.3.2 Katalogy knihkupectví české

Cílem bylo vybrat ta největší a nejznámější česká knihkupectví, která splňují výše zmíněné požadavky. Jelikož pro knihkupectví neexistují žádné žebříčky ani soutěže, byl pro výběr stanoven následující postup. V české mutaci internetového vyhledávače Google byl vyhledán dotaz „knihkupectví“, následně se zobrazil seznam několika knihkupectví, zde byla vybrána položka „Další místa“. Zde pak ze seznamu byly sečteny počty recenzí jednotlivých knihkupectví (vyhledávač Google odděluje jednotlivé pobočky) a vybrána ta knihkupectví, která měla v součtu nad 80 recenzí. Hodnoceno bylo pouze prvních 10 stran výsledků vyhledávání. Kromě toho, že všechna vybraná knihkupectví musela splňovat podmínky uvedené v kapitole 5.3, tak z výběru byla vyřazena knihkupectví, která:

- vůbec nebyla knihkupectvími;

- měla pro analyzované knihy jedinou kategorii, a to „beletrie“¹⁰.

Knihkupectví, která byla z výběru vyřazena, jsou uvedena i s důvody jejich vyřazení v příloze 2. Zároveň je tam zdokumentován počet recenzí jednotlivých knihkupectví, které byly na Googlu v době výběru databází (listopad 2019).

5.3.3 Knihovní katalogy anglicky mluvících zemí

I zde bylo cílem vybrat největší a nejznámější knihovny anglicky mluvícího světa. Vybrány byly knihovny měst s více než 1 milionem obyvatel ze států uvedených v kapitole 5.3. Ta byla určena na základě dat dostupných v internetové encyklopedii Wikipedia dle jednotlivých zemí:

- List of United States cities by population (2020), česky Seznam měst ve Spojených státech podle počtu obyvatel;
- List of cities in the United Kingdom (2020), česky Seznam měst ve Spojeném království;
- List of largest Canadian cities by census (2020), kapitola 2016, česky Seznam největších kanadských měst podle sčítání lidu;
- List of cities in Australia by population (2020), česky Seznam měst v Austrálii podle počtu obyvatel;
- List of cities in New Zealand (2020), česky Seznam měst na Novém Zélandu;
- List of urban areas in the Republic of Ireland by population (2020), česky Seznam městských území v Irské republice podle počtu obyvatel.

Upřednostňovány byly „public libraries“ nacházející se v daných městech. Toto rozdělení zajistilo, že výběr respektoval rozložení anglicky mluvícího obyvatelstva po celém světě. Zároveň byly připojeny nejoblíbenější knihovny podle článku „The 12 Most Popular Libraries in the World“ (Temple, 2018a), které odpovídaly požadavkům stanoveným v kapitole 5.3 (tzn. opět se musely nacházet v jedné ze

¹⁰ Kategorie knih je nejčastějším metadatovým elementem v katalozích knihkupectví. Vzhledem k tomu, že analyzujeme pouze knihy z beletrie, dá se kategorie „beletrie“ předpokládat, a tudíž nepřináší pro výzkum žádnou novou informaci.

šesti anglicky mluvících zemí). Každá knihovna mohla být do výběru zařazena pouze jednou.

5.3.4 Katalogy knihkupectví anglicky mluvících zemí

Vzhledem k neexistenci oficiálních globálních informací o knihkupectvích (alespoň ve formě volně dostupných informací na webu) byly databáze z této kategorie vybrány z kombinace více článků nalezených na webu. Všechny jsou dostatečně relevantní i důvěryhodné, mají uvedeného autora i datum publikování, žádný z nich není starší deseti let. Jedná se o tyto články:

- 12 Of The Biggest Bookshops In The World For When You Want To Lose Yourself In Literature, česky 12 z největších knihkupectví světa, když se chcete ztratit v literatuře (Oulton, 2017);
- The world's greatest bookshops (2012), česky Největší světová knihkupectví;
- The 10 Most Famous Bookstores in the World, česky 10 nejznámějších knihkupectví světa (Temple, 2018b);
- 18 Bookstores Every Book Lover Must Visit At Least Once, česky 18 knihkupectví, která musí každý milovník knih alespoň jednou navštívit (Lutz, 2014).

Vybrány byly opět pouze ty databáze, které splňovaly požadavky stanovené v kapitole 5.3, vyřazené databáze najdete i s odůvodněním v příloze 2.

5.4 Výběr analyzovaných metadatových elementů

Pro tento výzkum nás zajímají obsahová metadata podle definice v kapitole 2.2, která jsou strukturovaná, tzn. jsou zapsaná ve formě polí, a/nebo nemají charakter volně psaného textu. V záznamech jednotlivých knih v rámci jedné databáze musí být stejné metadatové výroky vždy označeny stejně a musí být zřejmé, že se jedná o stejný metadatový element jako v ostatních záznamech. Výše uvedené platí pro užitelská metadata pouze v omezené míře, nakolik to jejich povaha umožňuje.

Před začátkem sběru dat byly v analyzovaných databázích identifikovány a sepsány typy metadatových výroků, které splňují výše zmíněné požadavky a nesou určitou informační hodnotu. Pro jednotlivé typy byly vytvořeny charakteristické názvy

a bylo zaznamenáno, ve které databázi se který typ metadatového výroku ke sběru nachází a ve které nikoliv. To proběhlo pomocí tabulky, kde názvy databází označovaly řádky, typy metadatových výroků sloupce a jednotlivé buňky tabulky byly vyplněny čísly 0 a 1, kde 0 znamená „databáze neobsahuje tento typ metadatového výroku“ a 1 znamená „databáze obsahuje tento typ metadatového výroku“. Cílem této přehledové tabulky bylo určit, ve kterých databázích se metadatový výrok daného typu vyskytuje alespoň jednou, tedy která databáze má schopnost ho zaznamenat a zpřístupnit. Není nutné, aby se daný typ metadatových výroků nacházel ve všech záznamech analyzovaných knih. Pro vytvoření tabulky byl použit nástroj Microsoft Excel, tabulka se nachází v příloze 3. Oproti kapitole 5.5.2 jsou v této tabulce uživatelská metadata reprezentována typy dle jejich formy (tagy a políčky), jelikož sloužila pouze ke sběru metadatových výroků.

5.5 Sběr dat

Pro potřeby této práce byly využity obsahové metadatové výroky sesbírány autorkou ze všech tří typů databází (viz kapitola 5.1.1). Data byla získána pouze elektronickou formou prostřednictvím webových prezentací databází. Při tom byly využity různé možnosti těchto prezentací a také bylo přihlédnuto k uživatelskému určení jednotlivých typů databází i jednotlivých typů zobrazení v rámci jedné databáze.

Sběr dat probíhal v období od 23. 12. 2019 do 15. 2. 2020.

Data byla zapisována do tabulky vytvořené pomocí nástroje Microsoft Excel, a to tím způsobem, že analyzované typy metadatových výroků označovaly sloupce a na řádek byly jednotlivě zapisovány názvy databází a ke každé byly do buněk v příslušných sloupcích vypsány všechny sebrané metadatové výroky. Další název databáze následoval vždy až po vepsání všech příslušných metadatových výroků, počet řádků alokovaných pro jednotlivé databáze se tedy lišil vždy v závislosti na počtu relevantních metadatových výroků v databázi zobrazených. Pro přehlednější vizualizaci byla vždy mezi každé dvě databáze vložena vodorovná čára. Pro každou knihu byla vytvořena nová tabulka na samostatném listě, listy byly označeny pořadovými čísly knih. Tabulky se nachází v příloze 4.

5.5.1 Výběr záznamu a zobrazení

Pro výběr konkrétního záznamu, ze kterého proběhne sběr dat, platí následující pravidla:

- Musí jít o záznam stejné knihy, kterou hledáme. Neuvažujeme žádná shrnutí ani kvízy, které mají stejný název.
- Musí jít o záznam tištěné knihy. Nebereme v potaz e-knihy ani audioknihy.
- Kniha musí být v jazyce databáze.

Pokud nejsou tato pravidla splněna, postupujeme, jako kdyby se kniha v databázi nenacházela. Pokud se i po aplikaci těchto pravidel máme k dispozici více záznamů, pak preferujeme „normální“ formát (nikoli velkoformátové vydání) a první vydání knihy, ale pokud nejsou k dispozici, můžeme vybrat jakýkoliv jiný záznam, který odpovídá pravidlům. Na vazbě knihy (pevná nebo brožovaná) nezáleží. Pro katalogy knihkupectví anglicky mluvících zemí dále platí, že jsou sbírána data ze všech záznamů, které odpovídají pravidlům (k tomu bylo použito tlačítko „View other formats and editions“ nebo obdobné, pokud bylo k dispozici). Důvodem je vysoká rozmanitost dat mezi jednotlivými záznamy a není možné určit, který záznam je ten „správný“.

Při sběru dat bylo zjištěno, že databáze mohou mít různé formy zobrazení záznamu. V takovém případě bylo preferováno zobrazení pro knihovnické nebo informační profesionály (zpravidla zobrazení záznamu ve formátu MARC), pokud nebylo k dispozici, tak návěšťový lístek (týká se českých knihovnických katalogů) a případně jakékoli další zobrazení, zejména čtenářské.

5.5.2 Přehled typů metadat

Níže uvádíme přehled jednotlivých typů metadat, které byly sbírány. Nejprve je však k této problematice třeba přidat poznámku:

Následující typologie má dvě funkce: (1) První funkcí je porovnání profesionálních a uživatelských metadat, které je primárním cílem práce. Proto typologie metadat¹¹ vychází z metadatových elementů, které se v analyzovaných databázích skutečně nachází.¹² A již předem je zřejmé, že ve čtenářských portálech se budou nacházet jiné typy metadat než v profesionálních databázích. Z toho důvodu typologie obsahuje kromě typů, do kterých se dají zařadit obě skupiny metadat, také typy, kterým odpovídají pouze profesionální nebo pouze uživatelská metadata. U každého typu bude blíže vysvětleno, metadata ze které skupiny do něj mohou patřit. Je to jediný způsob, jak zajistit přesné porovnání profesionálních a uživatelských metadat.

(2) Druhou, doplňkovou, funkcí typologie je popis sloupců tabulek, které sloužily ke sběru metadatových výroků (a nachází se v příloze 4). Základním rysem uživatelských metadat v rámci tohoto výzkumu je fakt, že není možné odlišit metadata obsahová (nebo faktická) od jiných typů, bereme-li v úvahu typologii tagů (viz kapitola 2.4). Proto jsou do výzkumu zahrnuty vždy všechna uživatelská metadata, která splňují podmínky (viz níže). Na tomto místě je důležité zmínit, že uživatelské metadatové výroky byly řazeny ještě do dvou dalších sloupců, které v typologii neuvádíme, jelikož jde o čistě formální dělení založené na formě uživatelských metadat. Slouží pouze pro sběr metadatových výroků a jejich řazení do sloupců a nemá v rámci výzkumu žádný další význam. Jedná se o tyto dvě kategorie:

1. **Tagy:** Jejich zdrojem byla pole *Štítky* a *Tags*. Vzhledem k velkému množství metadat tohoto typu ve čtenářských portálech anglicky mluvících zemí bylo rozhodnuto, že v případě, že se v databázi nachází více než 20 tagů, budou sbírána pouze ta metadata, která přidalo 20 a více uživatelů. Na české čtenářské portály se toto pravidlo nedá aplikovat, jelikož neobsahují informaci o počtu uživatelů, v tomto případě byla vždy sebrána všechna metadata tohoto typu. Jak již bylo řečeno výše, zajímá nás pouze text tagu a jeho popisná funkce, nikoli funkce kontextová.

¹¹ Mluvíme zde obecně, proto používáme termín metadata a nikoli metadatové výroky nebo elementy. Metadatovými elementy rozumíme v tomto kontextu konkrétní pole v záznamu, níže uvedené typy jsou zobecněným typem zahrnujícím několik různých polí.

¹² Doladění typologie tedy následovalo až po sběru dat.

2. **Poličky:** Metadata typu poličky byla sbírána obdobně jako tagy s tím rozdílem, že je muselo přidat více než 100 uživatelů (v databázích anglicky mluvících zemí). Zdrojem pro tento typ metadat byla pole *Knih XY je v, See top shelves..., Lists* a v jednom případě nebylo pole pojmenováno, ale pouze graficky znázorněno.

Nyní už byly vysvětleny všechny důležité okolnosti a můžeme přistoupit k typologii jako takové. Metadatové typy jsou doplněny názvy polí, pod kterými byly v databázích nalezeny (jak již bylo řečeno, preferováno bylo zobrazení pro profesionály, i zde tedy na prvním místě uvádíme kódy polí MARC). Tím jsou zároveň specifikována pravidla sběru dat, v některých případech jsou však potřeba i další poznámky ke sběru. Metadata jsou uvedena v pořadí, ve kterém se nachází v tabulkách pro sběr (příloha 4). Kódy a názvy polí a podpolí jsou převzaty z materiálu *Katalogizace podle RDA ve formátu MARC 21- tištěné a elektronické monografie* (Lichtenbergová, 2016) a/nebo *MARC 21 Format for Bibliographic Data* (2019).

A) Metadata, která reprezentují řízené slovníky: Tato metadata jsou základem pro obsahový popis informačních zdrojů, z hlediska aboutness jsou tím nejdůležitějším, co profesionální bibliografické záznamy nabízí. Jsou zde zastoupeny nejpoužívanější řízené slovníky v českém i anglickém prostředí. Navíc jsme zde zařadili metadatový typ klíčová slova, který reprezentuje obecnou formu obsahových metadat. Do těchto kategorií patří vždy pouze profesionální metadata.

1. **Konspekt:** Základním místem pro sběr metadat typu Konspekt bylo pole 072 ve formátu MARC, tedy *Kód předmětové kategorie*, dalším místem pak *Slovní vyjádření skupiny Konspektu* v návěšt'ovém lístku. V případě metadat tohoto typu bylo vždy sbíráno pouze slovní vyjádření, nikoli notace: zajímá nás tedy konkrétně podpole \$x, *slovní označení skupiny Konspektu*.
2. **MDT:** Základním místem pro sběr metadat typu MDT bylo pole 080 ve formátu MARC, tedy *Mezinárodní desetinné třídění (MDT)*, podpole \$a *Klasifikační znak MDT*, dalším místem pak *Třídění MDT* v návěšt'ovém lístku. U tohoto typu metadat se v záznamu zpravidla nachází pouze notace, slovní vyjádření bylo

přebíráno z portálu České MDT Online (2020). Poznámka: Notace (0:82-) znamená „ve formě...“, ale pro zjednodušení zápisu zapisujeme pouze formu jako samostatný termín.

3. **Předmět LCSH:** Základním místem pro sběr metadat typu předmět LCSH bylo pole 650 formátu MARC, tedy *Subject Added Entry-Topical Term*, pokud byl druhý indikátor 0, což značí použití Library of Congress Subject Headings. Dále mohlo jít o pole 610, tedy *Subject Added Entry – Corporate Name*, taktéž za podmínky, že druhý indikátor byl 0, a některé termíny z pole *Subject*, jak bude specifikováno v kapitole 5.5.3.
4. **DDT:** Základním místem pro sběr metadat typu DDT bylo pole 082 ve formátu MARC, tedy *Dewey Decimal Classification Number*, dalším místem pak pole *Dewey* nebo *Canonical DDC/MDS*. U tohoto typu metadat se v záznamu zpravidla nachází pouze notace, slovní vyjádření bylo přebíráno z 22. vydání publikace Dewey decimal classification and relative index, části 1 (Dewey, 2003a), 2 (Dewey, 2003b) a 3 (Dewey, 2003c) podle potřeby. Ve většině případů bylo v databázích deklarováno užití 23. vydání DDT, ale 22. vydání je nejnovějším vydáním volně dostupným z Internet Archive, proto bylo rozhodnuto o jeho užití pro dekódování notací. V některých případech bylo v databázi deklarováno jiné vydání DDT, seznam naleznete v příloze 5. U tohoto typu metadat je výjimka a může být zobrazen ve čtenářském portálu (konkrétně to umožňuje databáze LibraryThing), pak je s ním nakládáno jako s uživatelskými metadaty.
5. **Předmět FAST:** Jediným místem pro sběr metadat typu předmět FAST bylo pole 650 ve formátu MARC (*Subject Added Entry-Topical Term*) za předpokladu, že druhý indikátor byl 7 a v podpoli \$2 (*Source of heading or term*) bylo uvedeno „fast“.
6. **Klíčová slova:** Primárním zdrojem metadat typu klíčová slova bylo pole 653 ve formátu MARC, tedy *Index Term-Uncontrolled*, podpole \$a *Uncontrolled term*. V dalších případech také mohl být

tento typ metadat nalezen v poli *Uživatelská klíčová slova*¹³, *Štítky*¹⁴ nebo také *Kategorie* v případě, kdy kniha nebyla zařazena do beletrie (těchto případů nastal velmi omezený počet).

B) Metadata, která popisují základní entity obsahu knih: Tato metadata jsou také hodnocena jako velmi důležitá, a to jak z hlediska jejich výskytu v databázích, tak z hlediska obsahu informačního zdroje. Nachází se zde zejména metadata, která odpovídají na základní otázky: co, kdo, jak, kde, kdy a kam je to zařazeno. Pokud není stanoveno jinak, mohou do těchto typů být zařazena jak metadata profesionální, tak uživatelská.

7. **Téma:** Hlavním zdrojem metadat typu téma byla pole 650 a 610 ve formátu MARC (tedy *Subject Added Entry-Topical Term* a *Subject Added Entry – Corporate Name*) v případě, že metadata nesplňovala požadavky pro zařazení do metadat typu LCSH a/nebo předmět FAST. Dalšími zdroji mohla být pole *Téma (záhlaví)* v návštěvovém lístku, *Topics*, *Topical term*, *Theme*, *Subject Keywords*, *General Subjects* nebo některé termíny z pole *Subject*, jak bude specifikováno v kapitole 5.5.3.
8. **Postavy:** Primárním zdrojem pro metadata typu postavy byla pole 600 a 650 ve formátu MARC (tedy *Subject Added Entry – Personal Name* a *Subject Added Entry-Topical Term*) v případě, že se v poli vyskytoval termín „Fictitious character“. Toto pravidlo je nadřazené pravidlům v bodech 3 a 5. Termín „Fictitious character“ byl ze zápisu vynecháván, sloužil pouze k identifikaci metadat typu postavy. Dalším zdrojem pro tento typ metadat bylo také pole *Characters*. Vždy nás zajímají pouze jména postav, pro účel tohoto výzkumu považujeme za irelevantní např. pohlaví nebo další atributy postav.
9. **Typ postav:** Jediným zdrojem pro sběr metadat typu typ postav bylo pole *Character*. V případě kolize s pravidlem v bodu 8 bylo rozhodnuto podle konkrétního obsahu daných polí a jeho smyslu.

¹³ V tomto případě je „uživatel“ myšlen knihovník-katalogizátor, o uživatelská obsahová metadata se tedy nejedná.

¹⁴ Toto pole se vyskytuje v některých českých knihkupeckých katalozích a jeho význam je shodný s klíčovými slovy. Stále se tedy také jedná o profesionální metadata.

10. **Důležité události:** Tento typ metadat se nacházel v poli *Important events*. Jedná se pouze o uživatelská metadata, tento typ nebyl v rámci profesionálních metadat identifikován.
11. **Forma/žánr:** Zdrojem pro metadata typu forma/žánr bylo především pole 655 ve formátu MARC, tedy *Rejstříkový termín – žánr/forma*, dále také některá pole MARC vlastní pouze analyzovaným databázím, pokud se v uživatelském zobrazení nazývala forma/žánr. Dále se také mohlo jednat o pole *Forma/Žánr Záhlaví* v návěšt'ovém lístku nebo pole *Genre/Form*. Tento typ zahrnuje pouze profesionální metadata, jak je vysvětleno níže u typu žánr (volný).
12. **Žánr (volný):** Tento typ metadat vznikl z důvodu odlišení metadat vytvořených podle pravidel formátu MARC a volně tvořených termínů, které se zejména v knihovních katalozích anglicky mluvících zemí vyskytují v jednom záznamu současně. U těchto volně tvořených termínů není jasný jejich původ, ve většině případů jsou to data generovaná z externích služeb (více o této problematice v kapitole 8.2), což bylo důležitým faktorem pro jejich oddělení od dat vzniklých při katalogizaci. Typ metadat žánr (volný) tedy zahrnuje volně tvořený obsah polí *Genre* a zároveň všechny žánry ve čtenářských portálech. Jde o kombinaci profesionálních a uživatelských metadat.
13. **Kategorie:** Do tohoto typu metadat řadíme obsah pole *Kategorie* nebo *Categories*, a zároveň kategorie, do kterých byly jednotlivé knihy zařazeny, což může být vyznačeno např. drobečkovou navigací nebo grafickým znázorněním. Patří sem pouze metadata profesionální.
14. **Místo:** Primárním zdrojem pro metadata typu místo bylo pole 651 ve formátu MARC, tedy *Subject Added Entry – Geographic Name*, dalšími zdroji mohla být pole *Geogr. heslo (Záhlaví)* a *Geografické zpřesnění* v návěšt'ovém lístku a *BIC subject*

*qualifier*¹⁵ s notací 1*, což značí kvalifikátor místa. Obdobně jako u typu forma/žánr, i sem zahrnujeme pouze profesionální metadata, více informací u typu místo (volný).

15. **Místo (volný):** Obdobně jako typ žánr (volný) vznikl i typ místo (volný) s cílem odlišit metadata vzniklá při katalogizaci od metadat volně tvořených a/nebo s nezjištěným původem. Spadají tam pole *Location* a *Setting*, zároveň všechna označení místa ze čtenářských portálů.

16. **Časové období:** Zdrojem pro typ metadat časové období je primárně pole 648 ve formátu MARC (*Subject Added Entry – Chronological Term*) a dále také pole *Time period* a všechna označení časového období ze čtenářských portálů.

C) Metadata, která popisují personální kontext knihy: Tento typ metadat nám neříká nic z hlediska pojmů, které se v knize vyskytují, přesto s obsahem velmi úzce souvisí a může ho dále specifikovat. Nacházíme zde odpovědi na otázku, od koho a pro koho. Do obou typů řadíme jak metadata profesionální, tak uživatelská.

17. **Charakteristika autora:** Jediným zdrojem pro metadata tohoto typu je pole *Creator/Contributor Characteristics*.

18. **Uživatelské určení:** Hlavním zdrojem pro metadata typu uživatelské určení bylo pole 521 ve formátu MARC, tedy *Target Audience Note*, a dále pole *Age Range*, *Audience* a *Minimum age*.

D) Metadata, která popisují doplňkové vlastnosti textu: Tento typ metadat doplňuje základní pojmy, jelikož specifikuje vyjádření obsahu ve formě textu. Byl vybrán díky své přímé souvislosti s obsahem knihy, např. víme-li, že text je lyrického charakteru, dozvídáme se také informace o jeho obsahu, stejně tak, pokud víme, že je psaný napínavým stylem. Tato metadata podrobněji odpovídají na otázku jak. Ve všech případech jsou také generovaná z externích služeb, více o této problematice v kapitole

¹⁵ Book Industry Communication vytvořila seznam předmětových kategorií pro globální knižní trh, nyní už tento standard není dále rozvíjen a je nahrazen standardem Thema (BIC Standard Subject Categories, 2020).

8.2. Z těchto typů lze v rámci uživatelských metadat identifikovat pouze typ 20. dojem.

19. **Dějová linka:** Metadata tohoto typu se nacházela v poli *Storyline*.

20. **Dojem:** Metadata byla sbírána z polí *Appeal* a *Tone*.

21. **Styl psaní:** Zdrojem pro metadata typu styl psaní bylo pole *Writing style*.

22. **Tempo:** Metadata tohoto typu pochází z pole *Pace*.

E) Metadata, která popisují kontext knihy (jiný než personální): Tento typ metadat už zdánlivě nesouvisí s obsahem tak přímo. Typ ceny určuje určitou kvalitu nebo hodnotu knihy a typy zahrnující související informační zdroje jsou pro nás důležité jak z hlediska obsahového (název souvisejícího zdroje může nést také obsahovou informaci), tak z hlediska kontextového. Pokud víme, že kniha patří do série knih, může to pro nás být např. informace, že její obsah bude navazovat a/nebo na něj bude navazovat obsah jiné knihy. To je důležité zejména ve chvíli, kdy uživatel vyhledává knihu k četbě. Všechny tyto typy kromě souvisejícího filmu mohou obsahovat jak metadata profesionální, tak metadata uživatelská.

23. **Ceny:** Základním zdrojem pro metadata typu ceny bylo pole 586 ve formátu MARC (tedy *Awards Note*), mohlo se také jednat o pole 500 (*Všeobecná poznámka*), pokud jeho obsahem byla informace o cenách, které kniha získala. Dále sem také zahrnujeme data z polí *Award Winners*, *Awards* a *Literary awards*.

24. **Související dílo/série:** Zde se jedná o názvy knižních sérií, do kterých analyzované knihy patří, a názvy ostatních knih v takové sérii. Nezajímají nás pořadí knihy v sérii ani knihy od stejného autora, knihy, které si lidé také prohlíželi, koupili apod. Metadata tohoto typu pocházela primárně z pole 787 ve formátu MARC, *Other Relationship Entry*, z podpole \$t *Title*, dále se také mohlo jednat o některá vlastní pole ve formátu MARC (např. pole *DIL*, které používá Městská knihovna Louny). Další metadata tohoto typu se nacházela i v polích *Series*, *Books in This Series*, *Added title*, *Série*, *Series info* a *Work-to-work relationships*. Zde je důležité poznamenat, že „series“ znamená v angličtině i „edice“. Tento význam slova „series“ považujeme pro tento výzkum za

irelevantní, jelikož nic nevyovídá o obsahu knihy. Rozlišení mezi dvěma významy termínů „series“ záleželo na hodnocení autorky v dané situaci, v každé databázi byl tento termín použit jinak. Další důležitou poznámkou k tomuto typu metadat je to, že byla sbírána vždy jen ta data, která se nacházela přímo v záznamu jako textové informace, za relevantní tedy nebyla považována data na odkazovaných stránkách ani názvy souvisejících děl jako obrazové informace (fotografie obálky).

25. **Související film:** Zdrojem pro tento typ metadat bylo pole *Related movies*.

F) Specifické typy uživatelských metadat: Tyto typy byly v průběhu výzkumu nalezeny mezi uživatelskými metadaty a nedají se zařadit do žádné z výše uvedených kategorií. Jak již bylo řečeno, u uživatelských metadat nebylo možné oddělit obsahová metadata od jiných, proto se zde nachází i typy, které bychom mezi klasická obsahová metadata nezařadili. Z toho důvodu se k těmto kategoriím neřadí žádná profesionální metadata. Nemusí to ale nutně znamenat, že se dané metadatové výroky v profesionálních databázích nenacházely, pouze nebyly v rámci obsahových metadat sbírány. Těchto metadatových výroků je v porovnání s ostatními typy velmi malé množství, proto nebudou mít negativní vliv na výsledky porovnání. Jejich zdrojem byly výhradně tagy a políčky, o kterých bylo pojednáno samostatně výše.

26. **Nadřazené:** Tato kategorie metadat byla použita pro pojmy nadřazené slovu beletrie, např. literatura.

27. **Rok vydání**

28. **Autor**

29. **Jazyk**

30. **Hodnota obsahu:** Sem patří zejména metadatové výroky udávající hodnocení vnějších subjektů – např. časopisy zabývající se literaturou daného žánru – nebo označení „klasika“.

Pro přehlednost zde ještě uvedeme stručnou tabulku, ve které bude vyznačeno, které metadatové typy obsahují metadata profesionální, uživatelská nebo oboje.

Tabulka 2: Typy metadat a jejich aplikovatelnost na profesionální a uživatelská metadata

Skupina (A-F)	Typ metadat	Obsahuje profesionální metadata	Obsahuje uživatelská metadata
A) Metadata, která reprezentují řízené slovníky	Konspekt	X	
	MDT	X	
	Předmět LCSH	X	
	DDT	X	X
	Předmět FAST	X	
	Klíčová slova	X	
B) Metadata, která popisují základní entity obsahu knih	Téma	X	X
	Postavy	X	X
	Typ postav	X	X
	Důležité události	X	X
	Forma/žánr	X	
	Žánr (volný)	X	X
	Kategorie	X	
	Místo	X	
	Místo (volný)	X	X
	Časové období	X	X
C) Metadata, která popisují personální kontext knihy	Charakteristika autora	X	X
	Uživatelské určení	X	X
D) Metadata, která popisují doplňkové vlastnosti textu	Dějová linka	X	
	Dojem	X	X
	Styl psaní	X	
	Tempo	X	

Skupina (A-F)	Typ metadat	Obsahuje profesionální metadata	Obsahuje uživatelská metadata
E) Metadata, která popisují kontext knihy (jiný než personální)	Ceny	X	X
	Související dílo/série	X	X
	Související film	X	X
F) Specifické typy uživatelských metadat	Hodnota obsahu		X
	Nadřazené		X
	Rok vydání		X
	Autor		X
	Jazyk		X

5.5.3 Speciální případ: pole Subject

Jelikož v některých databázích nebylo k dispozici zobrazení ve formátu MARC, přestože se zjevně na pozadí záznamu nachází, docházelo v těchto případech k situaci, kdy nešlo snadno určit typ zobrazených metadat. Proto bylo stanoveno, jak v takových případech postupovat (vztahuje se na pole *Subject* a v některých databázích i na pole *Personal subject*, *Corporate subject*, *Subject Added Entry – Topical Term* a *Subject-Meeting Name*).

- Metadatové výroky, ve kterých se nenachází pojem „Fiction“ ani „fast“, byla zařazena jako typ téma, jelikož není možné určit, v jakém poli dle MARC se nacházela ani jaké schéma bylo při jejich vytváření použito.
- Metadatové výroky, ve kterých se nachází pojem „Fiction“ nebo „fast“:
 - řadíme jako typ postavy, pokud se v nich nachází pojem „Fictitious character“ (stále platí, že tento pojem v zápisu vynecháváme);
 - řadíme jako místo, pokud začínají označením místa;
 - řadíme jako předmět LCSH, pokud začínají jiným označením a zároveň obsahují pojem „Fiction“;
 - řadíme jako předmět FAST, pokud začínají jiným označením a zároveň obsahují pojem „fast“ (a tento pojem v zápisu vynecháváme).

- Vždy platí, že metadatové výroky rozdělujeme na části a zapisujeme každou část jednotlivě, zároveň v rámci jednoho typu metadat položky deduplikujeme, aby nedocházelo k jejich opakování.

5.5.4 Další pravidla zápisu dat

Pro zápis metadatových výroků do tabulky byla definována následující pravidla:

- Pokud se v databázi kniha nenacházela, bylo do příslušné buňky zapsáno „n“.
- Pokud se v databázi kniha nacházela, ale určitý typ metadatového výroku nikoli, bylo do příslušné buňky zapsáno „-“, např. v zahraničních knihovních katalozích nejsou používány metadatové výroky typu Konspekt.
- Na pořadí termínů nezáleží.
- Na interpunkci nezáleží, např. tečky na konci pojmů nemají význam, pojem „science fiction“ zcela odpovídá pojmu „science-fiction“.
- Uvádíme pouze termíny v jazyce databáze, v českých databázích české a slovenské (slovenština nebrání pochopení významu), v anglických databázích anglické termíny. Neuvádíme např. termíny z Gemeinsame Normdatei, které jsou použity v databázi Dublin Public Library. Tím je zároveň řečeno, že pokud je v záznamu termín dvojjazyčně, uvádíme ho pouze jednou v jazyce databáze. (Národní knihovna České republiky uvádí pole forma/žánr i v angličtině.)
- Pokud se notace a/nebo termín skládá z více částí, zapisujeme je odděleně, např. termín „World War, 1939-1945 -- Youth -- France -- Fiction.“ zapíšeme jako čtyři samostatné termíny.
- Analogicky se každá kategorie a/nebo podkategorie počítá sama za sebe a zapisuje se do samostatné buňky.
- V rámci jedné knihy, jedné databáze a jednoho typu metadat deduplikujeme termíny, které by se opakovaly. To se týká zejména pojmu „Fiction“, který se zpravidla vyskytuje na konci každého řádku metadatových výroků typu LCSH, ale i dalších případů z knihovních a knihkupeckých katalogů. Netýká se uživatelských metadatových výroků, kde mají případné duplicity svůj význam.

- Uvádíme pouze kategorie, jejichž názvy nesou informaci. Z tohoto důvodu nezapisujeme termíny jako „beletrie“, „fiction“, „A to Z“, apod.
- Vynecháváme také pojmy, které se vztahují k velkoformátovému tisku. Jejich uvádění by znevýhodňovalo databáze, kde se zrovna náhodou nenachází jiný výtisk než velkoformátový.

5.6 Analýza dat

Po sběru dat následovala fáze analýzy dat. Tato analýza probíhala vždy nejprve pro jednotlivé analyzované knihy a následně byl vytvořen průměr všech knih dohromady. Proto je důležité, aby bylo k dispozici dostatečně velké množství dat, aby tato zprůměrovaná čísla měla vypovídací hodnotu.

Celá analýza byla provedena v tabulkovém editoru Microsoft Excel a bylo pro ni využito více funkcí, které tento editor nabízí, např. součet, průměr, logické funkce nebo obarvení buněk pro lepší přehlednost. Všechny číselné hodnoty byly zaokrouhlovány na dvě desetinná místa. Pro přehlednost je kapitola rozdělena na více částí, ve kterých jsou popsány jednotlivé části analýzy.

5.6.1 Objektivita/subjektivita uživatelských metadatových výroků

Na uživatelská metadata se vztahuje analýza, která si klade za cíl určit poměr subjektivních a objektivních metadatových výroků podle definice v kapitole 5.1.5. Pro každý jednotlivý pojem bylo určeno, zda se jedná o pojem subjektivní nebo objektivní a v případě, že šlo o objektivní metadatový výrok, byl určen jeho typ podle typologie v kapitole 5.5.2.

Přiřazování typů u nejasných termínů se řídilo následujícími pravidly, která platila pro všechny knihy:

- „série“, „series“, „stand alone“ nebo „prvotina“ označujeme jako související dílo/série;
- „zfilmované“, „film“ nebo „book to movie“ označujeme jako související film;
- „bestseller“ a „edice světový bestseller“ označujeme jako ceny;
- emoce čtenáře (vč. „suspense“) označujeme jako dojem;
- „klasika“ nebo „classics“ označujeme jako hodnota obsahu;

- „knihy“, „literatura“, „knihy abecedně“ apod. označujeme jako nadřazené;
- „drama“ označujeme jako žánr (volný).

Všechna uživatelská metadata, která se nacházela ve čtenářských portálech v jiných „polích“ než tagy a poličky, byla automaticky hodnocena jako objektivní. Dále byl určen celkový počet objektivních uživatelských metadatových výroků a jejich procento z celkového počtu uživatelských metadatových výroků. Procenta objektivních a subjektivních uživatelských metadatových výroků musí dohromady dávat 100 %. Tabulky s výsledky se nachází v příloze 6, každá kniha má vlastní list označený svým pořadovým číslem.

Zde je vhodné upozornit na jeden konkrétní případ metadatových výroků, pro které bylo rozhodnuto, že nebudou hodnoceny ani jako objektivní, ani jako subjektivní. Jedná se o poličku „default“, která se u některých knih vyskytuje v databázi Goodreads, název této poličky nenesení žádný smysl, nepřináší žádnou novou informaci. Na tuto skutečnost je vždy u konkrétní knihy upozorněno, u každé knihy se tato polička nacházela maximálně jednou.

5.6.2 Shoda uživatelských metadat s profesionálními

V této fázi bylo určeno, kolik z objektivních uživatelských metadat odpovídá metadatovým výročkům profesionálním. Pro všechny jednotlivé hodnoty objektivních uživatelských metadatových výroků bylo určeno, zda se hodnoty se stejným obsahem (smyslem) nachází u stejné knihy i v profesionálních metadatach. Podle toho byly příslušné buňky označeny buď zelenou barvou (hodnoty odpovídají profesionálním) nebo oranžovou barvou (hodnoty neodpovídají profesionálním). Opět byl pro oba případy určen celkový počet hodnot, stejně jako rozdělení podle jednotlivých typů metadat. Procenta hodnot, která odpovídají a neodpovídají profesionálním metadatovým výročkům, musí dohromady dát 100 %. Tabulky s výsledky se nachází také v příloze 6.

Pro určování, které uživatelské hodnoty odpovídají profesionálním, byla stanovena následující pravidla, aby u různých knih zůstalo hodnocení vždy stejné a objektivní:

- musí se shodovat podstata obou hodnot, např. „vražda“ není „oběť vraždy“;

- pokud je alespoň jedna část hodnoty subjektivní, označíme ji celou jako subjektivní, např. „top thrillery“;
- pokud alespoň jedna část hodnoty neodpovídá profesionálním metadatovým výrokům, označíme ji celou jako „neodpovídá“ (oranžovou barvou);
- všechna témata považujeme za objektivní;
- „série“ nebo „series“ odpovídá profesionálním metadatům, pokud se mezi nimi vyskytuje jakýkoli metadatový výrok typu související dílo/série;
- „zfilmované“, „film“ nebo „book to movie“ odpovídá profesionálním metadatům, pokud se mezi nimi vyskytuje jakýkoli metadatový výrok typu související film;
- „bestseller“ a „edice světový bestseller“ odpovídá profesionálním metadatům, pokud se mezi nimi (nebo v nich) vyskytuje slovo bestseller, např. „The New York Times Bestseller“;
- „prvotina“ odpovídá profesionálním metadatům, pokud se mezi nimi objevuje slovo prvotina;
- emoce čtenáře odpovídají profesionálním metadatům, pokud se mezi nimi vyskytuje slovo s odpovídajícím významem, např. hodnotám napětí nebo překvapení odpovídá hodnota suspense;
- „stand alone“ odpovídá profesionálním metadatům, pokud se mezi nimi vyskytuje slovo stand alone;
- „klasika“ nebo „classics“ odpovídá profesionálním metadatům, pokud se mezi nimi vyskytuje slovo klasika, classic nebo classics;
- „knihy“, „literatura“, „knihy abecedně“ apod. vždy odpovídají profesionálním metadatům.

Byly zde také vypsány konkrétní hodnoty objektivních metadatových výroků, které neodpovídají profesionálním metadatovým výrokům, byly však deduplikovány podle jejich smyslu, a zároveň i mezi jazyky. U hodnot složených z více částí byly také vypsány pouze ty části, které profesionálním metadatům neodpovídaly. Zde je cílem odhalit přidanou hodnotu uživatelských metadat oproti profesionálním. Tyto konkrétní hodnoty byly vždy také sečteny a byl u nich uveden typ metadatových výroků. Někdy se zde vyskytují typy metadatových výroků, které se v záznamech v databázích nacházely, pouze nebyly určeny ke sběru, tyto hodnoty nebudou

započítány do odpovědi na příslušnou výzkumnou otázku (viz kapitola 7.5). Jedná se o tyto typy metadatových výroků: nadřazené, rok vydání, autor, jazyk.

5.6.3 Počet metadatových výroků pro jednotlivé knihy

Pro profesionální a uživatelská metadata byl vypočítán celkový počet metadatových výroků dle jednotlivých typů. Výsledky jsou uloženy v tabulce, kde řádky označují typy metadatových výroků a sloupce analyzované knihy. Pro typy metadatových výroků, které se v profesionálních nebo uživatelských metadatach nevyskytují (viz tabulka v kapitole 5.5.2), je vždy uvedena nula. Tabulka se nachází v příloze 7.

5.6.4 Mapování na Thema

V této fázi jsme se zabývali mapováním sebraných metadatových výroků na standard Thema, který byl blíže popsán v kapitole 5.1.2. Mapování můžeme definovat jako sémantický vztah mezi dvěma pojmy (Vocabulary Encoding Schemes, 2020). Jelikož Thema slouží k obecnému stanovení předmětu knih bez nutnosti používat specifická metadatová schémata, používáme ho zde jako prostředníka mezi světy knihoven, knihkupectví a čtenářů. Tento standard byl díky svým vlastnostem nejvhodnějším kandidátem. Cílem je zjistit míru sémantické interoperability profesionálních a uživatelských metadat. Opět platí, že využíváme dostatečné množství knih, abychom byli schopni výsledky alespoň částečně generalizovat. A jelikož metadatové výroky mohou standardu odpovídat na více úrovních (některé odpovídají přesněji a jiná volněji), rozhodli jsme se také využít dva pojmy, které specifikují přesnost mapování (viz kapitola 5.1.3 a níže).

Pro každou knihu byl vytvořen vlastní list určený pro mapování metadatových výroků na Thema. Byly sem zkopírovány všechny profesionální metadatové výroky, byly deduplikovány podle jejich významu (bez ohledu na typ metadat a databázi, ze které pochází, tyto údaje pro účely mapování považujeme za irelevantní) a rozděleny na české a anglické. Byly seřazeny abecedně podle sloupce obsahujícího anglické metadatové výroky a těmto byly podle významu přiřazeny odpovídající metadatové výroky české, pokud zde byly zastoupeny. Výsledkem je neznámý počet řádků, kde každý řádek obsahuje český termín, anglický termín nebo oba. Následně byl stejný postup zopakován pro metadatové výroky

uživatelské. Tabulky s hodnotami a jejich příslušnými předmětovými kategoriemi Thema naleznete v příloze 8.

Při deduplikaci bylo postupováno podle následujících pravidel:

- nezáleží na velikosti písmen, např. „americká literatura“ odpovídá pojmu „Americká literatura“;
- nezáleží na jednotném nebo množném čísle, např. „společenský román“ odpovídá pojmu „společenské romány“;
- nezáleží na slovosledu, např. „americká literatura“ odpovídá pojmu „literatura americká“;
- pokud je jeden termín kompletně obsažený v druhém, ale není mu nadřazený, nebereme ho v potaz, např. „domestic“ odpovídá pojmu „domestic fiction“, ale „fiction“ neodpovídá pojmu „domestic fiction“;
- jako „fiction“ zároveň počítáme i termíny „fiction & related“, „fiction: general“ nebo „general fiction“;
- do jednoho termínu slučujeme i termíny, které mají shodný význam, ale jiný text, v případě nejistoty můžeme jeden z termínů uvést do závorky, např. „USA (Spojené státy americké)“.

Pro všechny řádky (bez ohledu na typ metadat) byl stanoven typ termínu podle toho, z kolika myšlenkových částí, resp. z kolika termínů Thema se skládají. To slouží k určení, jak složité koncepty jsou ve zkoumaných metadatových výrocích obsaženy, zejména pro porovnání profesionálních a uživatelských metadat. Bylo možné zde zapsat tyto hodnoty: „jednoduchý“, „složený ze dvou“, „složený ze tří“ nebo „složený ze čtyř“. V počtu případů menším než pět byl termín složen z více částí, ale pro tyto výjimky nebyla vytvářena nová kategorie, z důvodu zjednodušení vyhodnocování byly zařazeny do kategorie „složený ze čtyř“.

Následují sloupce Přesnost, Notace a Popis. Pokud odpovídající termín (nebo termíny v případě složených pojmů) nebyl v standardu Thema nalezen¹⁶, bylo do příslušné buňky ve sloupci Přesnost zapsáno „-“ a sloupce Notace a Popis zůstaly

¹⁶ Pro mapování byla použita česká verze standardu Thema (https://www.editeur.org/files/Thema/1.3/Thema_v1.3.3_cs.html). Předpokládá se, že různé jazykové verze standardu mají shodný obsah.

prázdné. V případě, kdy odpovídající termín (resp. termíny) Thema nalezeny byly, bylo pro každou část pojmu určeno, zda se jedná o exactMatch nebo closeMatch (viz kapitola 5.1.3) a do sloupců Notace a Popis byly doplněny přesné notace a slovní vyjádření odpovídajících termínů Thema. Z toho vyplývá, že v případě metadatových výroků složených z více částí bylo nutné vložit řádek, resp. řádky. Pokud však nebylo možné namapovat všechny části složeného termínu, pak byl celý termín hodnocen jako nesplněno.

Všechny názvy souvisejících děl a/nebo sérií byly hodnoceny jako názvy díla, nikoli jako termíny s významem, např. pojem „Levhart“ jako název díla nelze namapovat na Thema, přestože u pojmu „levhart“ jako zvíře by nebyl problém. U geografických míst, která nemají přesně odpovídající protějšek, vybíráme nejbližší nadřazený termín a označujeme ho jako closeMatch.

Na základě této tabulky bylo pro profesionální metadatové výroky sečteno, kolik z nich odpovídá, a naopak neodpovídá Thema, kolik termínů je jednoduchých, složených ze dvou, tří a čtyř a kolik z těchto kategorií odpovídá Thema jako exactMatch, kolik jako closeMatch (a kolik jako kombinace exactMatch a closeMatch v případě složených termínů) a kolik neodpovídá Thema vůbec. Bylo také vypočítáno, kolik z metadatových výroků, která odpovídají Thema, bylo zařazeno jako exactMatch, closeMatch nebo kombinace těchto možností. Pomocí kontrolních výpočtů bylo ověřeno, že byly započítány všechny vypsane metadatové výroky.

Stejný postup byl zopakován i pro uživatelské metadatové výroky a na závěr bylo pro všechny výše uvedené případy porovnáno, který typ metadatových výroků (profesionální nebo uživatelské) má vyšší relativní hodnotu (v procentech). Tabulka s výsledky této části analýzy se nachází v příloze 9.

6. Vyhodnocení analýzy

V následujících několika podkapitolách budou uvedeny zjištěné hodnoty analýzy, která byla popsána v kapitole 5.6 a podkapitolách. Zjištěné hodnoty byly u jednotlivých knih velmi rozdílné, proto předpokládáme, že jejich zprůměrováním získáme určitý obecný vzorec, kterému se budou blížit také metadatové výroky dalších knih, které nebyly v této práci analyzovány. Konkrétní hodnoty můžete nalézt v přílohách, není cílem této práce zde uvádět každou jednotlivou číselnou hodnotu. Tím, co nás zajímá nejvíce, je již zmiňovaný obecný vzorec. Pro účely zápisu a usnadnění porozumění budou všechna čísla níže zaokrouhlena na dvě desetinná místa.

6.1 Objektivní uživatelská metadata

Nyní už se zaměřujeme na uživatelská metadata a určujeme jejich objektivitu nebo subjektivitu podle definice v kapitole 5.1.5. V následující tabulce se nachází rozmezí a průměrný počet metadatových výroků a následně rozmezí a průměr procentuálních podílů objektivních metadat u jednotlivých analyzovaných knih. Z tabulky je zřejmé, že průměrně je objektivních 61 % uživatelských metadatových výroků.

Tabulka 3: Objektivní uživatelská metadata – přehled

Počet metadatových výroků	
rozmezí od-do	31-219
průměrný	94,85
Procentuální část	
rozmezí od-do	35-71 %
průměrný	61 %

Následují zjištěné hodnoty objektivních metadatových výroků podle formy uživatelských metadat, tedy tagů a políček. Doplnkové tabulky se subjektivními metadatovými výroky zde již podle našeho názoru nejsou potřeba. Kompletní

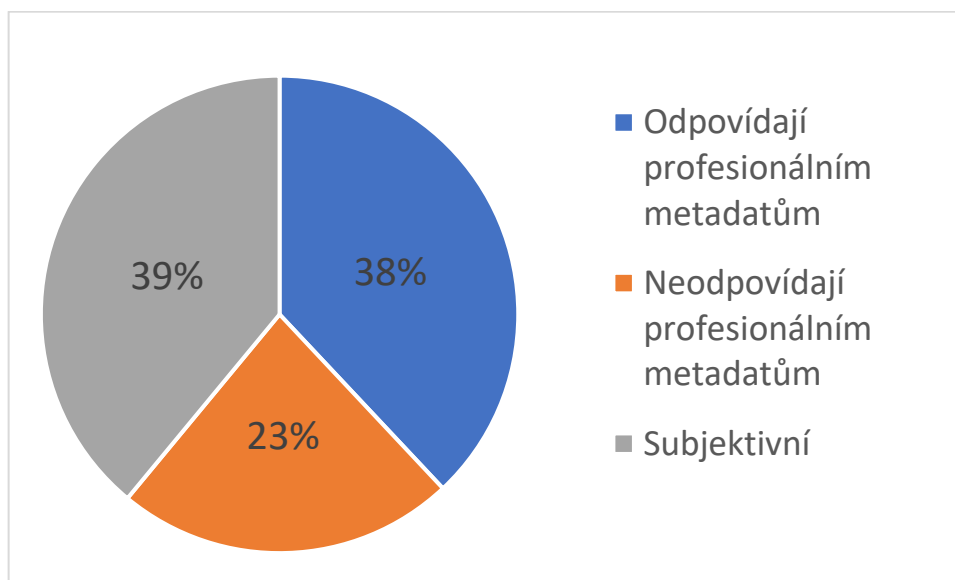
informace můžete v případě potřeby nalézt v příloze 6. Z tabulky vyplývá, že tagy jsou přibližně o 53 % objektivnější než poličky.

Tabulka 4: Objektivní uživatelská metadata podle formy

	Tagy	Poličky
Průměrný počet metadatových výroků	30,48	21,96
Procentuální počet metadatových výroků	81 %	28 %

A následující graf ilustruje, kolik uživatelských metadat je subjektivních a kolik objektivních. Pro objektivní metadata potom dále určuje, kolik z nich odpovídá a neodpovídá profesionálním metadatům. Tomu se pak budeme věnovat více v dalších kapitolách.

Graf 1: Rozložení uživatelských metadat



6.1.1 Objektivní uživatelské metadatové výroky, které odpovídají profesionálním

Pro objektivní uživatelské metadatové výroky, které odpovídají profesionálním, uvádíme tabulku, která zobrazuje zjištěné informace. Opět připomínáme, že byly porovnávány hodnoty metadatových výroků a výsledná čísla jsou uváděna pouze z metadatových výroků, které již byly v předchozí kapitole určeny jako objektivní.

Tabulka 5: Objektivní uživatelská metadata, která odpovídají profesionálním – přehled

Počet metadatových výroků	
rozmezí od-do	20-125
průměrný	57,33
Procentuální část	
rozmezí od-do	43-84 %
průměrný	62 %

Pro porovnání metadatových forem uvádíme pro přehlednost také tabulku, ze které vyplývá, že poličky odpovídaly profesionálním metadatovým výročkům přibližně o 5 % častěji než tagy.

Tabulka 6: Objektivní uživatelská metadata, která odpovídají profesionálním, podle formy

	Tagy	Poličky
Průměrný počet metadatových výroků	21,63	16,74
Procentuální počet metadatových výroků	71 %	76 %

6.1.2 Objektivní uživatelské metadatové výroky, které neodpovídají profesionálním

Pro objektivní uživatelské výroky, které neodpovídají výročkům profesionálním, resp. jejich hodnotám, opět uvedeme zjištěné údaje pro přehlednost v tabulkách. Dají se tak snadno porovnat s tabulkami v předchozí kapitole. Opět je zřejmé, že uživatelská metadata, která odpovídají a neodpovídají profesionálním metadatům, dávají dohromady 100 %. Pro porovnání je také vhodný výšečový Graf 1 výše.

Tabulka 7: Objektivní uživatelská metadata, která neodpovídají profesionálním – přehled

Počet metadatových výroků	
rozmezí od-do	9-96
průměrný	37,52
Procentuální část	
rozmezí od-do	16-57 %
průměrný	38 %

Tabulka 8: Objektivní uživatelská metadata, která neodpovídají profesionálním, podle formy

	Tagy	Poličky
Průměrný počet metadatových výroků	8,81	5,22
Procentuální počet metadatových výroků	29 %	24 %

Z tabulky vyplývá, že mezi tagy se našly originální metadatové výroky přibližně o 5 % častěji než mezi poličkami.

Zajímalo nás také, které typy metadatových výroků nejčastěji neodpovídaly profesionálním metadatovým výročkům, resp. jejich hodnotám. Jednalo se o tyto typy metadatových výroků (v závorce je uveden absolutní počet, podle kterého jsou typy také seřazeny):

- postavy (462), např. Tom Watson;
- téma (166), např. osud nebo civil rights;
- ceny (71), např. světový bestseller nebo Australian Book Industry Award (ABIA) Nominee for International Book (2016);
- místo (volný) (69), např. Europe nebo Ashbury, Buckinghamshire, England, UK;
- žánr (volný) (35), např. realistic fiction nebo drama;
- související dílo/série (25), např. rozhlasové zpracování nebo TV seriál;

- autor (14), např. Jo Nesbo;
- rok vydání (12), např. vydané 2018;
- dojem (11), např. ironie nebo dark;
- související film (9), např. zfilmované nebo A Man Called Ove [2015 movie];
- uživatelské určení (8), např. pro ženy;
- jazyk (7), např. ich forma nebo English;
- důležité události (5), např. holocaust;
- hodnota obsahu (5), např. nejlepší dle časopisu Pevnost nebo classics;
- charakteristika autora (5), např. female author nebo British;
- časové období (3), např. 40.-50. léta 20. století;
- typ postav (2), blind girl a female protagonist;
- nadřazené (0)¹⁷.

6.2 Porovnání profesionálních a uživatelských metadat podle typu metadatových výroků

V tuto chvíli budeme již porovnávat přímo nasbírané metadatové výroky mezi profesionálními a uživatelskými metadaty. V tabulce je vždy uveden celkový počet metadatových výroků daného typu v profesionálních databázích, poté průměr metadatových výroků na jeden metadatový záznam, ve kterém je možné tento typ metadatového výroku zobrazit. Tedy celkový počet metadatových výroků byl vydělen hodnotou zobrazenou na grafech 7 a 8 níže a toto číslo bylo vyděleno počtem analyzovaných knih (tedy dvaceti sedmi). A stejné hodnoty i pro metadata uživatelská. Díky tomu můžeme mezi profesionálními a uživatelskými metadaty porovnat jak absolutní počet metadatových výroků, tak i poměrný počet na jeden metadatový záznam. Výpočet této hodnoty je navržen tak, aby byl porovnatelný mezi oběma typy metadat i přesto, že nemají stejný počet zdrojových databází. Tabulky s výpočty uvedených hodnot naleznete v příloze 7. Jsou zde na konci tabulky zařazeny i ty typy metadatových výroků, které nejsou obsahové, ale byly ve čtenářských portálech nalezeny. U těchto typů je v buňkách pro profesionální

¹⁷ Nadřazené metadatové výroky byly vždy hodnocené jako odpovídající profesionálním metadatům.

metadata vždy nula, přestože je možné, že se tyto údaje v profesionálních záznamech vyskytovaly (např. rok vydání). Do uživatelských metadat jsou započítána pouze metadata objektivní, jelikož metadata subjektivní nemůžeme zařadit do žádného z typů.

Tabulka 9: Porovnání profesionálních a uživatelských metadat podle typů metadatových výroků

	Celkový počet profesionálních metadat	Průměrný počet profesionálních metadat na jeden metadatový záznam	Celkový počet uživatelských metadat	Průměrný počet uživatelských metadat na jeden metadatový záznam
Konspekt	604	0,93	0	0
MDT	1049	2,43	0	0
Předmět LCSH	2922	4,33	0	0
DDT	461	0,85	1	0
Předmět FAST	155	0,72	0	0
Klíčová slova	871	2,15	0	0
Téma	1096	0,94	505	4,68
Postavy	377	0,56	528	9,78
Typ postav	299	1,58	5	0,09
Důležité události	0	0	25	0,23
Forma/žánr	1869	1,61	0	0
Žánr (volný)	672	2,26	728	6,74
Kategorie	1191	2,59	0	0
Místo	521	0,58	0	0
Místo (volný)	55	0,41	213	1,97
Časové období	323	0,44	12	0,11
Charakteristika autora	2	0,07	14	0,26

	Celkový počet profesionálních metadat	Průměrný počet profesionálních metadat na jeden metadatový záznam	Celkový počet uživatelských metadat	Průměrný počet uživatelských metadat na jeden metadatový záznam
Uživatelské určení	38	0,09	64	0,59
Dějová linka	48	0,89	0	0
Dojem	911	3,75	68	0,63
Styl psaní	64	1,19	0	0
Tempo	22	0,81	0	0
Ceny	447	0,59	206	1,91
Související dílo/série	1112	0,9	126	1,17
Související film	0	0	35	0,32
Hodnota obsahu	0	0	7	0,09
Nadřazené	0	0	14	0,17
Rok vydání	0	0	16	0,2
Autor	0	0	24	0,44
Jazyk	0	0	13	0,12

Z tabulky vidíme opět velké rozdíly mezi profesionálními a uživatelskými metadaty. Nejpoužívanější typy metadatových výroků u profesionálních metadat jsou předmět LCSH (s více jak 4 výroky v jednom záznamu), dojem (více než 3 výroky) a průměrně více než dva výroky v jednom metadatovém záznamu mají kategorie, MDT, žánr (volný) a klíčová slova.

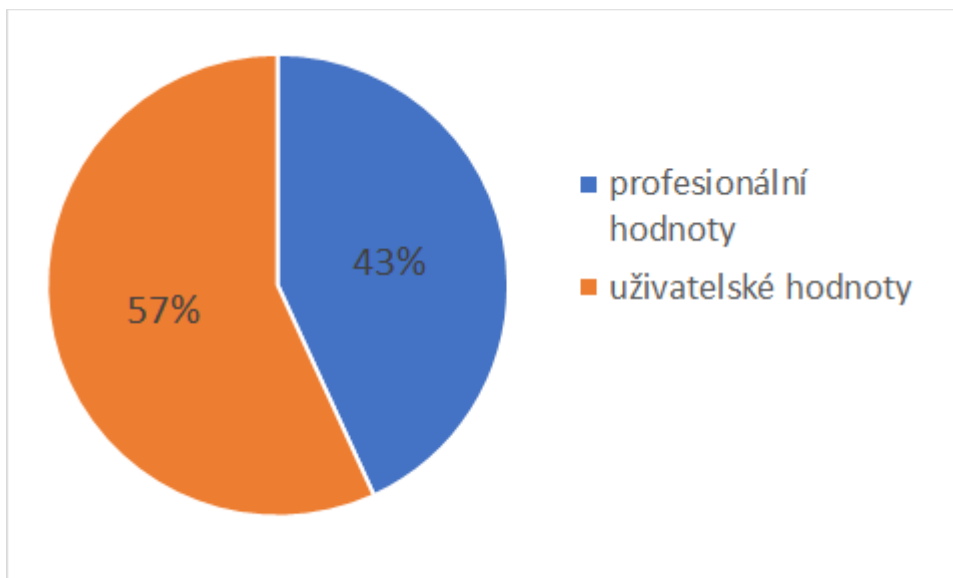
U uživatelských metadat je situace na první pohled jiná. Nejpoužívanějším typem metadatových výroků jsou postavy s průměrně více než 9 výskyty v jednom metadatovém záznamu. Za nimi následují žánr (volný) (více než 6 výskytů), téma

(více než 4 výskyty) a přibližně po jednom výskytu měly typy místo (volný), související dílo/série a ceny.

6.3 Metadata a Thema

Zde je důležité upozornit, že se jedná o metadata objektivní i subjektivní dohromady. Následující údaje vychází z dat v příloze 9, vlastní hodnoty a jejich odpovídající předmětové kategorie standardu Thema najdete v příloze 8. Na následujícím grafu vidíme poměr profesionálních a uživatelských metadatových hodnot, které byly mapovány. Jedná se o počty po deduplikaci hodnot podle jejich významu (viz kapitola 5.6.4), tzn. součty metadatových výroků nebudou odpovídat součtům v předchozích kapitolách, kde šlo o prostý součet buněk.

Graf 2: Poměr unikátních hodnot profesionálních a uživatelských metadat pro zjišťování sémantické interoperability



6.3.1 Hodnoty profesionálních metadatových výroků

V této fázi byly profesionální metadatové výroky, resp. jejich hodnoty, mapovány na předmětové kategorie standardu Thema. V tabulce uvádíme údaje o počtu a procentuálním poměru, které standardu odpovídaly.

Tabulka 10: Profesionální hodnoty, které odpovídaly Thema – přehled

Počet hodnot	
rozmezí od-do	14-41
průměrný	28,96
Procentuální část	
rozmezí od-do	30-71 %
průměrný	51 %

V další tabulce pak uvádíme naopak hodnoty, které Thema neodpovídaly. Odpovídající a neodpovídající hodnoty dávají dohromady opět 100 %.

Tabulka 11: Profesionální hodnoty, které neodpovídaly Thema – přehled

Počet hodnot	
rozmezí od-do	12-48
průměrný	27,96
Procentuální část	
rozmezí od-do	29-70 %
průměrný	49 %

Z tabulek můžeme vyhodnotit, že více než polovina hodnot profesionálních metadatových výroků standardu Thema odpovídá. Dále uvádíme, kolik hodnot bylo jednoduchých a složených.

Tabulka 12: Rozdělení profesionálních hodnot podle komplexnosti

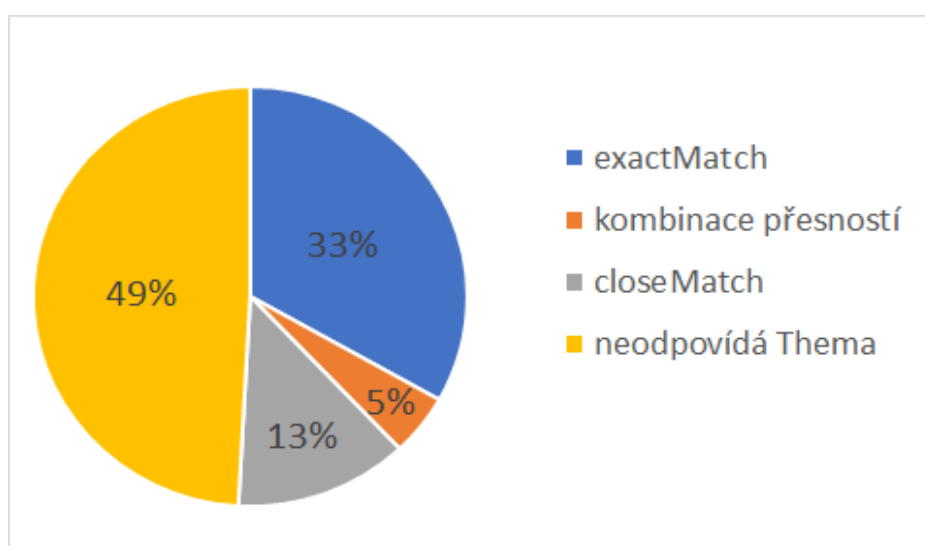
	Průměrný počet hodnot	Procentuální část
Jednoduché hodnoty	38,89	68 %
Hodnoty složené ze dvou	13,89	25 %
Hodnoty složené ze tří	4,04	7 %

	Průměrný počet hodnot	Procentuální část
Hodnoty složené ze čtyř	0,11	0 %

Dále uvádíme, kolik kterých hodnot odpovídalo s jakou přesností (tedy exactMatch, closeMatch nebo jejich kombinací) nebo neodpovídalo standardu Thema vůbec. Pro přehlednost jsou údaje zpracovány do několika grafů.

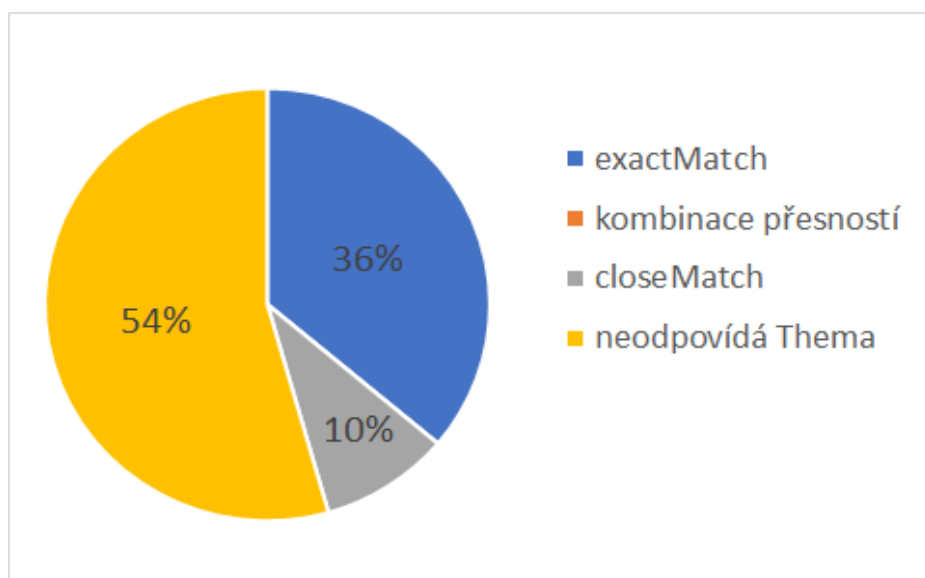
Všechny hodnoty dohromady

Graf 3: Profesionální hodnoty dle přesnosti mapování na Thema



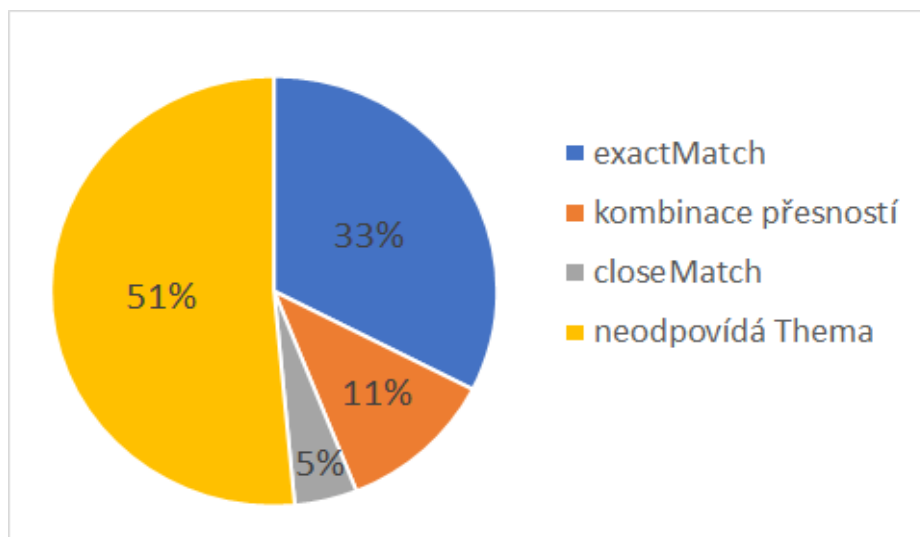
Jednoduché hodnoty

Graf 4: Jednoduché profesionální hodnoty dle přesnosti mapování na Thema



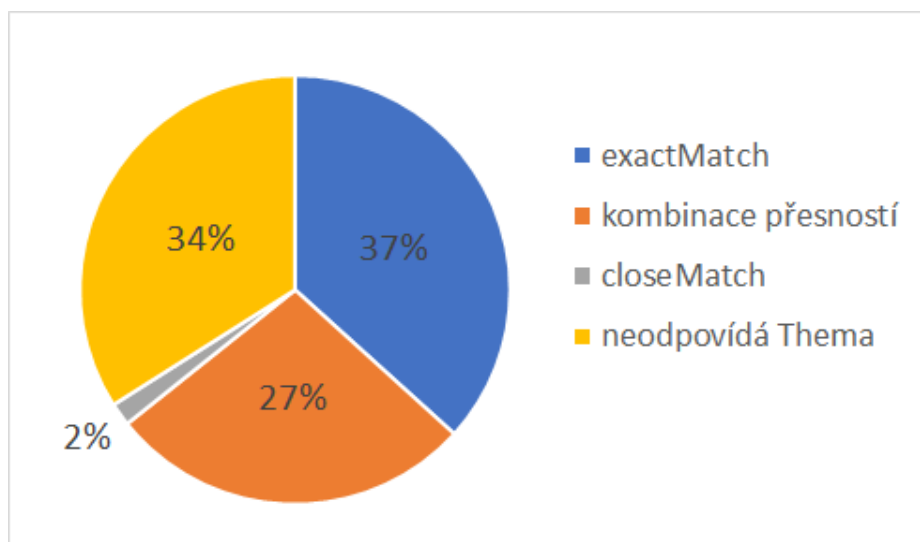
Hodnoty složené ze dvou

Graf 5: Profesionální hodnoty složené ze dvou dle přesnosti mapování na Thema



Hodnoty složené ze tří

Graf 6: Profesionální hodnoty složené ze tří dle přesnosti mapování na Thema



Hodnoty složené ze čtyř v profesionálních metadatech nebyly téměř žádné, proto se jimi pro tuto chvíli zabývat nebudeme.

6.3.2 Hodnoty uživatelských metadatových výroků

Obdobně byly všechny údaje zjištěny také pro uživatelská metadata. V tabulce uvádíme údaje o počtu a procentuálním poměru, které standardu odpovídaly.

Tabulka 13: Uživatelské hodnoty, které odpovídaly Thema – přehled

Počet hodnot	
rozmezí od-do	13-54
průměrný	26,37
Procentuální část	
rozmezí od-do	22-52 %
průměrný	37 %

V další tabulce pak uvádíme naopak hodnoty, které Thema neodpovídaly. Odpovídající a neodpovídající hodnoty dávají dohromady opět 100 %.

Tabulka 14: Uživatelské hodnoty, které neodpovídaly Thema – přehled

Počet hodnot	
rozmezí od-do	16-104
průměrný	48,63
Procentuální část	
rozmezí od-do	48-78 %
průměrný	63 %

Z těchto čísel vidíme, že standardu Thema odpovídalo méně než polovina uživatelských metadatových výroků, a to objektivních i subjektivních dohromady. Dále uvádíme, kolik hodnot bylo jednoduchých a složených.

Tabulka 15: Rozdělení uživatelských hodnot podle komplexnosti

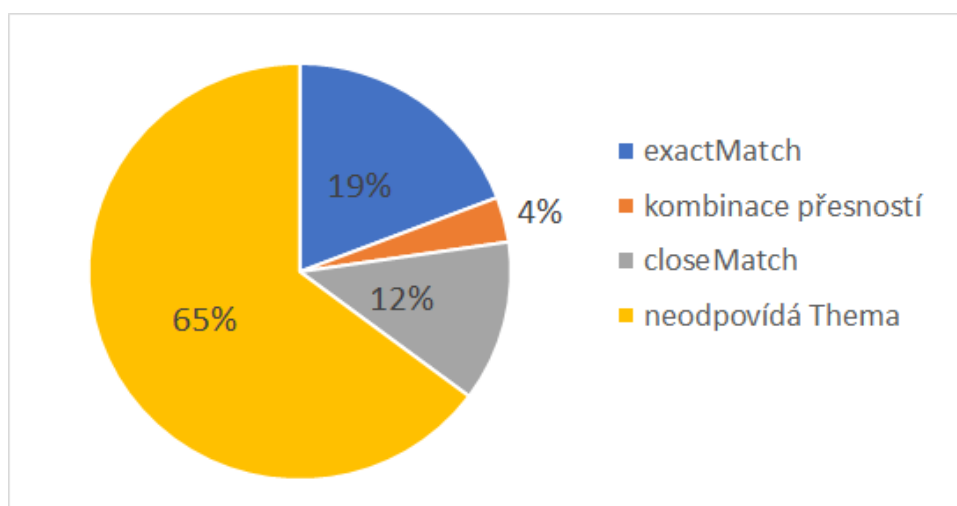
	Průměrný počet hodnot	Procentuální část
Jednoduché hodnoty	53,63	72 %
Hodnoty složené ze dvou	15,85	21 %
Hodnoty složené ze tří	3,81	5 %

	Průměrný počet hodnot	Procentuální část
Hodnoty složené ze čtyř	1,78	2 %

Dále uvádíme, kolik kterých hodnot odpovídalo s jakou přesností (tedy exactMatch, closeMatch nebo jejich kombinací) nebo neodpovídalo standardu Thema vůbec. Pro přehlednost jsou údaje opět zpracovány do několika grafů.

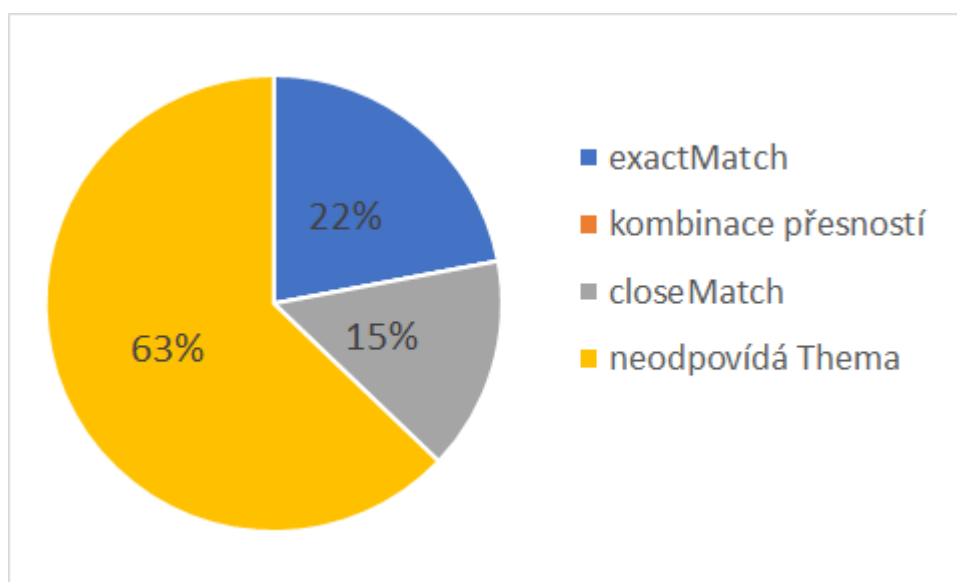
Všechny hodnoty dohromady

Tabulka 16: Uživatelské hodnoty dle přesnosti mapování na Thema



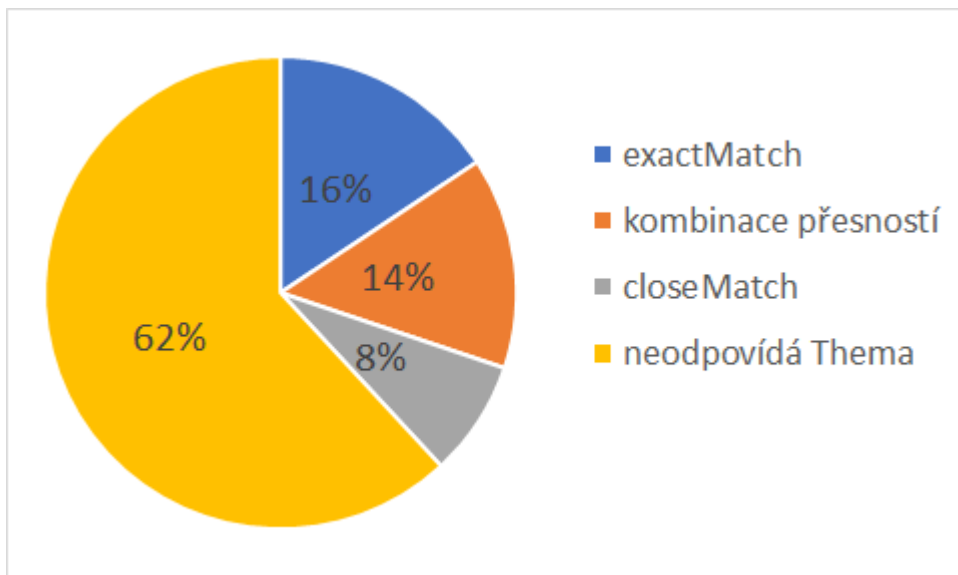
Jednoduché hodnoty

Tabulka 17: Jednoduché uživatelské hodnoty dle přesnosti mapování na Thema



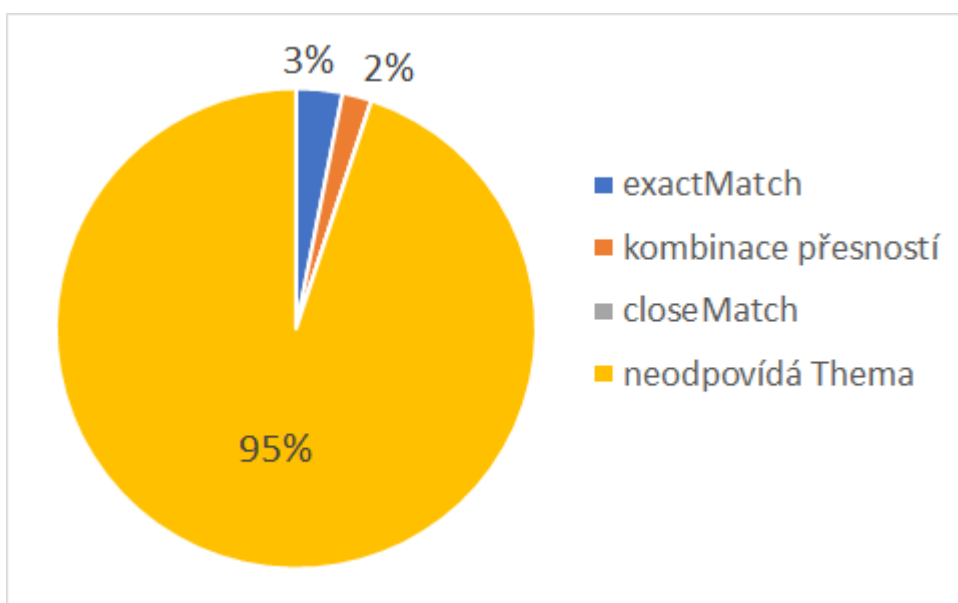
Hodnoty složené ze dvou

Tabulka 18: Uživatelské hodnoty složené ze dvou dle přesnosti mapování na Thema



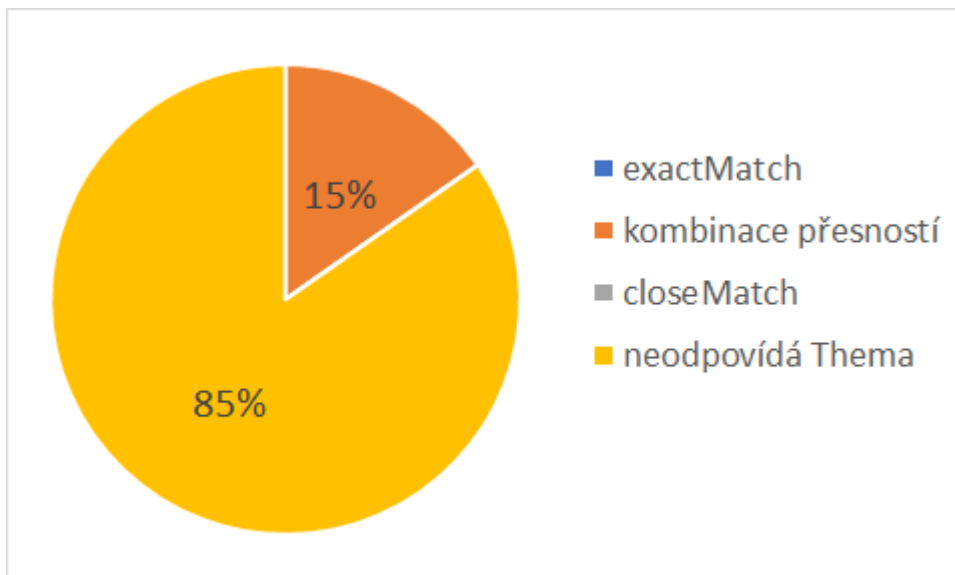
Hodnoty složené ze tří

Tabulka 19: Uživatelské hodnoty složené ze tří dle přesnosti mapování na Thema



Hodnoty složené ze čtyř

Tabulka 20: Uživatelské hodnoty složené ze čtyř dle přesnosti mapování na Thema



6.3.3 Porovnání profesionálních a uživatelských metadatových výroků

Následně bylo pro každou knihu porovnáno, zda standardu Thema odpovídají více profesionální nebo uživatelské metadatové výroky. Zde opět uvedeme celkové, průměrné výsledky, výsledky konkrétních knih se nachází v příloze 9. Thema odpovídaly více profesionální metadatové výroky v 89 % případů (konkrétně u 24 z 27 knih). Více mu odpovídaly metadatové výroky uživatelské v 11 % případů.

Jednoduchých termínů bylo více mezi uživatelskými metadatovými výroky, a to v 63 % případů. Z toho odpovídaly Thema s přesností exactMatch častěji profesionální metadatové výroky (u 93 % knih), uživatelské více odpovídaly ve dvou případech. S přesností closeMatch z jednoduchých metadatových výroků častěji odpovídaly opět profesionální metadatové výroky, ale pouze v 56 % případů, ve zbylých 44 % byly přesnější metadatové výroky uživatelské. Thema neodpovídaly častěji jednoduché metadatové výroky uživatelské, v 81 % případů.

Složených termínů bylo více mezi profesionálními metadatovými výroky, analogicky 63 %. Z termínů složených ze dvou částí bylo více profesionálních metadatových výroků v 67 % případů. Z toho odpovídaly s přesností exactMatch častěji profesionální metadatové výroky (u 89 % knih), u zbylých tří knih odpovídaly častěji metadatové výroky uživatelské. S kombinací obou typů

přesností odpovídaly Thema častěji uživatelské metadatové výroky (v 70 % případů), ve 26 % případů metadatové výroky uživatelské a u 4 % knih (1 případ) bylo shodné množství profesionálních i uživatelských metadatových výroků tohoto typu. S přesností closeMatch odpovídaly Thema v 74 % případů metadatové výroky uživatelské. A vůbec neodpovídaly častěji metadatové výroky uživatelské (u 67 % knih).

Z termínů složených ze tří částí bylo více profesionálních metadatových výroků v 78 % případů. Z toho odpovídaly s přesností exactMatch častěji profesionální metadatové výroky (u 96 % knih), u zbylé knihy byl počet uživatelských a profesionálních metadatových výroků stejný. S kombinací obou typů přesností odpovídaly Thema častěji profesionální metadatové výroky, a to v 70 % případů. S přesností closeMatch v 93 % případů žádné metadatové výroky neexistovaly, ve zbylých dvou případech více odpovídaly metadatové výroky profesionální. A vůbec neodpovídaly častěji metadatové výroky uživatelské (u 89 % knih).

Z termínů složených ze čtyř částí bylo více uživatelských metadatových výroků v 70 % případů a ve 26 % případů žádné takové metadatové výroky neexistovaly. metadatové výroky odpovídající s přesností exactMatch v 92 % případů neexistovaly, u zbylé knihy odpovídaly více metadatové výroky profesionální. Metadatové výroky odpovídající s kombinací přesností v 81 % případů neexistovaly, u zbylých pěti knih odpovídaly více metadatové výroky uživatelské. Metadatové výroky odpovídající s přesností closeMatch neexistovaly ve 100 % případů. Z tohoto typu termínů Thema neodpovídaly nejčastěji metadatové výroky uživatelské (v 63 % případů).

Celkem z metadatových výroků, které odpovídaly Thema, se nejčastěji jednalo o přesnost exactMatch u profesionálních metadatových výroků (v 78 % případů), o kombinaci přesností u uživatelských metadatových výroků (v 52 % případů) a o přesnost closeMatch u metadatových výroků uživatelských, v 81 % případů.

7. Odpovědi na výzkumné otázky

7.1 Které typy obsahových metadatových výroků existují? Liší se to mezi profesionálními a uživatelskými metadaty?

V rámci této otázky budeme brát v potaz pouze metadata obsahová, tzn. budou vyřazeny typy nadřazené, rok vydání, autor a jazyk. Identifikovali jsme 26 kategorií obsahových metadatových výroků. Mohou se nacházet jak mezi profesionálními nebo uživatelskými metadaty, tak mezi oběma typy současně. Přehled toho, které typy metadatových výroků obsahují profesionální a uživatelské metadatové výroky, přikládáme níže v tabulce.

Tabulka 21: Typy obsahových metadat a jejich aplikovatelnost na profesionální a uživatelská metadata

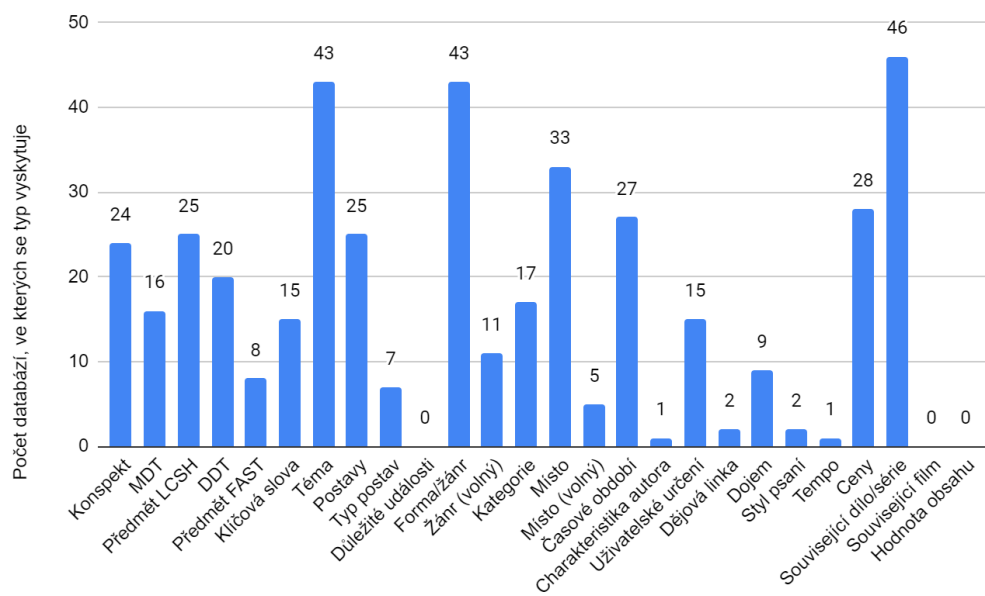
Skupina (A-F)	Typ metadat	Obsahuje profesionální metadata	Obsahuje uživatelská metadata
A) Metadata, která reprezentují řízené slovníky	Konspekt	X	
	MDT	X	
	Předmět LCSH	X	
	DDT	X	X
	Předmět FAST	X	
	Klíčová slova	X	
B) Metadata, která popisují základní entity obsahu knih	Téma	X	X
	Postavy	X	X
	Typ postav	X	X
	Důležité události	X	X
	Forma/žánr	X	
	Žánr (volný)	X	X
	Kategorie	X	
	Místo	X	
	Místo (volný)	X	X

	Časové období	X	X
C) Metadata, která popisují personální kontext knihy	Charakteristika autora	X	X
	Uživatelské určení	X	X
D) Metadata, která popisují doplňkové vlastnosti textu	Dějová linka	X	
	Dojem	X	X
	Styl psaní	X	
	Tempo	X	
E) Metadata, která popisují kontext knihy (jiný než personální)	Ceny	X	X
	Související dílo/série	X	X
	Související film	X	X
F) Specifické typy uživatelských metadat	Hodnota obsahu		X

V příloze 7 naleznete také rozložení těchto typů metadatových výroků přímo v konkrétních databázích. Jsou z něj patrné velké rozdíly v tom, jaká obsahová metadata se nachází v profesionálních a uživatelských databázích. V tuto chvíli bereme v potaz typ metadatových výroků, které se v dané databázi vyskytly alespoň jednou, tedy daná databáze umožňuje jejich zápis. Pro ilustraci přikládáme dva grafy, které znázorňují množství databází, v kterých se nachází jednotlivé typy metadatových výroků, a to zvlášť pro metadata profesionální (Graf 7) a zvlášť pro obsahová (Graf 8).

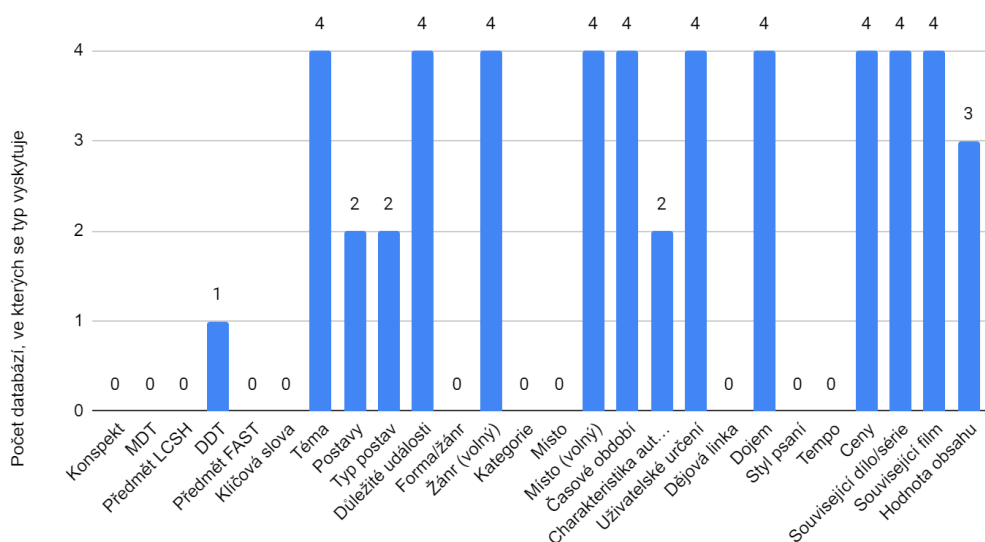
Typy metadatových výroků v profesionálních metadatech

Graf 7: Počet databází, ve kterých se vyskytují různé typy profesionálních metadat



Typy metadatových výroků v uživatelských metadatech

Graf 8: Počet databází, ve kterých se vyskytují různé typy uživatelských metadat



Z grafů je zřejmé, že mezi profesionálními a uživatelskými metadaty je velký rozdíl. Databází s profesionálními metadaty bylo celkem 68, žádný typ metadatových výroků se však nevyskytoval ve všech analyzovaných databázích. Nejvyšší počet databází je 46, v tolika databázích se nacházel typ metadatových

výroků související dílo/série. Naopak se mezi profesionálními metadaty v těchto databázích nevyskytovaly žádné metadatové výroky typů důležité události, související film a hodnota obsahu.

Na rozdíl od profesionálních metadat, v rámci uživatelských metadat se hned deset typů metadatových výroků nacházelo ve všech analyzovaných databázích. Jedná se o typy téma, důležité události, žánr (volný), místo (volný), časové období, uživatelské určení, dojem, ceny, související dílo/série a související film. Všimněme si, že mezi těmito deseti nejčastějšími typy se nachází dva, které se v profesionálních metadatech nevyskytovaly vůbec. Naopak se v uživatelských metadatech nevyskytovaly tyto typy metadatových výroků: většina metadatových výroků, které reprezentují řízené slovníky, forma/žánr, místo, kategorie a tři typy metadatových výroků, které byly v profesionálních metadatech generovány z externích služeb. Některé z těchto jmenovaných typů jsou však v uživatelských metadatech reprezentovány jinými typy: namísto typu forma/žánr uživatelská metadata mají žánr (volný), místo místa mají místo (volný), místo kategorií mají metadatové výroky ve formě políček. To může naznačovat, že metadatové typy skupiny A a některá data generovaná z externích služeb nejsou pro uživatele příliš důležité.

Závěr

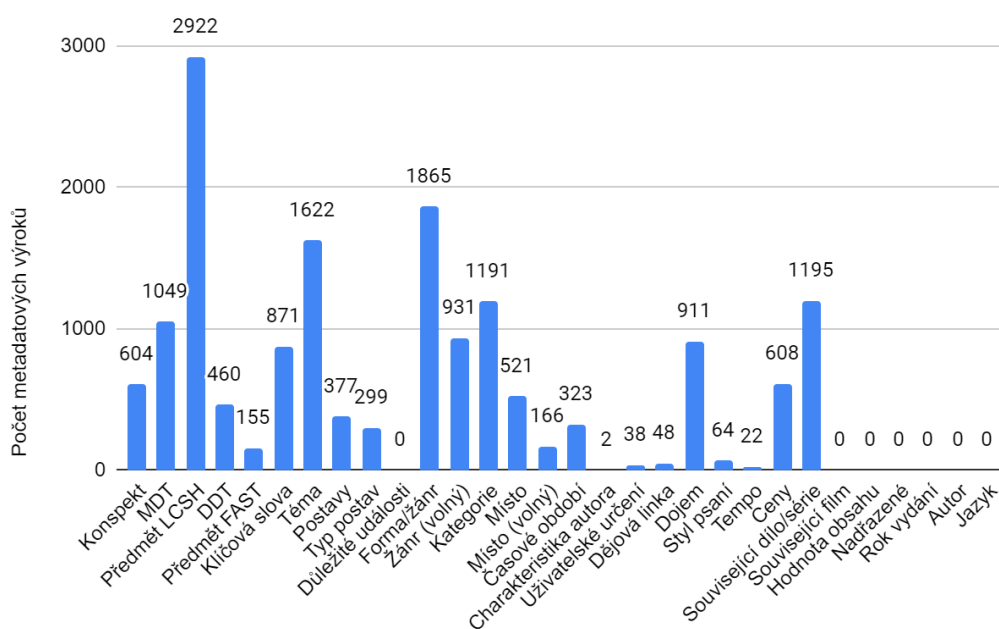
Existuje 26 typů obsahových metadatových výroků, což je podle nás velké množství. Tyto metadatové výroky jsou velmi různorodé, od výroků reprezentujících řízené slovníky až po doplňkové informace o textu a kontextu popisovaného informačního zdroje.

Porovnáváme, které databáze umožňují zápis kterých typů metadatových výroků. Už na této úrovni existují velké rozdíly mezi profesionálními a uživatelskými metadaty. Mezi profesionálními metadaty je více typů celkem, avšak jejich rozdělení v databázích je velmi různorodé. Neexistuje typ, který by se vyskytoval ve všech analyzovaných databázích. U uživatelských metadat je tomu naopak: celkově je typů méně, ale častěji se shodují napříč různými databázemi. Nejčastěji se vyskytující typy naznačují, které informace o knihách uživatele nejvíce zajímají.

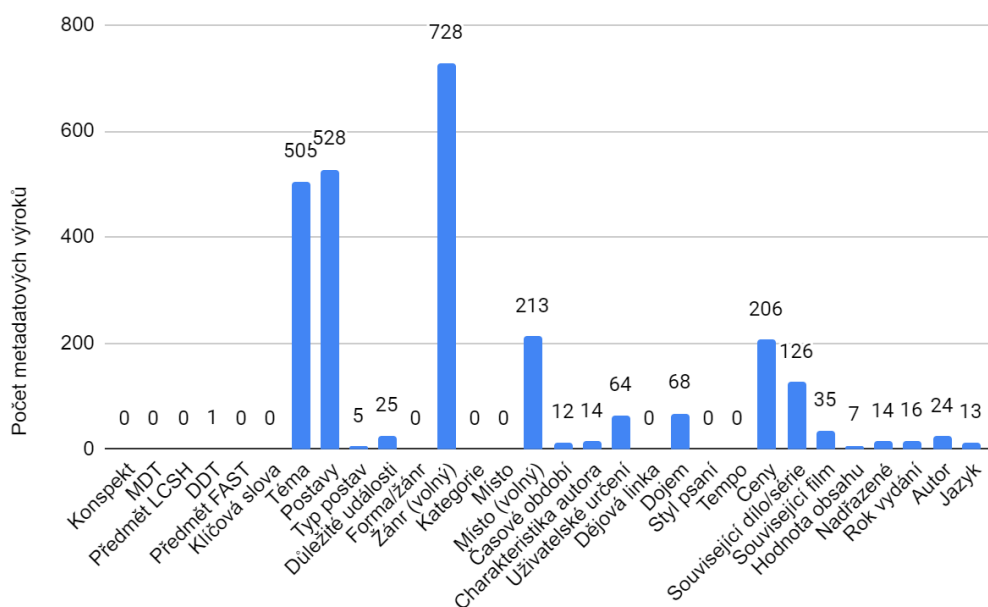
7.2 Kolik metadatových výroků kterého typu databáze obsahují?

Zde se už nacházíme na úrovni konkrétních metadatových výroků, které byly v metadatových záznamech analyzovaných knih nalezeny. Uvádíme grafy obdobné grafům 7 a 8 výše, ale tentokrát s konkrétními počty metadatových výroků. Při pohledu na grafy je však potřeba mít v paměti, že profesionálních databází bylo 68 a uživatelské 4.

Graf 9: Profesionální metadatové výroky dle typů

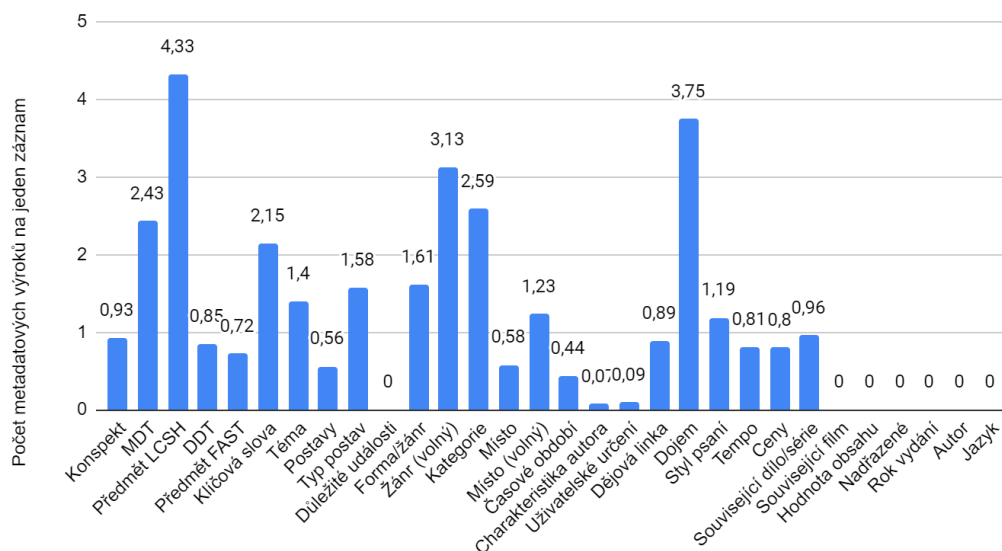


Graf 10: Uživatelské metadatové výroky dle typů

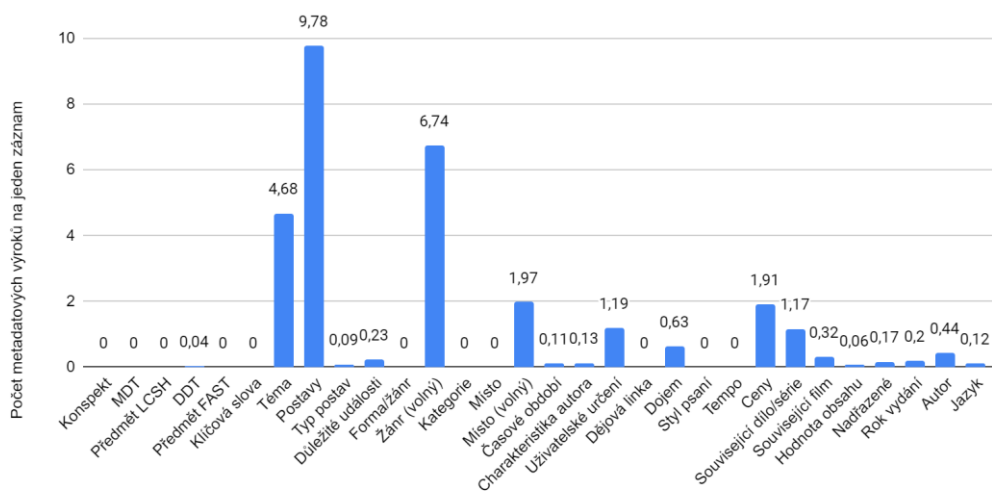


Abychom porovnali ještě i čísla, která se porovnat dají, uvádíme opět stejné dva grafy, ale tentokrát s koeficienty, které znázorňují využívanost daného typu metadatových výroků v daném typu databázi (viz kapitola 6.2).

Graf 11: Využívanost typů metadatových výroků – profesionální metadata



Graf 12: Využívanost typů metadatových výroků – uživatelská metadata



Závěr

Rozdíly mezi typy metadatových výroků profesionálních a uživatelských jsou obrovské. Jediný typ metadatového údaje, v jehož častém používání se profesionální a uživatelská shodují, je žánr (volný). V profesionálních databázích je tento typ ale zpravidla generovaný z externích služeb. Naopak u uživatelských metadat je velmi výrazné používání metadatových výroků typu postavy, které se

v profesionálních metadatech nachází jen málo. Z toho je patrný velký zájem uživatelů o tyto údaje a v menší míře také o téma, uživatelské určení a ceny dané knihy. Knihovny a knihkupectví by se těmito informacím v popisu beletrie měly věnovat více. U jiných informačních zdrojů než beletristických knih by už takovou váhu tyto údaje s největší pravděpodobností neměly.

7.3 Z jaké části jsou uživatelská metadata objektivní? Liší se to mezi jejich formami?

Průměrně jsou uživatelská metadata objektivní z 61 %. Metadata typu tagy jsou objektivní z 81 %, zatímco poličky z 28 %. Z toho plyne, že rozdíly mezi formami metadatových výroků jsou značné, tagy jsou o 53 % objektivnější než poličky.

Závěr

Uživatelská metadata jsou objektivní více než z poloviny. To je podle nás výborný výsledek, který by mohl ilustrovat skutečnou hodnotu uživatelských metadat. Přestože existuje velké množství motivací k jejich tvorbě (srovnej i s typy tagů v kapitole 2.4), uživatelé přidávají převážně objektivní informace. Z toho je také zřejmé, že pro ně jsou takové informace důležité, když knihy vyhledávají.

Velmi překvapivý je pro nás také rozdíl mezi formami uživatelských metadat, tagy jsou o 50 % objektivnější než poličky. Z toho vyvozujeme, že v analyzovaných čtenářských portálech mají odlišnou funkci: tagy slouží spíše k objektivnímu popisu, zatímco poličky k osobní organizaci knih.

7.4 Z jaké části se profesionální a uživatelská metadata shodují?

V našem výzkumu jsme zjišťovali, kolik hodnot objektivních uživatelských metadatových výroků průměrně sémanticky odpovídá hodnotám metadatových výroků profesionálních. Dospěli jsme k výsledku 62 %. To odpovídá přibližně 38 %, pokud bychom brali v úvahu všechna uživatelská metadata, objektivní i subjektivní dohromady. Pokud by nás zajímaly absolutní hodnoty čísel, u jedné knihy odpovídalo průměrně 57 uživatelských metadatových výroků, nicméně absolutní počet se u jednotlivých knih velmi lišil, čísla se pohybují od 20 po 125 uživatelských metadatových výroků odpovídajících profesionálním.

Přibližně každá třetí hodnota, kterou uživatelé vytvořili, byla sémanticky shodná s profesionálními hodnotami metadatových výroků u stejné knihy.

Závěr

Více než jedna třetina uživatelských metadat sémanticky odpovídala metadatům profesionálním, z čehož vyplývá, že je zde velký prostor pro obohacování profesionálních metadatových záznamů. Jelikož uživatelská metadata ilustrují informace, které uživatelé u beletrie chtějí a očekávají, bylo by užitečné, kdyby se jejich typy a množství v budoucnosti více promítaly do knihovních a knihkupeckých katalogů, čímž by mělo dojít ke zlepšení uživatelské přívětivosti za předpokladu, že nedojde ke zvýšení komplikovanosti uživatelského rozhraní.

7.5 Existují uživatelská metadata, která jsou potenciálně užitečná a mohla by doplnit profesionální metadata?

Celkem bylo identifikováno 1012 hodnot metadatových výroků, které neodpovídaly profesionálním metadatům, jde průměrně o 37 metadatových výroků na jednu knihu, tedy 38 % z objektivních uživatelských metadat a 23 % ze všech uživatelských metadat. V tomto počtu jsou započítány některé hodnoty vícekrát, je to prostý součet buněk, a také jsou tam započítány hodnoty, které se v profesionálních metadatových záznamech nacházely, pouze nebyly sbírány, jelikož nejde o obsahová metadata. Rozložení typů metadatových výroků, ze kterých tyto hodnoty pocházejí, naleznete v kapitole 6.1.2, nejčastějším typem byly postavy a potom téma.

Seznam všech hodnot, které neodpovídaly profesionálním metadatům, naleznete v příloze 10. Metadatové výroky v příloze jsou rozděleny na dvě části, podle toho, zda pocházejí z čistě obsahových metadatových výroků, které přinášejí nové informace (695 hodnot), nebo z jiných typů popisných metadatových výroků, které se v knihovních a/nebo knihkupeckých katalozích pravděpodobně nacházely (20 hodnot). Jsou řazeny abecedně (bez ohledu na jazyk výroků) a pokud by se některé výroky přesně opakovaly, jsou uvedeny pouze jednou. Proto také součet neodpovídá číslu 1012. Mezi jazyky výroky deduplikovány nejsou, abychom ilustrovali skutečný výskyt metadatových výroků v databázích. V tuto chvíli nás

nezajímá, ke které knize byly tyto metadatové výroky přiřazeny, tuto informaci lze dohledat v příloze 4.

Pokud bychom chtěli brát v úvahu pouze ty metadatové výroky, které byly ve čtenářských portálech skutečně unikátní, zůstane nám 695 hodnot uživatelských metadatových výroků, které přinášely oproti profesionálním metadatovým výročkům zcela novou informaci. To je průměrně 25,7 hodnoty na jednu knihu.

Závěr

Existuje velké množství uživatelských metadat, která by mohla obohatit profesionální popis, konkrétně přibližně 25,7 hodnot na jednu knihu. Na základě tohoto výzkumu nemůžeme zcela objektivně tvrdit, že profesionální metadata nejsou dostačující, ale víme, že uživatelé sami pro sebe vytvářejí ještě jiné metadatové výroky, které by tak mohly sloužit jako podněty k obohacování profesionálních metadatových záznamů. Je to opravdu velký počet informací, které uživatelé chtějí znát a profesionálové jim je nemohou nebo nechtějí předávat.

7.6 Kdo při popisu informačních zdrojů tvoří složitější koncepty, profesionálové nebo uživatelé?

Díky rozdělení metadatových výroků na jednoduché a složené v rámci mapování na Thema můžeme určit, který typ tvůrců metadat tvoří termíny, které se skládají z více myšlenkových částí. V následující tabulce najdete procenta metadatových výroků, která tvůrci vytvořili jako jednoduchá, složená ze dvou, tří a čtyř pojmů. (Procenta jsou zaokrouhlená na celá čísla.)

Tabulka 22: Složitost metadatových hodnot podle tvůrců

	Profesionálové	Uživatelé
Jednoduché metadatové výroky	68 %	72 %
Metadatové výroky složené ze dvou	25 %	21 %
Metadatové výroky složené ze tří	7 %	5 %
Metadatové výroky složené ze čtyř	0 %	2 %

Vidíme, že uživatelé častěji tvořili metadatové výroky jednoduché, a to o 4 %. metadatové výroky složené ze dvou a tři pojmů tvořili častěji profesionálové, ale ze čtyř pojmů opět uživatelé.

Závěr

Pokud bychom měli zodpovědět výzkumnou otázku, řekli bychom, že složitější koncepty tvoří profesionálové s výjimkou velmi složitých konceptů, které tvoří častěji uživatelé. Můžeme se jen dohadovat, co je důvodem, proč profesionálové téměř netvoří metadatové výroky složené ze čtyř myšlenkových částí – zda to řízené slovníky neumožňují, či zda se zdají být již příliš složitá, aby mohla být pro uživatele knihovnických katalogů užitečná.

7.7 Jaká je míra sémantické interoperability uživatelských a profesionálních metadat?

Interoperabilitu definujeme jako vlastnost systému, která zajišťuje jeho spolupráci s dalšími systémy bez jakýchkoliv omezení (GDT Interop, 2020). Pro zjištění míry sémantické operability bylo využito mapování metadatových výroků, resp. jejich hodnot, na Thema. Jelikož tento standard slouží k propojování různých světů v oblasti knižního trhu, i zde ho použijeme k propojování různých světů, profesionálních a uživatelských metadat. Pro profesionální hodnoty platí, že odpovídají Thema z 51 %, zatímco hodnoty uživatelské pouze ze 37 %.

Zajímavé ale je, že u hodnot jednoduchých a složených ze dvou částí procentuální část namapovaných hodnot přibližně odpovídá těmto procentům (u profesionálních hodnot 51 % a 49 %, u uživatelských 39 % a 43 %), zatímco u hodnot složených ze tří částí se u obou skupin velmi liší. U hodnot profesionálních odpovídá Thema 65 %, což je o 14 % více než celkových 51 %. V případě hodnot uživatelských odpovídá Thema 12 %, což je o 25 % méně než celkových 37 %.

Závěr

Určení sémantické interoperability je obtížný proces, v tomto případě jsme zvolili způsob, který vyžaduje nástroj, který funguje jako „prostředník“ mezi dvěma odlišnými systémy, jejichž interoperabilitu chceme zjistit. Profesionální a uživatelská metadata jsme namapovali na předmětové kategorie standardu

Thema, což nám umožnilo udělat si představu o jejich sémantické interoperabilitě. Samozřejmě sám proces mapování vždy z určité části závisí na úsudku člověka, který ho provádí, proto nemůžeme tvrdit, že naše výsledky jsou v tomto směru všeobecně platné. Pokud by výzkum prováděla jiná osoba nebo by byl zvolen jiný způsob zjištění sémantické interoperability, výsledky by se mohly lišit.

Zjistili jsme, že míra sémantické interoperability uživatelských metadat je přibližně 37 %. Zajímavé je srovnání s výsledky čtvrté výzkumné otázky, jejíž vyhodnocení říká, že sémantická shoda uživatelských metadat s profesionálními je 35 %. To nám říká, že míra připravenosti uživatelských metadat ke spolupráci s dalšími systémy odpovídá jejich reálné shodě s profesionálními metadaty.

Míra sémantické interoperability profesionálních metadat je přibližně 51 %. Je tedy o 14 % vyšší než u metadat uživatelských. Očekávali bychom větší rozdíl mezi oběma typy metadat, z čehož vyplývá, že uživatelská metadata jsou mnohem více sémanticky interoperabilní, než je obecně předpokládáno. To má jistě velmi příznivý vliv na jejich možné budoucí využití v rámci profesionálních systémů.

8. Diskuse

8.1 Porovnání s dosavadními výzkumy

V kapitole 2 jsme uvedli některé výzkumy problematiky obsahových metadat. Nyní nás zajímá, zda výsledky našeho výzkumu odpovídají výzkumům dřívějším, nebo zda se v něčem odlišují. Přímou porovnat můžeme výsledky se třemi výzkumy, které se zabývaly nejpodobnější problematikou.

Šauperlová zjistila, že nejčastěji používanými tagy jsou: jména postav, geografické názvy, časové období, ve kterém se kniha odehrává, a literární žánr. Také uživatelé zmiňují ocenění, které kniha získala, svůj názor na román a propojují je s ostatními díly v sérii nebo s jinak příbuznými knihami nebo filmy. (Šauperl, 2012) My jsme jako nejčastější zjistili tyto uživatelské metadatové výroky: postavy, téma, žánr, místo, dojem, uživatelské určení, ceny, související dílo/série, související film, autor, důležité události, rok vydání, charakteristika autora, nadřazené, jazyk, časové období, hodnota obsahu. Všechny typy zmiňované Šauperlovou odpovídají. Nevíme, jak měla definovaný „názor“, pravděpodobně se jedná o subjektivní metadatové výroky, případně dojem. Navíc jsme definovali metadatové výroky typu téma, uživatelské určení, důležité události, rok vydání, charakteristiku autora, nadřazené a jazyk.

Výzkum Lawsonové se zabýval tím, zda existují objektivní uživatelské metadatové výroky, která neodpovídají profesionálním metadatovým výroky. Zjistila, že existuje dostatek objektivních tagů, pro které existují odpovídající hesla v LCSH (a jsou tedy objektivní), ale které v profesionálním popisu použity nebyly. (Lawson, 2009) Měli jsme jinak nastavená kritéria pro určení objektivnosti metadatových výroků, nicméně oba výzkumy došly k obdobným závěrům. Je zřejmé, že uživatelská metadata nabízejí něco, co se zatím profesionálním katalogizátorům nepodařilo zachytit.

Whiteová se ve svém výzkumu zabývá sice metadaty v repozitářích, nicméně došla k závěru, že přibližně čtvrtina metadatových výroků byla vložena stejně jak profesionály, tak vědci (tedy neprofesionály, v zásadě se jedná o uživatelské metadatové výroky s tím rozdílem, že zde dokumenty popisovali přímo jejich

autoři). Rozdíly v tvorbě metadat mezi těmito dvěma skupinami byly velké. (White, 2012) V našem výzkumu jsme došli k závěru, že 62 % objektivních metadatových výroků odpovídá profesionálnímu popisu, což odpovídá přibližně 38 % všech uživatelských metadatových výroků (objektivních i subjektivních dohromady). To je o 10 % více než v případě výzkumu Whiteové, nicméně nevíme přesně, jak přísná pravidla měla pro posuzování, zda uživatelské metadatové výroky odpovídají profesionálním.

8.2 Poznatky mimo výzkumné otázky

Dále bychom v tomto bodě ještě rádi uvedli některé skutečnosti, které sice nebyly předmětem výzkumu, nicméně by mohly sloužit pro doplnění kontextu výzkumu.

Zaznamenali jsme v knihovních katalozích anglicky mluvících zemí, že ve svých záznamech často využívají data generovaná z přidaných, externích služeb. Šlo především o službu NoveList, která vytváří obohacení pro knihovní katalogy za účelem pomoci najít čtenářům jejich další oblíbenou knihu. Tuto službu provozuje komerční firma EBSCO Industries, Inc. (About NoveList, 2020) NoveList spolupracuje při obohacování katalogů s dalšími partnery jako jsou Baker & Taylor nebo Goodreads. Některé katalogy byly přímo napojeny na Goodreads nebo LibraryThing, ze kterých přebíraly uživatelské recenze a/nebo doporučení podobných titulů. Některé knihovny si podobný obsah vytváří vlastní, např. British Library. Svůj obsah vytváří také komunita knihoven využívajících systém BiblioCommons, data od uživatelů jsou pak sdílená napříč sítí knihoven. Jde o zajímavou tematiku, neboť externí data ve svých záznamech zpřístupňuje knihovna, která ale není jejich původcem. V některých případech může dojít k zastření provenience dat směrem k uživateli, pokud není jasně uvedeno, že knihovna není tvůrcem těchto dat, ale pouze je přebírá. Nicméně obliba tohoto generovaného obsahu je mezi knihovnami anglicky mluvících zemí značná, z čehož je zřejmé, že stoupá zájem uživatelů i o další informace, než které se nacházejí v klasickém bibliografickém záznamu, jako jsou např. uživatelské i profesionální recenze, anotace, doporučení dalších knih, autorské poznámky a další.

Některé knihovny vytvářejí vlastní pole se stejným obsahem jako existující pole formátu MARC. Příkladem může být např. pole REJ v katalogu knihovny Bedřicha Beneše Buchlovana, které obsahuje stejný obsah jako pole 655, žánr/forma.

Několik anglických katalogů knihkupectví má zřejmě zcela shodnou databázi, byla zde stoprocentní shoda obsahových metadatových výroků. Jedná se o databáze Book People, Books & Books, Politics & Prose a Prairie Lights. Ve výzkumu byly všechny tyto databáze ponechány, protože se snažíme reprezentovat realitu.

Zastavíme se také u použití DDT v anglických knihovních katalozích. Ve formátu MARC se v podpoli \$2 zpravidla nachází deklarované vydání DDT, které bylo použito při katalogizaci. V různých knihovnách byly však u jedné knihy různé kódy DDT i přes to, že deklarovaly stejnou verzi. Je možné, že katalogizátoři používali starší verzi DDT, než je deklarovaná v záznamu, případně přidělovali notace „po paměti“. Patrné je to na rozdílu notací 813.54 a 813.6, které byly nalezeny pro některé knihy obě, přestože by se měly vylučovat.

8.3 Zpětné hodnocení výzkumu

Jedná se o kvalitativně-kvantitativní srovnávací analýzu. Kvalitativní význam mají původní sebraná data jako hloubková sonda do profesionálních a uživatelských obsahových metadatových výroků. Mnoho zajímavých informací se dá vyčíst již z kompletních tabulek v příloze 4, např. konkrétní formulace tagů a poliček. Kvantitativní význam má především následná analýza, kde dochází k vypočítání průměrných hodnot vybraných hodnocených parametrů.

Časová náročnost sběru dat a analýzy dat pro jednotlivé knihy byla zhodnocena jako 6,5 hodiny času na jednu knihu, celkem bylo analyzováno 27 knih, tj. 175,5 hodiny celkem. K tomu je nutné připočítat přibližně 4 hodiny času na závěrečnou analýzu.

Samozřejmě jsme se během výzkumu potýkali i s několika problémovými oblastmi. Při sběru dat hrálo velkou roli, že některé knihovny nezveřejňují volně záznamy ve formátu MARC, ale pouze v zobrazení pro čtenáře, přestože je zřejmé, že interně jsou na MARC založeny. Z toho důvodu bylo nutné vymyslet opatření pro sběr dat z takových katalogů, zejména pro pole *Subject*, ve kterém mohou být data z více

různých polí MARC (naleznete ho v kapitole 5.5.3). Je tak možné, že některé metadatové výroky byly zpracovány odlišně, než kdybychom měli formát MARC k dispozici. Této situaci se ale nedalo vyhnout, aniž bychom se museli uchýlit k tipování.

V Library of Trinity College Dublin nemají volně dostupné ani některé záznamy knih. Rozlišují typy materiálu „printed books“, jejichž záznamy jsou dostupné volně, a „books“, pro přístup k jejich záznamům je potřeba přihlášení. Z tohoto důvodu je knih nalezených v této databázi méně než u ostatních databází.

Další problémová oblast byla způsobena pravidly pro výběr knih, která původně nepočítala s tím, že se bude jednat o knihy v originálním jazyce jiném než angličtina. Tak došlo k tomu, že v průběhu výzkumu bylo nutné vyřadit dvě knihy, které nebyly vydány v angličtině v dostatečném časovém odstupu. Šlo o knihy od Larse Keplera (Playground a Lovce králíků), které se umístily v Českém bestselleru, a proto byly do výzkumu původně zařazeny.

Při hodnocení objektivitu nebo subjektivitu uživatelských metadatových výroků bylo zjištěno, že ne vždy je jasné, co uživatel daným termínem myslel. Příkladem je tag „fonetikum“, který se opakoval u více analyzovaných knih.

Chtěli jsme také ověřovat, jak se liší uživatelské metadatové výroky v češtině a angličtině, nicméně v průběhu výzkumu bylo zjištěno, že poměr metadatových výroků obou jazyků to neumožňuje. Metadatových výroků v angličtině je o tolik více, že jakékoli výsledky by touto skutečností musely být nutně ovlivněny. Dva analyzované čtenářské portály používají a metadata v angličtině přidávají lidé z celého světa, zatímco české portály mají nepoměrně menší uživatelskou základnu. Z tohoto důvodu se domníváme, že jakékoli porovnávání uživatelských metadat v češtině a v angličtině by vyžadovalo zcela jinak připravenou metodiku, která by tento rozdíl mezi jazyky zohledňovala.

A závěrečným úskalím výzkumu byl jeho zápis do textové podoby. Nebylo zcela jasné, kdy používat označení „metadata“ a kdy „data“, jelikož původní metadata se v průběhu výzkumu stávají sebranými daty a mají svá vlastní metadata.

Předložený výzkum má také své limity. Za hlavní limit považujeme jazykové hledisko, jelikož bylo stanovené hledisko pouze českého a anglického jazyka.

Předpokládáme, že knihovny z jiných jazykových oblastí by mohly k tomuto tématu také přispět, např. německá nebo francouzská národní knihovna.

Dalším limitem je nepochybně relativně malé množství analyzovaných knih, proto lze výsledky generalizovat jen do určité míry. Vzhledem k tomu, že byly vybrány knihy hodně kupované, předpokládá se, že výsledky by se u jiných, méně známých knih, lišily.

8.4 Možnosti dalšího zkoumání

Při dalším zkoumání by bylo určitě vhodné ověřit výsledky výzkumu i v jiných jazykových oblastech a na větším množství knih. Tím by se možnost výsledky generalizovat na všechny knihy obecně výrazně zvýšila.

Také je pro další výzkum určitě vhodná i oblast generování dat pro knihovní katalogy z externích služeb a jejich spolupráce se čtenářskými portály na zajištění co nejlepšího komfortu pro uživatele-čtenáře. Bylo by jistě zajímavé zjistit, proč knihovny tyto služby do svých katalogů zařazují, jaký názor na to mají uživatelé, a jestli se něco podobného chystá i v českém prostředí, případně proč ano/ne.

V některých knihovních katalozích nebo katalozích knihkupectví se rovněž nachází uživatelská metadata, která v tomto výzkumu nebyla hodnocena. Mohlo by být zajímavé zjistit, jestli odpovídají uživatelským metadatům ve čtenářských portálech.

Další oblastí vhodnou k výzkumu by byla nepochybně i problematika řešení problémů přirozeného jazyka při tvorbě uživatelských metadat. V našem výzkumu se často vyskytovaly uživatelské metadatové výroky se stejným významem, ale rozdílným zápisem, např. „abandoned“ a „dnf“. V analyzovaných čtenářských portálech nebyla tato problematika řešena, proto by bylo užitečné zjistit, jakým způsobem se řeší v jiných než knižních uživatelských komunitách.

Také by se dalo ověřit, zda existují podobné vzorce tvorby obsahových metadat i pro jiné dokumenty než tištěné knihy, např. elektronické knihy nebo zcela jiné dokumenty, obrázky nebo hudbu.

9. Závěr

Závěrem si dovolíme shrnout zjištěné poznatky a připojit stručné zamyšlení o budoucnosti. Cílem této práce bylo porovnání profesionálních obsahových metadat s metadaty uživatelskými. Výzkum však probíhal pouze v česko-anglickém jazykovém prostředí a na omezeném počtu knih, není tak možné jeho výsledky generalizovat na všechny knihy obecně.

Byly zjištěny velké rozdíly jak na úrovni typů metadat, které databáze umožňují zpřístupňovat, tak na úrovni konkrétní využívanosti těchto typů. Za nejzajímavější informaci považujeme, že uživatelská metadata jsou objektivní z 61 %, což převyšuje jejich hodnocení v obecném povědomí. Zároveň mezi těmito objektivními metadaty existuje 38 % uživatelských metadatových výroků, které mohou potenciálně přinášet užitečné informace o knihách a v profesionálních metadatových záznamech se nevyskytují. Celkem jsme analyzovali 4395 uživatelských metadatových výroků a 1012 z nich přinášelo nové informace o dané knize oproti profesionálním metadatům. Toto číslo bychom se nebáli nazvat až alarmujícím. Z výše zmíněných výsledků usuzujeme na důležitou roli uživatelských metadat pro budoucnost metadat obecně a předpokládáme, že jejich význam bude ještě stoupat.

Podle našeho názoru mají knihovny a knihkupectví dvě hlavní možnosti, jak se se současnou situací vyrovnat: buď vylepší obsah a uživatelskou přívětivost svých katalogů, nebo navážou spolupráci se čtenářskými portály nebo dalšími službami, čímž se uživatelům přiblíží a přinesou jim přesně ty informace, které uživatelé chtějí. Samozřejmě v tom budou důležitou roli hrát také finance.

Otázkou současnosti a blízké budoucnosti pak jsou konkrétní pokusy o kombinaci profesionálních a uživatelských metadat při vyhledávání dokumentů. Z dlouhodobějšího hlediska považujeme kombinaci profesionálních a uživatelských metadat nejen za možnou, ale za prospěšnou pro obě strany. Pro uživatele je důležité zejména snadnější vyhledávání a pro knihovny (nebo informační instituce obecně) snížení nákladů na indexaci informačních zdrojů, ke kterému může dojít díky citlivému navýšení podílu uživatelských metadat.

Použitá literatura

About NoveList: We Transform Lives Through Reading. In: *NoveList* [online]. Ipswich (Massachusetts): EBSCO Information Services, c2020 [cit. 2020-04-25]. Dostupné z: <https://www.ebscohost.com/novelist/about-novelist/about-novelist>.

Aboutness, n. In: *Oxford English Dictionary* [online]. Oxford: Oxford University Press, 2009 [cit. 2020-04-25]. Dostupné z: <https://www.oed.com/view/Entry/247514?redirectedFrom=aboutness>.

Adresáře knihoven v ČR. In: *Souborný katalog České republiky* [online]. Praha: Národní knihovna ČR, 2016 [cit. 2020-04-25]. Dostupné z: <https://www.caslin.cz/caslin/odkazy/adresare-knihoven/odkazy-na-jine-adresare>.

AHMED, K., D. AYERS, M. BIRBECK, J. COUSINS, D. DODDS, J. LUBELL, M. NIC, D. RIVERS-MOORE, A. WATT, R. WORDEN, A. WRIGHTSON. *Professional XML Meta Data*. Birmingham (Alabama): Wrox Press, 2001.

AL-KHALIFA, Hend S. a Hugh C. DAVIS. Towards better understanding of folksonomic patterns. In: *Proceedings of the 18th conference on Hypertext and hypermedia - HT '07* [online]. New York: Association for Computing Machinery, 2007 [cit. 2020-04-25]. ISBN 978-1-59593-820-6. Dostupné z (DOI): <https://doi.org/10.1145/1286240.1286288>.

ALSARHAN, Hesham. *The effects of target audience on social tagging* [online]. Bloomington, 2013 [cit. 2020-04-25]. ISBN 978-1-303-53486-7. Komerčně dostupné z: <https://search.proquest.com/docview/1468434380>. Disertační práce. Department of Information and Library Science, Indiana University.

AMES, Morgan a Mor NAAMAN. Why we tag: motivations for annotation in mobile and online media. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07* [online]. New York: Association for Computing Machinery, 2007 [cit. 2020-04-25]. ISBN 978-1-59593-593-9. Dostupné z (DOI): <https://doi.org/10.1145/1240624.1240772>.

ANGEL, Christine Marie. A Comparison of Descriptive Tagging Practices by Library, Archive and Museum Professionals Using an Inter-Indexing Consistency

Approach [online]. Columbia (Jižní Karolína), 2012 [cit. 2020-04-25]. ISBN 978-1-2673-6577-4. Dostupné z: <https://search.proquest.com/docview/1020132307>.
Disertační práce. University of South Carolina.

AUZKÁ, Vendula. *Fanfikce - internetový fenomén literární tvorby* [online]. Plzeň, 2016 [cit. 2020-04-25]. Dostupné z: <http://hdl.handle.net/11025/24559>.
Bakalářská práce. Západočeská univerzita v Plzni. Vedoucí práce Jiří Staněk.

BALÍKOVÁ, Marie. Katalog. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2020-04-25]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000001443&local_base=KTD.

BEARDSLEY, Richard. Outsourcing and the Role of Strategic Alliances. *Publishers Weekly* [online]. New York: PWxyz, 2016, **263**(51), s. 6-7 [cit. 2020-04-25]. ISSN 0000-0019. Dostupné z: <https://www.publishersweekly.com/pw/by-topic/industry-news/manufacturing/article/72249-outsourcing-and-the-role-of-strategic-alliances.html>.

BENOIT, Edward a Amanda L. MUNSON. Proceed with Caution: Deepening Practitioner Concerns about Social Tagging within Digital Collections. *Portal: Libraries and the Academy* [online]. Baltimore (Maryland): Johns Hopkins University Press, 2018, **18**(4), s. 759-779 [cit. 2020-04-25]. DOI: 10.1353/pla.2018.0044. ISSN 1530-7131. Komerčně dostupné z: <https://muse.jhu.edu/article/704941>.

BENZ, Dominik a Andreas HOTH. Position Paper: Ontology Learning from Folksonomies. In: HINNEBURG, Alexander, ed. *Lernen - Wissen – Adaption: Workshop Proceedings LWA 2007*, Halle, September 2007, Martin Luther University Halle-Wittenberg, Institute for Informatics, Databases and Information Systems [online]. Halle-Wittenberg: Martin Luther University, 2007 [cit. 2020-04-25]. Dostupné z: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/pub/pdf/benz2007position.pdf>.

Bestselling Books of 2015. In: *Publishers Weekly* [online]. New York: PWxyz, 2016 [cit. 2020-04-25]. Dostupné z: <https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/69054-bestselling-books-of-2015.html>.

Bestselling Books of 2016. In: *Publishers Weekly* [online]. New York: PWxyz, 2017 [cit. 2020-04-25]. Dostupné z: <https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/72566-the-best-selling-books-of-2016.html>.

Bestselling Books of 2017. In: *Publishers Weekly* [online]. New York: PWxyz, 2018 [cit. 2020-04-25]. Dostupné z: <https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/75817-the-best-sellers-of-2017.html>.

BIC Standard Subject Categories. In: *Book Industry Communication* [online]. London: Book Industry Communication, c2020 [cit. 2020-04-25]. Dostupné z: <https://bic.org.uk/7/BIC-Standard-Subject-Categories>.

BOGERS, Toine a Vivien PETRAS. Tagging vs. Controlled Vocabulary: Which is More Helpful for Book Search? In: *IConference 2015 Proceedings* [online]. 2015 [cit. 2020-04-25]. Dostupné z: <http://hdl.handle.net/2142/73673>.

Book of the Month [online]. Book of the Month, c2020 [cit. 2020-04-25]. Dostupné z: <https://www.bookofthemonth.com/>.

BROUGHTON, Vanda. 8 - Content analysis 1: document description. In: *Essential classification* [online]. Second edition. London: Facet Publishing, 2015 [cit. 2020-04-25]. ISBN 978-1-78330-031-0. Komerčně dostupné z (DOI): <https://doi.org/10.29085/9781783302383.008>.

BUNDZA, Maira. The Choice Is Yours! Researchers Assign Subject Metadata to Their Own Materials in Institutional Repositories. *Cataloging & Classification Quarterly* [online]. New York: Routledge, 2014, **52**(1), s. 110-118 [cit. 2020-04-25]. ISSN 0163-9374. Komerčně dostupné z (DOI): <https://doi.org/10.1080/01639374.2013.852439>.

České MDT Online [online]. Haag: UDC Consortium, c2020 [cit. 2020-04-25]. Dostupné z: <http://cz.udc-hub.com/cs/login.php>.

ČESKO. Zákon č. 257/2001 Sb., o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon). In: *Sbírka zákonů České republiky*. 29. 6. 2001. ISSN 1211-1244. Dostupný také z: https://ipk.nkp.cz/legislativa/01_LegPod/knihovni-zakon-257-2001-sb.-a-

[navazne-provadeci-prepisy/Zakon257.htm](#). Znění se zapracováním všech novel přijatých do července 2017.

DEMPSEY, Lorcan a Rachel M. HEERY. Metadata: a current view of practice and issues. *Journal of Documentation* [online]. Bradford (Anglie): Emerald Publishing, 1998, **54**(2), s. 145-172 [cit. 2020-04-25]. ISSN 0022-0418. Komerčně dostupné z (DOI): <https://doi.org/10.1108/EUM0000000007164>.

DENNEN, Vanessa P., Lauren M. BAGDY a Michelle L. CATES. Effective Tagging Practices for Online Learning Environments: An Exploratory Study of Approach and Accuracy. *Online Learning* [online]. Newburyport (Maryland): Online Learning Consortium, 2018, **22**(3), s. 103-120 [cit. 2020-04-25]. ISSN 2472-5749. Dostupné z: <https://eric.ed.gov/?id=EJ1191492>.

DEWEY, Melvil. *Dewey decimal classification and relative index: Volume 1* [online]. Edition 22. Dublin (Ohio): Online Computer Library Center, 2003a [cit. 2020-04-25]. Dostupné z: https://archive.org/details/deweydecimalclas01dewe_0.

DEWEY, Melvil. *Dewey decimal classification and relative index: Volume 2* [online]. Edition 22. Dublin (Ohio): Online Computer Library Center, 2003b [cit. 2020-04-25]. Dostupné z: https://archive.org/details/deweydecimalclas02dewe_0.

DEWEY, Melvil. *Dewey decimal classification and relative index: Volume 3* [online]. Edition 22. Dublin (Ohio): Online Computer Library Center, 2003c [cit. 2020-04-25]. Dostupné z: https://archive.org/details/deweydecimalclas03dewe_0.

Dlouhý chvost. In: *Wikipedie: otevřená encyklopedie* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2019-06-19 [cit. 2020-04-25]. Dostupné z: https://cs.wikipedia.org/wiki/Dlouh%C3%BD_chvost.

DUBINKO, Micah, Ravi KUMAR, Joseph MAGNANI, Jasmine NOVAK, Prabhakar RAGHAVAN a Andrew TOMKINS. Visualizing tags over time. In: *Proceedings of the 15th international conference on World Wide Web - WWW '06*

[online]. New York: Association for Computing Machinery, 2006 [cit. 2020-04-25]. ISBN 978-1-59593-323-2. Dostupné z (DOI): <https://doi.org/10.1145/1135777.1135810>.

Dublin Core™ Metadata Element Set, Version 1.1: Reference Description. In: *Dublin Core Metadata Initiative* [online]. Dublin Core Metadata Initiative, 2012 [cit. 2020-04-25]. Dostupné z: <https://www.dublincore.org/specifications/dublin-core/dces/>.

DUFF, Wendy a Sue MCKEMMISH. Metadata and ISO 9000 Compliance. *Information Management Journal* [online]. Overland Park (Kansas): A R M A International, 2000, **34**(1), s. 4-16 [cit. 2020-04-25]. ISSN 1535-2897. Komerčně dostupné z: <http://search.ebscohost.com/>.

Freedom of Information Act 2000. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-04-15 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/Freedom_of_Information_Act_2000.

FURNER, Jonathan. Folksonomies. In: BATES, Marcia J. a Mary NILES MAACK, eds. *Encyclopedia of Library and Information Sciences* [online]. 3rd ed. Boca Raton (Florida): CRC Press, 2010 [cit. 2020-04-25]. ISBN 978-0-203-75763-5. Komerčně dostupné z (DOI): <https://doi.org/10.1201/9780203757635>.

GDT INTEROP, John MCCREESH a Ed DANIEL. Definition of Interoperability. In: *Définition de l'Interopérabilité* [online]. Paris: Association Francophone des Utilisateurs de Logiciels Libres, c2020 [cit. 2020-04-25]. Dostupné z: <http://interoperability-definition.info/en/>.

GILLIAND-SWETLAND, A.. Defining Metadata. In: BACA, M., ed. *Introduction to Metadata: pathways to digital information*. Los Angeles (Kalifornie): Getty Information Institute, 1998.

GILLILAND, Anne J. Setting the Stage. In: *The Getty Research Institute* [online]. Los Angeles (Kalifornie): Getty, 2008 [cit. 2020-04-25]. Dostupné z: https://web.archive.org/web/20090422033023/http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html.

GOLDER, Scott A. a Bernardo A. HUBERMAN. Usage patterns of collaborative tagging systems. *Journal of Information Science* [online]. London: Sage Publications, 2006, **32**(2), s. 198-208 [cit. 2020-04-25]. ISSN 0165-5515. Komerčně dostupné z (DOI): <https://doi.org/10.1177/0165551506062337>.

GOLUB, Koraljka. Potential and Challenges of Subject Access in Libraries Today on the Example of Swedish Libraries. *International Information & Library Review* [online]. London: Taylor & Francis Group, 2016, **48**(3), s. 204-210 [cit. 2020-04-25]. ISSN 1057-2317. Komerčně dostupné z: <https://doi.org/10.1080/10572317.2016.1205406>.

GRZEGORZ, Gmiterek. Sieciowe sposoby społecznościowego zarządzania informacją o dokumentach. *ZIN – Zagadnienia Informacji Naukowej. Studia Informacyjne* [online]. Varšava: Stowarzyszenie Bibliotekarzy Polskich i Wydział Dziennikarstwa, Informacji i Bibliologii. Uniwersytet Warszawski, 2016, **54**(1), s. 108-133 [cit. 2020-04-25]. ISSN 0324-8194. Dostupné z: http://193.0.122.118/wp-content/uploads/SBP/Zin_2016_01.pdf.

GUPTA, Manish, Rui LI, Zhijun YIN a Jiawei HAN. An Overview of Social Tagging and Applications. In: *Social Network Data Analytics* [online]. Boston (Massachusetts): Springer US, 2011, s. 447-497 [cit. 2020-04-25]. ISBN 978-1-4419-8461-6. Dostupné z (DOI): http://link.springer.com/10.1007/978-1-4419-8462-3_16.

HANRATH, Scott a Erik RADIO. User search terms and controlled subject vocabularies in an institutional repository. *Library Hi Tech* [online]. Bingley (Anglie): Emerald Group Publishing, 2017, **35**(3), s. 360-367 [cit. 2020-04-25]. DOI: 10.1108/LHT-11-2016-0133. ISSN 0737-8831. Komerčně dostupné z: <https://www.emerald.com/insight/content/doi/10.1108/LHT-11-2016-0133/full/html>.

HAYNES, David. Metadata for Information Management and Retrieval: Understanding metadata and its use. London: Facet Publishing, 2004.

HAYNES, David. *Metadata for Information Management and Retrieval: Understanding metadata and its use* [online]. Second edition. London: Facet

Publishing, 2018 [cit. 2020-04-25]. ISBN 978-1-78330-216-1. Komerčně dostupné z: <http://search.ebscohost.com>.

HEDDEN, Heather. Controlled vocabularies, thesauri, and taxonomies. *The Indexer* [online]. Liverpool: Liverpool University Press, 2008, **26**(1), s. 33–34 [cit. 2020-04-25]. ISSN 0019-4131. Komerčně dostupné z: <http://search.ebscohost.com>.

HILLMAN, Diane. Using Dublin Core. In: *Dublin Core Metadata Initiative* [online]. Dublin Core Metadata Initiative, 2005 [cit. 2020-04-25]. Dostupné z: <https://www.dublincore.org/specifications/dublin-core/usageguide/>.

Historie za rok 2015. In: *Český bestseller* [online]. Ostrava, 2015 [cit. 2020-04-25]. Dostupné z: <http://www.ceskybestseller.cz/products/2015/>.

Historie za rok 2016. In: *Český bestseller* [online]. Ostrava, 2016 [cit. 2020-04-25]. Dostupné z: <http://www.ceskybestseller.cz/products/2016>.

Historie za rok 2017. In: *Český bestseller* [online]. Ostrava, 2017 [cit. 2020-04-25]. Dostupné z: <http://www.ceskybestseller.cz/products/2017>.

HJØRLAND, Birger. Subject (of Documents). *Knowledge Organization* [online]. Würzburg: Ergon Verlag, 2017, **44**(1), s. 55-64 [cit. 2020-04-25]. ISSN 0943-7444. Komerčně dostupné z: <http://search.ebscohost.com>. Odvozeno z článku stejného názvu v ISKO Encyclopedia of Knowledge Organization, verze 1.0, publikováno 2016-10-04, dostupné z: <https://www.isko.org/cyclo/subject.htm>.

HUTCHENS, Chad. Open access metadata: current practices and proposed solutions. *Learned Publishing* [online]. Hoboken (New Jersey): Wiley, 2013, **26**(3), s. 159-165 [cit. 2020-04-25]. ISSN 0953-1513. Komerčně dostupné z (DOI): <https://doi.org/10.1087/20130302>.

HYPÉN, Kaisa a Eetu MÄKELÄ. An ideal model for an information system for fiction and its application. *Library Review* [online]. Bradford (Anglie): Emerald Publishing, 2011, **60**(4), s. 279-292 [cit. 2020-04-25]. ISSN 0024-2535. Komerčně dostupné z (DOI): <https://doi.org/10.1108/00242531111127848>.

CHAN, Lois Mai a Athena SALABA. *Cataloging and Classification: An Introduction* [online]. Fourth edition. Lanham: Rowman & Littlefield Publishers, 2016 [cit. 2020-04-25]. ISBN 978-1-4422-3250-1. Komerčně dostupné z: <http://search.ebscohost.com>.

CHOI, Kahyun, Jin Ha LEE, Xiao HU a J. Stephen DOWNIE. Music subject classification based on lyrics and user interpretations: Will open peer review transform scientific and scholarly publishing? *Proceedings of the Association for Information Science and Technology* [online]. Silver Spring (Maryland): Association for Information Science and Technology, 2016, **53**(1), s. 1-10 [cit. 2020-04-25]. ISSN 2373-9231. Dostupné z (DOI): <https://doi.org/10.1002/pr2.2016.14505301041>.

CHOWDHURY, G. G. a Sudatta CHOWDHURY. *Organizing Information: From the Shelf to the Web* [online]. London: Facet Publishing, 2007 [cit. 2020-04-25]. ISBN 978-1-85604-578-0. Komerčně dostupné z: <http://search.ebscohost.com>.

ISAAC, Antoine a Ed SUMMERS, ed.. SKOS Simple Knowledge Organization System Primer. In: *W3C* [online]. 2009 [cit. 2020-04-25]. Dostupné z: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>.

JOHNSON, Shannon Fay. Fan fiction metadata creation and utilization within fan fiction archives: Three primary models. *Transformative Works and Cultures* [online]. New York, Organization for Transformative Work, 2014, **17** [cit. 2020-04-25]. ISSN 1941-2258. Dostupné z (DOI): <https://doi.org/10.3983/twc.2014.0578>.

JOORABCHI, Arash, Michael ENGLISH a Abdulhussain E. MAHDI. Automatic mapping of user tags to Wikipedia concepts. *Journal of Information Science* [online]. Thousand Oaks (Kalifornie): SAGE Publishing, 2015, **41**(5), s. 570-583 [cit. 2020-04-25]. ISSN 0165-5515. Komerčně dostupné z (DOI): <https://doi.org/10.1177/0165551515586669>.

KROSKI, Ellyssa. The Hive Mind: Folksonomies and User-Based Tagging. In: *InfoTangle* [online]. 2005 [cit. 2020-04-25]. Dostupné z: <http://web20bp.com/13f1b6019/wp-content/uploads/2013/03/The-Hive-Mind-Folksonomies-2005.pdf>.

KUČEROVÁ, Helena. Databáze. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2020-04-25]. Dostupné z:

https://aleph.nkp.cz/F/?func=direct&doc_number=000000089&local_base=KTD.

KUČEROVÁ, Helena. Co analyzujeme při obsahové analýze dokumentů?

K pojmu aboutness v organizaci znalostí. *Knihovna* [online]. Praha: Knihovnický institut Národní knihovny České republiky, 2014, **25**(1), s. 36-54 [cit. 2020-04-25]. ISSN 1801-3252. Dostupné z:

<http://oldknihovna.nkp.cz/knihovna141/141036.htm>.

KUČEROVÁ, Helena. *Organizace znalostí : klíčová témata* [online]. Praha:

Karolinum Press, 2017 [cit. 2020-04-25]. ISBN 978-80-246-3597-2. Komerčně dostupné z: <http://search.ebscohost.com>.

KUČEROVÁ, Helena. Pojmový model bibliografických informací IFLA LRM. *Bulletin SKIP* [online]. Praha: Svaz knihovníků a informačních pracovníků ČR, 2018, **27**(2) [cit. 2020-04-25]. ISSN 1213-5828. Dostupné z:

<https://bulletinskip.skipcr.cz/node/403>.

LAWSON, Karen G. Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers. *The Journal of Academic Librarianship* [online]. Amsterdam: Elsevier, 2009, **35**(6), s. 574-582 [cit. 2020-04-25]. ISSN 0099-1333. Komerčně dostupné z (DOI): <https://doi.org/10.1016/j.acalib.2009.08.020>.

LICHTENBERGOVÁ, Edita, Marie BALÍKOVÁ, Ludmila BENEŠOVÁ, Jarmila PŘIBYLOVÁ a Jaroslava SVOBODOVÁ. Katalogizace podle RDA ve formátu MARC 21 - tištěné a elektronické monografie: katalogizace na úrovni minimálního/doporučeného záznamu [online]. Aktualizované vydání. Praha: Národní knihovna České republiky, 2016 [cit. 2020-04-25]. Dostupné z: <https://text.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/katalogizace-podle-rda-ve-formatu-marc-21-tistene-a-elektronicke-monografie-katalogizace-na-urovni-minimalniho-doporuceneho-zaznamu>.

LIN, Yi-Ling. *Enhancing Image Findability through a Dual-Perspective Navigation Framework* [online]. Pittsburgh: 2013 [cit. 2020-04-25]. ISBN 978-1-303-77980-0. Komerčně dostupné z:

<https://search.proquest.com/docview/1512424040>. Disertační práce. University of Pittsburgh.

List of cities in Australia by population. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-04-10 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_cities_in_Australia_by_population.

List of cities in New Zealand. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-02-12 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_cities_in_New_Zealand.

List of cities in the United Kingdom. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-04-14 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_cities_in_the_United_Kingdom.

List of largest Canadian cities by census. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-04-01 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_largest_Canadian_cities_by_census.

List of United States cities by population. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-04-15 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population.

List of urban areas in the Republic of Ireland by population. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-03-18 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/List_of_urban_areas_in_the_Republic_of_Ireland_by_population.

LUTZ, Ashley. 18 Bookstores Every Book Lover Must Visit At Least Once. In: *Business Insider* [online]. New York: Insider, 2014 [cit. 2020-04-25]. Dostupné z: <https://www.businessinsider.com/best-bookstores-in-the-world-2014-2014-2>.

MAI, Jens-Erik. Folksonomies and the New Order: Authority in the Digital Disorder. *Knowledge Organization* [online]. Würzburg: Ergon Verlag, 2011, **38**(2), s. 114-122 [cit. 2020-04-25]. ISSN 0943-7444. Komerčně dostupné z: <http://search.ebscohost.com>.

MANZO, Christina, Geoff KAUFMAN, Sukdith PUNJASTHITKUL a Mary FLANAGAN. "By the People, For the People": Assessing the Value of Crowdsourced, User-Generated Metadata. *Digital Humanities Quarterly* [online]. The Alliance of Digital Humanities Organizations, 2015, **9**(1) [cit. 2020-04-25]. ISSN 1938-4122. Dostupné z: <http://www.digitalhumanities.org/dhq/vol/9/1/000204/000204.html>.

MARC 21 Format for Bibliographic Data: 1999 Edition, Update No. 1 (October 2000) through Update No. 29 (November 2019). In: *Library of Congress* [online]. Washington, D.C.: Library of Congress, 2019 [cit. 2020-04-25]. Dostupné z: <https://www.loc.gov/marc/bibliographic/>.

MCCLUNG, Julie. Herding cats: indexing British Columbia's political debates using controlled vocabulary. *The Indexer* [online]. Liverpool: Liverpool University Press, 2009, **27**(2), s. 66-69 [cit. 2020-04-25]. ISSN 1756-0632. Dostupné z (DOI): <https://doi.org/10.3828/indexer.2009.22>.

MERTES, Kate. On aboutness. *The Indexer: The International Journal of Indexing* [online]. London: Liverpool University Press, 2017, **35**(2), s. 77-78 [cit. 2020-04-25]. DOI: 10.3828/indexer.2017.24. ISSN 0019-4131. Komerčně dostupné z: <http://search.ebscohost.com>.

Městská knihovna roku. In: *Svaz knihovníků a informačních pracovníků České republiky* [online]. Praha: Svaz knihovníků a informačních pracovníků České republiky (SKIP), 2010 [cit. 2020-04-25]. Dostupné z: <https://www.skipcr.cz/akce-a-projekty/akce-skip/cena-skip-a-skanska-mestska-knihovna-roku>.

Μετά. In: *Dict.com* [online]. Lingea, c2020 [cit. 2020-04-25]. Dostupné z: <https://www.dict.com/greek-english/meta>.

MILES, Alistair a Sean BECHHOFER, ed., SKOS Simple Knowledge Organization System Reference. In: *W3C* [online]. 2009 [cit. 2020-04-25]. Dostupné z: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.

MÜLLER-PROVE, Matthias. Modell und Anwendungsperspektive des Social Tagging. In: *Good tags - bad tags. Social Tagging in der Wissensorganisation* [online]. Münster u.a.: Waxmann, 2008 [cit. 2020-04-25]. ISBN 978-3-8309-2039-7. ISSN 1434-3436. Dostupné z (URN) : <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-127570>.

O'NEILL, Edward T., Rick BENNETT a Kerre KAMMERER. Using Authorities to Improve Subject Searches. *Cataloging & Classification Quarterly* [online]. Abington: Taylor & Francis, 2014, **52**(1), s. 6-19 [cit. 2020-04-25]. ISSN 0163-9374. Komerčně dostupné z (DOI): <https://doi.org/10.1080/01639374.2013.850018>.

OULTON, Emma. 12 Of The Biggest Bookshops In The World For When You Want To Lose Yourself In Literature. In: *Bustle* [online]. New York: BDG Media, 2017 [cit. 2020-04-25]. Dostupné z: <https://www.bustle.com/p/12-of-the-biggest-bookshops-in-the-world-for-when-you-want-to-lose-yourself-in-literature-40100>.

PABLO, Luis a María GOICOECHEA. A Survey of Electronic Literature Collections. *CLCWeb: Comparative Literature and Culture: A WWWeb Journal* [online]. Edmonton: Purdue University Press, 2014, **16**(5), s. 1-9 [cit. 2020-04-25]. ISSN 1481-4374. Dostupné z (DOI): <https://doi.org/10.7771/1481-4374.2558>.

PARK, Jung-ran a Yuji TOSAKA. Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability. *Information Technology and Libraries* [online]. Chicago: American Library Association, 2010, **29**(3), s. 104-116 [cit. 2020-04-25]. ISSN 2163-5226. Dostupné z (DOI): <https://doi.org/10.6017/ital.v29i3.3136>.

PETERS, Isabella. *Folksonomies: Indexing and Retrieval in Web 2.0* [online]. Berlin: De Gruyter Saur, 2009 [cit. 2020-04-25]. ISBN 978-3-598-25179-5. Komerčně dostupné z: <http://search.ebscohost.com>.

PETERS, Isabella a Wolfgang G. STOCK. Folksonomy and information retrieval. In: *Proceedings of the American Society for Information Science and Technology* [online]. Medford (New Jersey): Information Today, 2007, **44**(1), s. 1-28

[cit. 2020-04-25]. ISSN 0044-7870. Dostupné z (DOI):

<https://doi.org/10.1002/meet.1450440226>.

POMERANTZ, Jeffrey. *Metadata* [online]. Cambridge (Massachusetts): MIT Press, 2015 [cit. 2020-04-25]. ISBN 978-0-262-52851-1. Komerčně dostupné z: <http://search.ebscohost.com>.

RAFFERTY, Pauline. Tagging. *Knowledge Organization* [online]. Würzburg: Ergon Verlag, 2018, **45**(6), s. 500-516 [cit. 2020-04-25]. DOI: 10.5771/0943-7444-2018-6-500. ISSN 0943-7444. Komerčně dostupné z:

<http://search.ebscohost.com>.

RIVA, Pat, Patrick LE BŒUF a Maja ŽUMER. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information* [online]. Revidované vydání. Haag: International Federation of Library Associations and Institutions, 2017 [cit. 2020-04-25]. Dostupné z: https://www.ifla.org/files/assets/cataloguing/frbr-irm/ifla-irm-august-2017_rev201712.pdf.

SAYNOR, Chris. Thema: Developing A Subject Category Scheme for A Global Book Trade. In: *Book Industry Study Group* [online]. New York: Book Industry Study Group, 2019 [cit. 2020-04-25]. Dostupné z:

<https://bisg.org/news/459676/Thema-Developing-A-Subject-Category-Scheme-for-A-Global-Book-Trade.htm>.

SICILIA, Miguel-Angel, ed. *Handbook Of Metadata, Semantics And Ontologies* [online]. Singapore: World Scientific, 2014 [cit. 2020-04-25]. ISBN 978-981-283-629-8. Komerčně dostupné z: <http://search.ebscohost.com>.

SMITH, Gene. Information architecture: Tagging: Emerging trends. *Bulletin of the American Society for Information Science and Technology* [online]. Silver Spring (Maryland): American Society for Information Science and Technology, 2008, **34**(6), s. 14-17 [cit. 2020-04-25]. ISSN 0095-4403. Dostupné z (DOI):

<http://doi.wiley.com/10.1002/bult.2008.1720340607>.

STANWICKS, Kabel Nathan. *Music Social Tagging as a Validation Tool for the FRBR Conceptual Model* [online]. Albany (New York), 2017 [cit. 2020-04-25]. ISBN 978-0-355-50970-0. Dostupné z:

<https://search.proquest.com/docview/1975481985>. Magisterská práce. State University of New York at Albany.

STEWART, Brian. Pictures into words. *The Indexer: The International Journal of Indexing* [online]. Liverpool: Liverpool University Press, 2015, **33**(1), s. 8-25 [cit. 2020-04-25]. ISSN 0019-4131. Komerčně dostupné z (DOI):

<https://doi.org/10.3828/indexer.2015.3>.

ŠAUPERL, Alenka. Pinning down a novel: characteristics of literary works as perceived by readers. *Library Review* [online]. Bingley: Emerald Publishing, 2012, **61**(4), s. 286-303 [cit. 2020-04-25]. ISSN 0024-2535. Komerčně dostupné z (DOI): <https://doi.org/10.1108/00242531211267581>.

TEMPLE, Emily. The 12 Most Popular Libraries in the World. In: *Literary Hub* [online]. New York, 2018a [cit. 2020-04-25]. Dostupné z: <https://lithub.com/the-12-most-popular-libraries-in-the-world/>.

TEMPLE, Emily. The 10 Most Famous Bookstores in the World. In: *Literary Hub* [online]. New York, 2018b [cit. 2020-04-25]. Dostupné z: <https://lithub.com/the-10-most-famous-bookstores-in-the-world/>.

Thema v1.4: Thema – the subject category scheme for a global book trade. In: *EDItEUR* [online]. London: EDItEUR, c2009 [cit. 2020-04-25]. Dostupné z: <https://www.editeur.org/151/Thema/>.

The New York Times Fiction Best Sellers of 2015. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2019-07-21 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/The_New_York_Times_Fiction_Best_Sellers_of_2015.

The New York Times Fiction Best Sellers of 2016. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2018-07-06 [cit. 2020-04-25]. Dostupné z:

https://en.wikipedia.org/wiki/The_New_York_Times_Fiction_Best_Sellers_of_2016.

The New York Times Fiction Best Sellers of 2017. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (Kalifornie): Wikimedia Foundation, 2001-, naposledy edit. 2020-03-07 [cit. 2020-04-25]. Dostupné z: https://en.wikipedia.org/wiki/The_New_York_Times_Fiction_Best_Sellers_of_2017.

The world's greatest bookshops. In: *Lonely Planet* [online]. Melbourne, 2012 [cit. 2020-04-25]. Dostupné z: <https://www.lonelyplanet.com/articles/the-worlds-greatest-bookshops>.

TOSAKA, Yuji a Cathy WENG. Reexamining Content-Enriched Access: Its Effect on Usage and Discovery. *College & Research Libraries* [online]. Chicago: Association of College and Research Libraries, 2011, **72**(5), s. 412-427 [cit. 2020-04-25]. ISSN 2150-6701. Dostupné z (DOI): <https://doi.org/10.5860/crl-137>.

VAIDYA, Praveenkumar a N. S. HARINARAYANA. The role of social tags in web resource discovery: an evaluation of user-generated keywords. *Annals of Library and Information Studies* [online]. New Delhi: National Institute of Science Communication and Information Resources, 2016, **36**(4), s. 289-297 [cit. 2020-04-25]. ISSN 0975-2404. Dostupné z: <http://op.niscair.res.in/index.php/ALIS/article/view/14494>.

VANDER WAL, Thomas. Explaining and Showing Broad and Narrow Folksonomies. In: *Vanderwal.net* [online]. 2005 [cit. 2020-04-25]. Dostupné z: <http://www.vanderwal.net/random/entrysel.php?blog=1635>.

VANDER WAL, Thomas. Folksonomy Coinage and Definition. In: *Vanderwal.net* [online]. 2007 [cit. 2020-04-25]. Dostupné z: <http://www.vanderwal.net/folksonomy.html>.

Vocabulary Control. In: *The Oxford Companion to the English Language* [online]. 2 ed. Oxford: Oxford University Press, 2018 [cit. 2020-04-25]. ISBN 978-0-19-174438-9. DOI: 10.1093/acref/9780199661282.001.0001. Komerčně dostupné z:

<https://www.oxfordreference.com/view/10.1093/acref/9780199661282.001.0001/acref-9780199661282>.

Vocabulary Encoding Schemes: RDA Terms. In: *RDA Toolkit* [online]. American Library Association, Canadian Federation of Library Associations, CILIP: Chartered Institute of Library and Information Professionals, c2020 [cit. 2020-04-25]. Komerčně dostupné z: <https://beta.rdatoolkit.org/>.

WALSH, John. The use of Library of Congress Subject Headings in digital collections. *Library Review* [online]. Bradford (Anglie): Emerald Publishing, 2011, **60**(4), s. 328-343 [cit. 2020-04-25]. ISSN 0024-2535. Komerčně dostupné z (DOI): <https://doi.org/10.1108/0024253111127875>.

WINGET, Megan. User-defined classification on the online photo sharing site Flickr... or, How I learned to stop worrying and love the million typing monkeys. *Advances in Classification Research Online* [online]. Medford (New Jersey): Information Today, 2006, **17**(1), s. 1-16 [cit. 2020-04-25]. DOI: 10.7152/acro.v17i1.12496. ISSN 2324-9773. Dostupné z: <https://journals.lib.washington.edu/index.php/acro/article/view/12496/10995>.

WELLISCH, H. H. *Indexing from A to Z*. 2nd edition. New York: H.W. Wilson, 1996. ISBN 978-0-8242-0882-0.

What the Freedom of Information Act does. In: *UK Parliament* [online]. London: UK Parliament, c2020 [cit. 2020-04-25]. Dostupné z: <https://www.parliament.uk/mps-lords-and-offices/offices/lords/freedom-of-information-in-the-house-of-lords/foi-lords-what-the-act-does/>.

WHITE, Hollie C. *Organizing Scientific Data Sets: Studying Similarities and Differences in Metadata and Subject Term Creation* [online]. Chapel Hill, 2012 [cit. 2020-04-25]. ISBN 978-1-2674-2049-7. Komerčně dostupné z: <https://search.proquest.com/docview/1024563199>. Disertační práce. University of North Carolina.

WYRWOLL, Claudia. *Social Media: Fundamentals, Models, and Ranking of User-Generated Content* [online]. Wiesbaden: Springer Fachmedien

Wiesbaden, 2014 [cit. 2020-04-25]. ISBN 978-3-658-06984-1. Komerčně dostupné z: <https://link.springer.com/book/10.1007%2F978-3-658-06984-1>.

YANG, Sharon Q. Tagging for Subject Access: A Glimpse into Current Practice by Vendors, Libraries, and Users. *Computers in Libraries* [online]. Medford: Information Today, 2012, **32**(9), s. 19-23 [cit. 2020-04-25]. ISSN 1041-7915. Dostupné z: <http://www.infotoday.com/cilmag/nov12/index.shtml>.

ZAVALINA, Oksana L. Exploring the role of scale: Comparative analysis of digital library user searching. *Proceedings of the American Society for Information Science and Technology* [online]. Silver Spring (Maryland): American Society for Information Science and Technology, 2014, **51**(1), s. 1-4 [cit. 2020-04-25]. ISSN 0044-7870. Dostupné z (DOI): <http://doi.wiley.com/10.1002/meet.2014.14505101120>.

ZERVAS, Panagiotis, Sampson DEMETRIOS G. a Maria ARISTEIDOU. Investigating the Effect of Users' Tagging Motivation on the Digital Educational Resources Metadata Descriptions. *International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in Digital Age (CELDA)* [online]. Madrid: International Association for Development of the Information Society, 2012 [cit. 2020-04-25]. Dostupné z: <https://eric.ed.gov/?id=ED542710>.

Seznamy tabulek a grafů

Seznam tabulek

Tabulka 1: Seznam analyzovaných knih	53
Tabulka 2: Typy metadat a jejich aplikovatelnost na profesionální a uživatelská metadata.....	70
Tabulka 3: Objektivní uživatelská metadata – přehled	79
Tabulka 4: Objektivní uživatelská metadata podle formy.....	80
Tabulka 5: Objektivní uživatelská metadata, která odpovídají profesionálním – přehled	81
Tabulka 6: Objektivní uživatelská metadata, která odpovídají profesionálním, podle formy	81
Tabulka 7: Objektivní uživatelská metadata, která neodpovídají profesionálním – přehled	82
Tabulka 8: Objektivní uživatelská metadata, která neodpovídají profesionálním, podle formy	82
Tabulka 9: Porovnání profesionálních a uživatelských metadat podle typů metadatových výroků	84
Tabulka 10: Profesionální hodnoty, které odpovídaly Thema – přehled	87
Tabulka 11: Profesionální hodnoty, které neodpovídaly Thema – přehled.....	87
Tabulka 12: Rozdělení profesionálních hodnot podle komplexnosti	87
Tabulka 13: Uživatelské hodnoty, které odpovídaly Thema – přehled.....	90
Tabulka 14: Uživatelské hodnoty, které neodpovídaly Thema – přehled	90
Tabulka 15: Rozdělení uživatelských hodnot podle komplexnosti	90
Tabulka 16: Uživatelské hodnoty dle přesnosti mapování na Thema	91
Tabulka 17: Jednoduché uživatelské hodnoty dle přesnosti mapování na Thema	91
Tabulka 18: Uživatelské hodnoty složené ze dvou dle přesnosti mapování na Thema	92
Tabulka 19: Uživatelské hodnoty složené ze tří dle přesnosti mapování na Thema	92
Tabulka 20: Uživatelské hodnoty složené ze čtyř dle přesnosti mapování na Thema	93

Tabulka 21: Typy obsahových metadat a jejich aplikovatelnost na profesionální a uživatelská metadata.....	95
Tabulka 22: Složitost metadatových hodnot podle tvůrců	103

Seznam grafů

Graf 1: Rozložení uživatelských metadat.....	80
Graf 2: Poměr unikátních hodnot profesionálních a uživatelských metadat pro zjištění sémantické interoperability	86
Graf 3: Profesionální hodnoty dle přesnosti mapování na Thema	88
Graf 4: Jednoduché profesionální hodnoty dle přesnosti mapování na Thema.....	88
Graf 5: Profesionální hodnoty složené ze dvou dle přesnosti mapování na Thema	89
Graf 6: Profesionální hodnoty složené ze tří dle přesnosti mapování na Thema ..	89
Graf 7: Počet databází, ve kterých se vyskytují různé typy profesionálních metadat	97
Graf 8: Počet databází, ve kterých se vyskytují různé typy uživatelských metadat	97
Graf 9: Profesionální metadatové výroky dle typů.....	99
Graf 10: Uživatelské metadatové výroky dle typů	99
Graf 11: Využívanost typů metadatových výroků – profesionální metadata	100
Graf 12: Využívanost typů metadatových výroků – uživatelská metadata	100

Seznam příloh

1. Seznam analyzovaných databází
2. Materiály k výběru databází
3. Typy metadat dle databází
4. Analyzované metadatové výroky
5. Databáze, které používají jinou verzi DDT než 23
6. Objektivita metadatových výroků a jejich shoda s profesionálními metadaty
7. Počet metadatových výroků dle typů
8. Mapování metadatových výroků na Thema
9. Sémantická interoperabilita dle typů metadat
10. Seznam uživatelských objektivních metadatových výroků, které neodpovídaly profesionálním