

Univerzita Karlova v Praze

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní program: Informační studia a knihovnictví

Studijní obor: Informační studia a knihovnictví

Diplomová práce

Bc. Jana Křížová

Digitální knihovna HathiTrust, její funkce a služby

Digital Library HathiTrust, its functions and services

Praha, 2017

Vedoucí práce Mgr. Adéla Jarolímková, PhD.

Vedoucí diplomové práce: Mgr. Adéla Jarolímková, PhD.

Oponent diplomové práce:

Datum obhajoby:

Hodnocení:

Prohlášení:

Prohlašuji, že jsem 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 31. 7. 2017

.....
Podpis studenta

Abstrakt:

Cílem diplomové práce je popsat Digitální knihovnu HathiTrust a analyzovat její obsah se zřetelem k česky psaným dokumentům. První a část druhé kapitoly je napsána na základě prostudovaných zdrojů a jsou v ní shrnuty informace o vzniku partnerství a Digitální knihovny HathiTrust. Dále popisuje její cíle, různé dílčí projekty a služby (například Výzkumné centrum HathiTrust, dále projekt zaměřený na vládní dokumenty USA a také systém pro správu metadat), možnosti a výhody členství a velmi stručně shrnuje některé z technických parametrů této digitální knihovny. Druhá kapitola je částečně (zejména pokud jde o popis uživatelského rozhraní Digitální knihovny HathiTrust) vytvořena na základě praktických zkušeností s jejím používáním. Kromě analýzy uživatelského rozhraní (i jeho mobilní verze) obsahuje tato část také popis metadat, jejich formáty, standard i pohyby. Poslední část druhé kapitoly je věnována problematice autorských práv, možnostem zpřístupňování plných textů a také soudnímu sporu Authors Guild v. HathiTrust. Třetí kapitola popisuje analýzu obsahu Digitální knihovny HathiTrust, s důrazem na česky psané dokumenty. Dále jsou popsány a interpretovány výsledky této analýzy. Česky psané dokumenty byly analyzovány na základě čtyř kritérií: datum vydání, místo vydání, typ dokumentu a také jeho obsah ve vztahu k oboru Informační studia a knihovnictví. V závěru práce jsou zhodnoceny výsledky obsahové analýzy.

Klíčová slova: partnerství HathiTrust, Digitální knihovna HathiTrust, obsahová analýza, bohemikální dokumenty

Abstract:

The aim of the diploma thesis is to describe the HathiTrust Digital Library and analyze its content with respect to documents written in Czech language. The first and part of the second chapter is written on the basis of the studied sources and summarizes information about the partnership and the HathiTrust Digital Library. It also describes its objectives, various sub-projects and services (such as the HathiTrust Research Center, a US federal documents project and a metadata management system), membership options and benefits, and briefly summarizes some of the technical parameters of this digital library. The second chapter is partly based on practical experience with use of the HathiTrust Digital Library user interface. In addition to analyzing the user interface (including its mobile version), this section also includes description of metadata, its formats, standards, and movements. The last part of the second chapter is devoted to copyright issues, the possibilities of accessing full texts, and also to the lawsuit Authors Guild v. HathiTrust. The third chapter describes the analysis of the content of the HathiTrust Digital Library, with an emphasis on documents written in Czech. The results of this analysis are described and interpreted. The Czech written documents were analyzed on the basis of four criteria: the date of publication, the place of publication, the type of document and its content in relation to the field of Information Studies and Librarianship. At the end of the thesis the results of the content analysis are evaluated.

Keywords: HathiTrust partnership, HathiTrust Digital Library, content analysis, bohemical documents

Identifikační záznam:

KŘÍŽOVÁ, Jana. *Digitální knihovna HathiTrust, její funkce a služby*. Praha, 2017. 87 s., příloha CD. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce Mgr. Adéla Jarolímková, Ph.D.

Obsah

Použité zkratky.....	9
Úvod.....	10
1. HathiTrust.....	12
1.1 Cíle a funkce.....	12
1.2 Struktura.....	14
1.2.1 Správní rada a Výkonný výbor.....	15
1.2.2 Programová řídicí komise.....	17
1.2.3 Ostatní komise a pracovní skupiny.....	17
1.3 Vznik a vývoj.....	18
1.3.1 Projekty.....	19
1.3.2 Výzkumné centrum HathiTrust.....	27
1.4 Členství v HathiTrust.....	29
1.4.1 Podmínky přijetí.....	30
1.4.2 Výhody členství.....	32
1.5 Technické parametry a standardy HTDL.....	32
2. Možnosti přístupu k obsahu HTDL.....	34
2.1 Analýza uživatelského rozhraní HTDL - Online katalog.....	35
2.1.1 Registrace uživatelů.....	35
2.1.2 Možnosti vyhledávání v HTDL.....	36
2.1.3 Výsledky vyhledávání.....	37
2.1.4 Záznam dokumentu v HTDL.....	38
2.1.5 Funkce vytváření sbírek.....	41
2.1.6 Mobilní verze HTDL.....	42
2.2 Metadata v HTDL.....	45
2.2.1 Příjem metadatových záznamů do HTDL.....	46
2.2.2 Standard pro metadata v HTDL.....	48
2.2.3 Možnosti získání záznamů z HTDL.....	50
2.2.4 Spolupráce s OCLC.....	51
2.3 Ukládání digitálních objektů do HTDL.....	52
2.4 Autorská práva.....	53
2.4.1 Dostupnost plných textů.....	54

2.4.2	Soudní spor Authors Guild v. HathiTrust	56
3.	Analýza obsahu digitální knihovny HathiTrust	59
3.1	Cíle a metodika obsahové analýzy HTDL.....	59
3.1.1	Metody sběru dat.....	59
3.1.2	Metody analýzy dat	60
3.2	Dostupné statistiky obsahu HTDL	61
3.2.1	Jazyk dokumentu.....	61
3.2.2	Datum vydání	63
3.3	Výsledky analýzy bohemikálních dokumentů v HTDL	65
3.3.1	Datum vydání dokumentu	66
3.3.2	Místo vydání.....	70
3.3.3	Typ dokumentu	74
3.3.4	Obsah dokumentů ve vztahu k oboru Informační studia a knihovnictví.....	75
	Závěr.....	77
	Seznam použitých zdrojů	79
	Seznam obrázků	84
	Seznam tabulek	84
	Seznam grafů.....	84
	Seznam příloh na CD	84
	Přílohy	85

Použité zkratky

CDL	California Digital Library
CIC	Committee on Institutional Cooperation
CRL	Center for Research Libraries
CRMS	Copyright Review Management System
DC	Dublin Core
DPN	Digital Preservation Network
GPO	Government Printing Office
HT	HathiTrust
HTDL	HathiTrust Digital Library
HTRC	HathiTrust Research Center
IMLS	Institute of Museum and Library Service
LCC	Library of Congress
OCLC	Online Computer Library Center
OWP	Orphan Works Project

Úvod

Digitální knihovna HathiTrust je v dnešní době, co do obsahu, jednou z největších na světě. Spravuje a rozvíjí ji stejnojmenné partnerství dnes již více než sta převážně univerzitních knihoven a jiných vzdělávacích či výzkumných institucí z různých zemí světa, nejčastěji však stále ze Spojených států. Minimálně pro své členy je tato digitální knihovna již neodmyslitelným informačním zdrojem, který pro ně může být přístupný i z jejich vlastního katalogu.

Toto téma mě zaujalo, protože mě zajímá problematika digitálních knihoven, ať už ze studijního, tak také z pracovního hlediska. Od června roku 2016 totiž pracuji na částečný úvazek v Oddělení pro digitální knihovnu v Knihovně Akademie věd, v. v. i.

Diplomová práce je členěna do tří částí. První část představuje popis vzniku a vývoje partnerství a Digitální knihovny HathiTrust. Část této kapitoly je věnována také cílům partnerství HathiTrust a možnostem a podmínkám členství. Na závěr první kapitoly jsou velmi stručně popsány některé technické parametry HTDL.

Druhá kapitola je zaměřena na obsah HTDL a je již spíše praktičtější povahy. Nejprve je analyzováno uživatelské rozhraní HTDL, jeho možnosti, funkce a také mobilní verze. Zbytek této kapitoly je věnován metadatům a autorskoprávní problematice. Mimo jiné je zde stručně popsán soudní spor, který partnerství vedlo se zástupci autorů a držitelů autorských práv.

Poslední kapitola zahrnuje analýzu HTDL z hlediska obsahu. Nejprve jsou stručně shrnuty již dostupné obsahové statistiky. Tyto statistiky zahrnují všechny dokumenty v HTDL. Následně jsou popsány výsledky obsahové analýzy bohemikálních (česky psaných) dokumentů. Hlavní zdrojem dat pro první část analýzy byly již zmíněné statistiky. Pro získání dat pro rozbor česky psaných dokumentů v HTDL bylo použito funkce vytváření sbírek, kterou nabízí uživatelské rozhraní HTDL. Všechny údaje byly následně analyzovány prostřednictvím tabulkového editoru Excel.

Jako hlavní zdroje teoretické části práce byly použity webové stránky HathiTrust, kde je o digitální knihovně i celém partnerství a dalších vedlejších projektech opravdu mnoho informací. Dále jsem čerpala z článků v odborných periodících, případně z kratších zpráv na webu. Zajímavým zdrojem byl profil HTDL na sociální síti Youtube, kde je v současné době

nahráno deset videí¹, která se týkají této digitální knihovny. Jde o přednášky z konferencí a také o nahrané webináře, které pracovníci HathiTrust pořádají pro nové členy. Posledním zdrojem byla také emailová komunikace s pracovníky HTDL, kdy byly získávány další informace a také nahlašované nalezené chyby, jak v rozhraní, tak v metadatových záznamech.

Předkládaná diplomová práce má celkem 87 stran, což je 86 normostran a 155 666 znaků s mezerami. V práci je použito jedenáct obrázků, osm tabulek a pět grafů. Obrázky tvoří ve většině případů snímky obrazovky uživatelského rozhraní HTDL pro ilustraci jeho popisu. Grafy a tabulky byly vkládány pro přehlednost. Diplomová práce je doplněna o tři přílohy, z nichž jedna je kvůli velikosti svého obsahu připojena na CD. Jde o data pro obsahovou analýzu česky psaných dokumentů v HTDL.

Diplomová práce byla vytvořena v souladu s vnitřními pravidly pro evidenci, odevzdávání, obhajování a zveřejňování závěrečných prací Filozofické fakulty Univerzity Karlovy. Použité zdroje byly v práci citovány v souladu s normou ISO-690:2010. Seznam použitých zdrojů na konci práce je řazen jmenně abecedně, v případech, kdy se první údaje shodují, jsou k rozlišení použita malá písmena abecedy. Pro citování byl použit Harvardský styl citování, tedy citování pomocí odkazů v textu s uvedeným příjmením autora, datem vydání a případně údajem o stránkování.

Na závěr bych chtěla velmi poděkovat oběma vedoucím své diplomové práce: PhDr. Evě Bratkové, Ph.D. a Mgr. Adéle Jarolímkové, Ph.D za všechny rady a připomínky, které byly vždy velmi užitečné a ochotně poskytnuté. Dále patří poděkování pracovníkům HTDL za jejich ochotu při zodpovídání mých dotazů a za rychlé reakce na všechny mé podněty a hlášení případných chyb.

¹ Informace k 16. 6. 2017, kanál HathiTrust je dostupný zde: <https://www.youtube.com/user/HathiTrust>

1. HathiTrust

HathiTrust je název partnerství převážně amerických univerzitních knihoven, které vytvořilo a nadále spravuje digitální knihovnu HathiTrust (HTDL) a další přidružené služby. Partnerství je podle práva Spojených států obchodní spojení dvou a více subjektů, které sdílejí a spoluvlastní společný majetek a zisk, na základě velikosti jejich počátečního příspěvku. Zákon nespécifikuje, o jaké subjekty se jedná, může jít o jednotlivé fyzické osoby, jejich skupiny, firmy, korporace apod. (CORNELL LAW SCHOOL, 2015). V případě HathiTrust jde o výzkumné instituce, knihovny a další instituce, které vlastní fond odborné literatury.

Název HathiTrust je odvozen z hindského slova *hathi*, což je označení pro slona. Jelikož jde o digitální knihovnu, jejímž cílem je uchovávat dokumenty po velmi dlouhou dobu (nejlépe navždy), nabízí se zde spojení s pověstnou sloní pamětí. Zároveň je vyobrazení slona základním prvkem loga HathiTrust (CHRISTENSON, 2011). Protože jde o v angličtině cizí slovo, v mnoha zdrojích se objevují různé varianty zápisu². Na webu HathiTrust jsou pro tyto případy uvedeny preferované podoby a významy názvu HathiTrust. Uvádí se bez mezery s velkým počátečním písmenem a s velkým písmenem ve slově „Trust“ (UNIVERSITY OF MICHIGAN, 2016d).

V následující kapitole jsou popsány cíle a funkce partnerství HathiTrust, následně jeho struktura, vývoj a možnosti a podmínky členství. Vývoj HathiTrust je popsán pomocí krátkého představení některých projektů, které ho po celou dobu existence doprovázejí. V závěru kapitoly jsou stručně představeny některé ze základních technických parametrů HTDL.

1.1 Cíle a funkce

Cíle partnerství HathiTrust jsou zakotveny v jeho stanovách. Hlavním cílem je podpora výzkumu, studia a všeobecného blaha pomocí sběru, organizace, ochrany a sdílení informací a znalostí (jejich záznamů). Tyto procesy probíhají za aktivní spolupráce všech členů partnerství HathiTrust. Tento obecný cíl je ve stanovách společnosti rozdělen do devíti následujících bodů.

² Nejčastější chybou je zápis s mezerou uprostřed slova: „Hathi Trust“, například v názvu a abstraktu tohoto článku s názvem *Comparison of three digital library interfaces: open library, Google books, and Hathi Trust* z roku 2012: <https://dx.doi.org/10.1145/2232817.2232894>

- „Budování spolehlivého digitálního repozitáře, který obsahuje digitalizované dokumenty z fondů členských institucí (které tento archiv spoluvlastní a podílejí se na jeho správě).
- Rozšiřování přístupu k těmto dokumentům, v co největším možném souladu s potřebami členských institucí. Zároveň je cílem co možná v největším měřítku zpřístupňovat dokumenty uživatelům s handicapem, který jim neumožňuje nebo výrazně znesnadňuje používat tištěné dokumenty³.
- Vývoj infrastruktury pro ukládání, zpřístupňování a další procesy zahrnující digitální dokumenty, která by byla jednak dostupná (co se týče všech možných zdrojů), jednak přínosná pro uživatele. Toto řešení by zároveň mělo zahrnovat podporu různých formátů digitálních dokumentů.
- Zajištění dlouhodobé archivace všech dokumentů, které jsou obsaženy v HTDL.
- Zlepšování kooperativní strategie všech členských knihoven pro správu a uchování tištěných fondů, čímž dojde ke snižování nákladů na výše uvedené.
- Vytvoření a udržování podpůrných služeb, které budou umožňovat používat HTDL a vyhledávat v ní.
- Vytvoření technického rámce, který umožňuje jak centralizovaný, tak distribuovaný vývoj nástrojů a služeb.
- Udržení partnerství HathiTrust v nekomerční sféře a zároveň definování služeb, ze kterých budou mít prospěch členské instituce“ (UNIVERSITY OF MICHIGAN, 2016c).

Z těchto devíti cílů je nejčastěji zmiňováno zpřístupňování plných textů dokumentů, ať už lidem s handicapem nebo běžným uživatelům. Velká část níže zmiňovaných projektů se také tímto problémem zabývá, většinou v nich jde o snahu rozšířit přístup k dokumentům v HTDL. Také se touto problematikou zabíral vleklý soudní spor, šlo v něm zejména o to, určit, zda zpřístupnění plného textu dokumentu uživatelům s handicapem, je porušením autorského práva či nikoliv (více v kapitole o [autorskoprávní problematice](#)). Uživatel s handicapem, který mu neumožňuje nebo výrazně znesnadňuje studium tištěných dokumentů, může prostřednictvím HTDL studovat dokumenty, které již nejsou chráněny autorským zákonem v pro něj mnohem přívětivějším provedení, ve formě prostého textu. Takový text je také snáze zpracovatelný různými čtecími zařízeními, než text ve formě naskenovaného obrazu. Co se týče děl stále spadajících pod ochranu autorského zákona, jsou

³ Uživatelům s tzv. Print disabilities

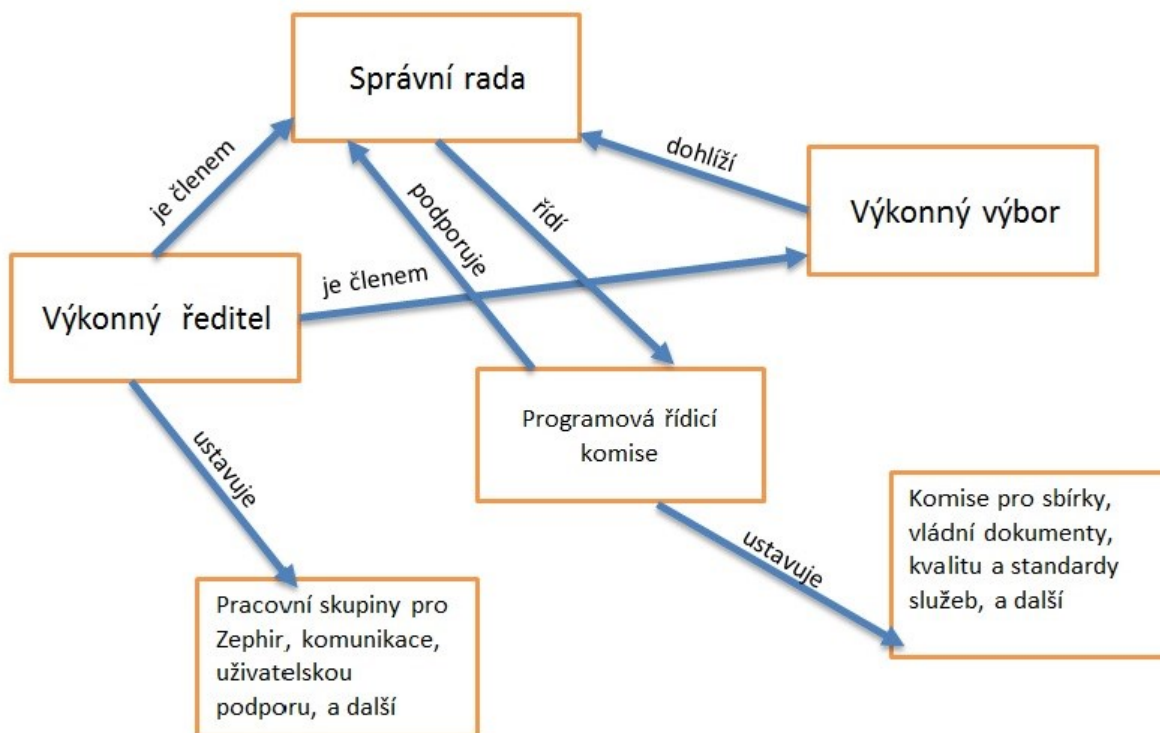
pro handicapované uživatele dostupné v plném textu. Ovšem stejně jako u běžných uživatelů, i tento přístup má svá omezení. Handicapovaný uživatel musí být ve svojí instituci (která je členem partnerství HathiTrust) identifikován jako uživatel s handicapem, který mu neumožňuje nebo výrazně znesnadňuje studium dokumentů v tištěné podobě. Tato informace musí být zanesena v záznamu studenta v příslušné instituci a také tuto informaci musí nějakým způsobem obdržet HTDL. Za těchto podmínek je systém schopen po přihlášení do katalogu HathiTrust zaznamenat, že jde o handicapovaného uživatele a na tomto základě upravit přístupová práva k určitým dokumentům. Ani tyto uživatelé totiž nemají přístup k plným textům všech dokumentů, které stále chrání autorský zákon. Mohou studovat plné texty pouze těch dokumentů, jež se nacházejí, nebo někdy nacházely ve fondu jejich instituce. Informace o vlastnících je uvedena v záznamu každého dokumentu v HTDL (ZAYTSEV, 2015).

1.2 Struktura

Jelikož má partnerství HathiTrust již mnoho členů, je také jeho organizační struktura poměrně široká a rozvětvená. V následujícím schématu (viz [Obrázek 1](#)) je nastíněno, ze kterých orgánů se HathiTrust skládá a jak na sebe navazují. Současná organizační struktura a vedení HathiTrust bylo ustaveno v rámci Ústavního shromáždění, které se konalo v roce 2011. Následkem tohoto setkání byl vznik Správní rady v dubnu 2012. Do té doby byl hlavním vedoucím orgánem HathiTrust Výkonný výbor (UNIVERSITY OF MICHIGAN, 2016g), na který dohlížel Strategický poradní výbor⁴. Na webu HathiTrust jsou stále uvedeny informace o tomto uskupení, jeho složení a funkcích. Dokonce zde lze nalézt zápisy ze setkání a schůzek, poslední je z 19. července 2012 (UNIVERSITY OF MICHIGAN, 2016q). Kromě níže uvedených orgánů organizační struktury se každoročně koná setkání zástupců všech členů partnerství HathiTrust. Na těchto setkáních se pořádají volby do správní rady a jsou zde prezentovány pokroky ve vývoji. Po každém setkání vznikne dokument, který obsahuje prezentace a poznámky ke všem tématům, které na něm byly diskutovány⁵.

⁴ Strategic Advisory Board

⁵ Dokumenty jsou dostupné na webu HathiTrust: <https://www.hathitrust.org/governance>



Obrázek 1: Schéma organizační struktury partnerství HathiTrust

Co se týče financování, partnerství HathiTrust disponuje vlastním rozpočtem, který je tvořen členskými příspěvky (více v podkapitole o [podmínkách přijetí](#)). Některé jednotlivé projekty, programy a další dílčí aktivity jsou finančně podporovány z nejrůznějších grantů. Institucí, která granty často poskytuje, je Ústav muzejních a knihovních služeb⁶ (dále IMLS) (UNIVERSITY OF MICHIGAN, 2016d).

1.2.1 Správní rada a Výkonný výbor

Orgánem, který má za činnost partnerství HathiTrust největší zodpovědnost je Správní rada⁷. Správní rada má běžně dvanáct členů (resp. třináct s nevolícím členem, kterým je výkonný ředitel). Šest míst náleží zástupcům zakládajících členů, dvě místa jsou obsazena zástupci systému Kalifornské univerzity⁸, další dvě místa se rozdělují mezi následující instituce: Chicagská univerzita, Iowská univerzita, Michiganská státní univerzita, Minnesotská univerzita, Severozápadní univerzita (Illinois), Ohijská státní univerzita, Pensylvánská státní univerzita a Wisconsinská univerzita v Madisonu. Po jednom místě je obsazeno Indianskou a Michiganskou univerzitou (UNIVERSITY OF MICHIGAN,

⁶ Institute of Museum and Library Service, viz: <https://www.imsls.gov/>

⁷ Board of Governors

⁸ Kalifornská univerzita zahrnuje celkem devět univerzitních kampusů a mnoho dalších součástí po celém státě i v zahraničí, viz: <https://www.universityofcalifornia.edu/uc-system/parts-of-uc>

2016c). Zbýlých šest míst je obsazeno volenými členy z celého partnerství HathiTrust. Tito zástupci jsou voleni na tříleté funkční období. V případě, že některý člen odstoupí, kvůli jiným důvodům, než je skončení volebního období (např. odchod do důchodu, zdravotní problémy apod.), konají se mimořádné volby, aby byl počet členů Správní rady stále stejný. V současné době⁹ se Správní rada skládá ze čtrnácti členů (dva z nich jsou nehlasující členové, mají pouze poradní hlas):

- John Butler (Minnesotská univerzita, předseda Programové řídicí komise¹⁰) – nehlasující člen
- Richard Clement (Univerzita Nového Mexika)
- John Culshaw (Iowská univerzita)
- Mike Fulrough (výkonný ředitel HathiTrust) – nehlasující člen
- James Hilton (Michiganská univerzita)
- Anne Kenney (Cornellova univerzita)
- Wendy Lougee (Minnesotská univerzita)
- Beth McNeil (Iovská státní univerzita)
- Brian Schottlaender (Kalifornská univerzita, San Diego)
- Winston Tabb (Univerzita Johna Hopkinse)
- Günter Waibel (Kalifornská digitální knihovna)
- Carolyn Walters (Indiánská univerzita)
- Betsy Wilson (Washingtonská univerzita)
- Robert Wolven (Kolumbijská univerzita) (UNIVERSITY OF MICHIGAN, 2016b)

Do činností, kterým se Správní rada věnuje, patří například schvalování a správa rozpočtu a financí, určování strategie vývoje, úprava a schvalování cílů, dohlížení na rozvoj Výzkumného centra HathiTrust (dále HTRC), zakládání pracovních skupin, jimž zároveň zadává úkoly, které jsou v souladu s cíli a posláním partnerství HathiTrust. Rozsah činností a kompetencí Správní rady určují stanovy HathiTrust. Pět členů správní rady jsou zároveň členy Výkonného výboru, který na Správní radu dohlíží. Zároveň přebírá její funkce a odpovědnosti během různých přechodných období, kdy není usnášeníschopná. Členy Výkonného výboru jsou předseda, bývalý předseda, pokladník, předseda Programové řídicí komise a výkonný ředitel HathiTrust (UNIVERSITY OF MICHIGAN, 2016c). Výkonným

⁹ Informace k 8. 11. 2016

¹⁰ Program Steering Committee

ředitelem je od roku 2014 Mike Furlough, před ním tuto funkci od počátku existence HathiTrust zastával John Wilkin (UNIVERSITY OF MICHIGAN, 2014).

1.2.2 Programová řídicí komise

Činnost Programové řídicí komise řídí Správní rada. Komise připravuje podklady a strategie pro diskuze a zasedání Správní rady a na pomoc si může přizvat člena některé z pracovních skupin. Jejím úkolem je také pomocí tiskových a jiných zpráv informovat členy partnerství o činnostech Správní rady a tyto činnosti se členy komunikovat a diskutovat o nich. Programová řídicí komise by měla mít nejméně osm, nejvíce však dvanáct členů, kteří jsou voleni na tři roky. Předseda komise je určen Správní radou a zůstává ve své funkci po dobu dvou let. Zároveň je součástí Správní rady a účastní se jejích zasedání jako nevolící člen. Komise je spolu se Správní radou zodpovědná za udržování a vývoj partnerství HathiTrust v rámci jeho mise a dílčích cílů (UNIVERSITY OF MICHIGAN, 2016c).

1.2.3 Ostatní komise a pracovní skupiny

Kromě výše zmíněných komisí a výborů, existují v rámci HathiTrust také pracovní skupiny a menší komise, z nichž některé se vztahují k níže zmíněným [projektům](#). Tyto skupiny mohou vzniknout z popudu Výkonného ředitele HathiTrust (či Výkonného výboru) nebo Programové řídicí komise.

Pod pravomoc Výkonného ředitele spadají následující skupiny. První z nich je Poradní skupina pro uživatelský komfort¹¹, jejímž hlavním úkolem je zajišťovat a neustále zlepšovat uživatelské rozhraní HTDL a všech jeho součástí z hlediska použitelnosti. Tato skupina také vytváří návody a manuály pro vyhledávání. Pracovní skupina pro komunikaci¹² zajišťuje komunikaci ve všech formách v rámci celého partnerství HathiTrust. Stará se také o tiskové zprávy a další výstupy, kterými jsou například webové stránky HathiTrust, blog, sociální sítě, také určuje, jakou formou bude probíhat komunikace s uživateli apod. Další skupinou, která spadá pod Výkonného ředitele, je Pracovní skupina pro uživatelskou podporu¹³, která má na starost zodpovídání všech dotazů, které uživatelé zasílají pomocí e-mailu a dalších prostředků, které jsou veřejně k dispozici uživatelům. Poslední skupina, která pracuje pod dohledem a na pokyn Výkonného ředitele je Poradní skupina Zephir¹⁴. Jak její název napovídá, má tato skupina poradní funkci pro vývojový tým systému pro správu metadat s názvem Zephir. Počet

¹¹ User Experience Advisory Group

¹² Communications Working Group

¹³ User Support Working Group

¹⁴ Zephir Advisory Group

členů ve výše uvedených skupinách je proměnlivý a pohybuje se mezi šesti (Poradní skupina pro uživatelský komfort) a sedmadvaceti (Pracovní skupina pro uživatelskou podporu) členy (UNIVERSITY OF MICHIGAN, 2016t).

Pracovních skupin a menších komisí zřízených Programovou řídicí komisí je celkem šest. Komise pro správu sbírek¹⁵ má za úkol dohlížet na vývoj sbírek a dodržování politiky partnerství HathiTrust v této oblasti. Poradní komise pro federální dokumenty¹⁶ dohlíží a připomínkuje Program federálních vládních dokumentů Spojených států. Podobně Poradní komise pro sdílení tištěných dokumentů¹⁷ dohlíží na Program sdílení tištěných dokumentů. Poradní skupina pro oblast metadat (jejich politiku, použití, strategie a sdílení)¹⁸ má na starosti, jak název napovídá, vše, co se týká metadat. Mimo jiné také dohlíží na vývoj systému pro správu metadat Zephir. Další pracovní skupina¹⁹ se zabývá kvalitou služeb a všemi možnými standardy, které jsou v provozu HTDL používány. Poslední skupinou je Pracovní skupina pro práva a přístup²⁰, která zajišťuje, aby byl v HTDL umožněn co nejširší možný přístup ke všem dokumentům, u nichž je to možné. Zároveň pod její dohled spadá projekt Systému pro správu a revizi autorsko-právních metadat (dále CRMS) (UNIVERSITY OF MICHIGAN, 2016t).

1.3 Vznik a vývoj

Předpokladem vzniku partnerství HathiTrust byl masový digitalizační projekt společnosti Google s názvem Google Books. Jeho cílem bylo převést co nejvíce knih (z fondů knihoven) do digitální podoby a zpřístupnit je pro fulltextové prohledávání, případně komerčně v plném rozsahu. První záměr vytvořit takový projekt se objevil již v roce 2002, ovšem první smlouvy s knihovnami byly uzavřeny až v roce 2004 (FRČKA, 2014, s. 32-33). Jednou z prvních byl Systém knihoven Michiganské univerzity, která je jednou ze zakládajících institucí HathiTrust.

Velmi zásadním důvodem pro vznik HathiTrust byl pravděpodobně také přístup společnosti Google k možnosti zpřístupňovat digitalizované dokumenty volně a bezplatně, a to i v případech, kdy se jednalo o volná nebo osiřelá díla a vládní dokumenty, na něž se již nevztahují autorská práva. Podle (JOHNSON, 2009) společnost Google nechtěla poskytovat

¹⁵ Collections Committee

¹⁶ Federal Document Advisory Committee

¹⁷ Shared Print Advisory Committee

¹⁸ Metadata Policy, Strategy, Use and Sharing Advisory Group

¹⁹ Quality Assurance and Standards Working Group

²⁰ Rights and Access Working Group

výsledky svého digitalizačního projektu, protože samotné zájmy společnosti jsou jiné, než zájmy a poslání knihoven. Jedním z hlavních cílů knihoven je poskytovat uživatelům přístup k informacím a dokumentům, naproti tomu zájmy společnosti Google jsou spíše komerčního rázu.

Vzniku partnerství s názvem HathiTrust předcházela existence tzv. *Sdíleného digitálního repozitáře*²¹, první informace o jeho vývoji se na, v té době stejnojmenném webu objevily 11. dubna 2008. Podle tohoto příspěvku byla právě finalizována jednání o spolupráci s Komisí pro institucionální spolupráci (dále CIC)²² a zároveň byly osloveny další instituce mimo CIC (UNIVERSITY OF MICHIGAN, 2008b). Jako HathiTrust bylo partnerství poprvé nazváno v příspěvku z 11. června 2008, šlo o čtvrtý pravidelný měsíční příspěvek (UNIVERSITY OF MICHIGAN, 2008a). Oznámení o spuštění nového webu HathiTrust²³ bylo zveřejněno 12. září roku 2008 (UNIVERSITY OF MICHIGAN, 2008). Jako oficiální doba vzniku se uvádí říjen roku 2008. Zakládajícími členy byly členské knihovny CIC, Kalifornská univerzita, Indianská univerzita a Michiganská univerzita. Záštitu nad celým projektem měla a dosud má poslední jmenovaná univerzita, kde HathiTrust také sídlí (YORK, 2012).

1.3.1 Projekty

V průběhu vývoje partnerství HathiTrust a jeho digitální knihovny vznikaly různé dílčí služby a projekty. Některé jsou již ukončeny (úspěšně či neúspěšně), některé stále běží. V následujícím přehledu jsou stručně popsány ty nejvýznamnější²⁴. Finanční zajištění těchto projektů a programů bylo řešeno buď z rozpočtu partnerství HathiTrust nebo z nejrůznějších grantů.

Osířelá díla

Projekt Osířelá díla²⁵ (dále pouze OWP) byl ve formě pilotního projektu spuštěn v květnu roku 2011. Jeho cílem bylo identifikovat osířelá díla v HTDL a následně je zpřístupnit v plném textu svým členským institucím (UNITED STATES. COURT OF APPEALS (2ND CIRCUIT), 2014, s. 8-9). OWP od svého vzniku velmi úzce souvisí s jiným projektem

²¹ Shared Digital Repository

²² Committee on Institutional Cooperation

²³ <https://www.hathitrust.org/>

²⁴ Podle názoru autorky

²⁵ HathiTrust Orphan Works Project

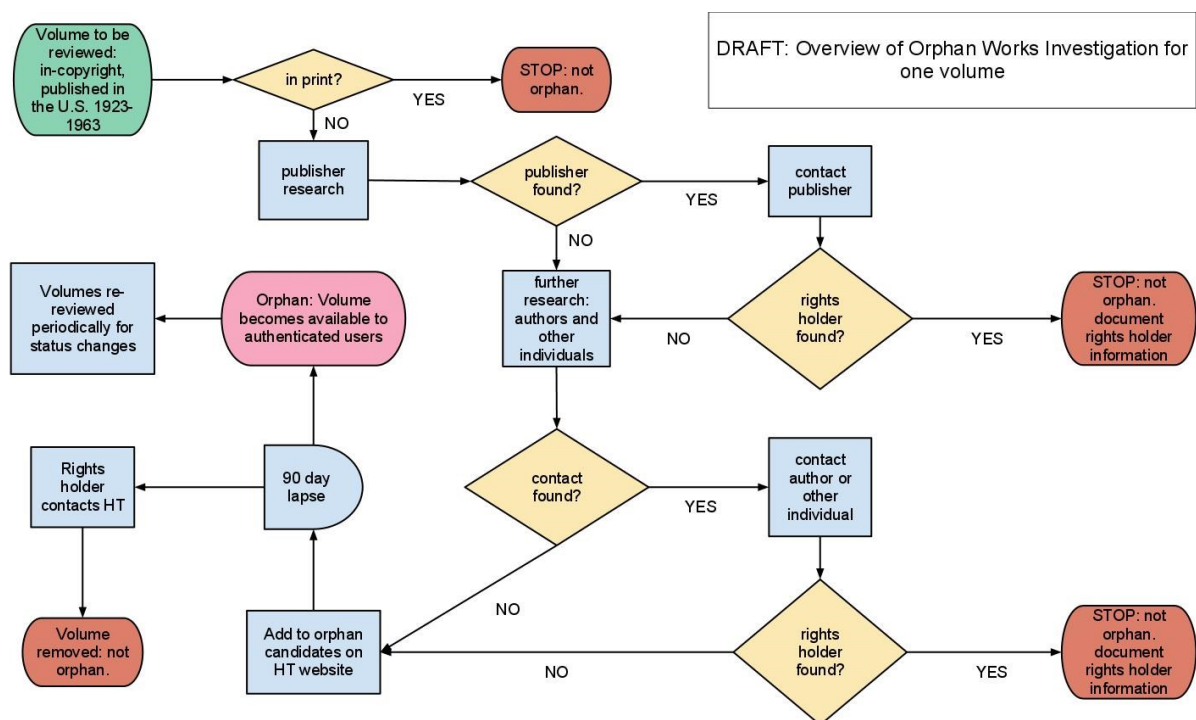
HathiTrust, s CRMS²⁶, který je popsán [níže](#). Celý proces spočíval v tom, že pomocí CRMS bylo určeno, zda jde o dílo (knihu) chráněné autorským zákonem (tzv. in copyright) a poté byl proveden výběr knih, které byly publikovány mezi lety 1923-1963²⁷ (UNIVERSITY OF MICHIGAN, 2014a).

Následný proces zjišťování a potvrzování, zda se skutečně jedná o osiřelé dílo, byl velmi zdoluhavý a náročný. Všichni pracovníci se řídili jednotným postupem, který je zveřejněn na webu Michiganské univerzity formou schématu (viz [Obrázek 2](#)). Prvním krokem bylo zjištění, zda je tištěný dokument stále dostupný ke koupi, pokud ano, je vyřazen z dalšího postupu ve schématu a ověřuje se další dílo. Pokud již kniha není dostupná (je tzv. „*out-of-print*“, zjišťuje se její vydavatel. Pakliže není možné vydavatele knihy kontaktovat (nakladatelství již neexistuje apod.), probíhá další vyhledávání, tentokrát autora nebo jiných spřízněných osob. Pokud je v některém z těchto kroků identifikován držitel autorských práv k příslušnému dílu, proces se přerušuje a pokračuje se od začátku dalším dílem. Jestliže však není držitel autorských práv k dílu nalezen, je toto zařazeno do seznamu kandidátů na osiřelá díla²⁸ na webu HathiTrust. Tento seznam byl vytvořen a spravován jako samostatná sbírka (viz níže kapitola o [sbírkách](#), které lze v rozhraní HTDL vytvářet) a po ukončení projektu byl odstraněn. Od zařazení do seznamu začala ubíhat devadesátidenní lhůta, kdy měl držitel autorských práv možnost se přihlásit a svoje dílo tak vyřadit ze seznamu kandidátů na osiřelá díla. Pokud by se držitel autorských práv v této lhůtě neozval, byl by plný text díla zpřístupněn členským institucím HathiTrust (po autentifikaci uživatele). I po zpřístupnění dokumentu bylo pomocí periodicky se opakujících kontrol a ověřování zjišťováno, zda se status dokumentu nezměnil (UNIVERSITY OF MICHIGAN, 2013a).

²⁶ Copyright Review Management System

²⁷ Projekt byl rozdělen do několika fází. Pro první fázi byly určeny právě knihy, chráněné autorským zákonem a vydané mezi lety 1923-1963.

²⁸ Orphan Work Candidate



Obrázek 2: Schéma ověřování osiřelých děl (UNIVERSITY OF MICHIGAN, 2013)

Tento projekt byl jedním z důvodů soudního sporu partnerství HathiTrust s „cechem“ autorů (Authors Guild), viz kapitola o problematice [autorských práv](#). Projekt byl ukončen v roce 2012, ještě než došlo ke konečnému rozhodnutí soudu. Na základě tohoto projektu nebyl zpřístupněn plný text ani jednoho dokumentu (UNITED STATES. COURT OF APPEALS (2ND CIRCUIT), 2014, s. 8-9).

System pro správu metadat Zephir

Zephir je název pro systém pro správu metadat, který byl vyvinut přímo pro digitální knihovnu HathiTrust. Stejně jako v případě HathiTrust byl název systému zvolen poměrně zábavným způsobem. Zephir je totiž jméno postavy (opice), která doprovází slona Babara ve francouzské pohádce Jean de Brunhoffové s názvem *Babar*. Tato skutečnost znázorňuje vztah HathiTrust a systému Zephir, jde o partnerský vztah, jehož obě součásti jsou dnes od sebe v podstatě neoddělitelné (STINE, 2013).

Z pohledu institucionální spolupráce v rámci partnerství HathiTrust jde o zásadní projekt, jelikož je vytvářen a dále spravován jinou organizací než je Michiganská univerzita (která do té doby zajišťovala veškeré služby a procesy, které se týkaly HTDL). Hlavním tvůrcem systému Zephir byla Kalifornská digitální knihovna²⁹ (CDL), s níž spolupracovali někteří profesionálové z Michiganské univerzity. Ostrá verze systému Zephir v rámci

²⁹ California Digital Library

HathiTrust byla spuštěna na konci října roku 2013, testována byla však již v průběhu roku 2012 (STINE, 2013). HathiTrust tímto systémem zároveň přispěl do Sítě pro ochranu digitálních dat (DPN)³⁰. Jde o kooperativní projekt, který spojuje nejrůznější vědecké a výzkumné instituce a pomáhá jim s dlouhodobou ochranou jejich digitálních a digitalizovaných dokumentů. Spolupráce spočívá ve společném vývoji a sdílení různých technologických a jiných infrastruktur, které zajišťují dlouhodobou archivaci (HathiTrust releases Zephir, 2013).

Projektový tým pro tvorbu tohoto systému měl celkem dvanáct členů, osm na Kalifornské a čtyři na Michiganské univerzitě. Jde o několikaletý projekt, který měl čtyři hlavní cíle:

- „Vytvořit systém pro správu metadat, který by odpovídal knihovnímu systému, který používá Michiganská univerzita (Aleph)
- Zefektivnit aktualizaci záznamů a celkově celý proces práce s metadaty
- Vytvořit takový systém, kde by bylo možné spravovat metadatové záznamy na více úrovních (dílo, vyjádření, jednotka, apod.)
- Zabezpečit správu metadat s hlediska verzování záznamů“ (UNIVERSITY OF MICHIGAN, 2016u).

Na udržování a správě systému Zephir pracují dva týmy, z nichž jeden působí na Kalifornské univerzitě (sedm členů) a druhý na Michiganské univerzitě (tři členové). Zephir zajišťuje procesy přijímání, zpracování, ukládání i export metadat (UNIVERSITY OF MICHIGAN, 2016u). Tento systém podporuje verzování (což bylo jedním z cílů projektu), shromažďuje tedy všechny dostupné verze metadatového záznamu digitálního objektu v HTDL. Někdy totiž mohou přicházet záznamy z více různých zdrojů a mohou se tedy od sebe navzájem lišit. Na základě speciálního algoritmu vybere Zephir nejlepší záznam, který je následně exportován pro použití v katalogu HTDL (HathiTrust releases Zephir, 2013). Informace o změnách a verzích záznamu jsou samozřejmě v systému dále uchovávány (více o funkcích systému Zephir v kapitole o [metadatech v HTDL](#)).

Vládní dokumenty

Projekt, zaměřený na vládní dokumenty Spojených států dostal název Program federálních vládních dokumentů Spojených států³¹. Byl zahájen v roce 2011 na shromáždění členů HathiTrust, kde bylo odhlasováno, že vznikne iniciativa, jejímž cílem bude co nejvíce

³⁰ Digital Preservation Network; více zde: <http://dpn.org/>

³¹ U.S. Federal Government Documents Program

rozšířit přístup k vládním dokumentům. Byla ustanovena poradní skupina, která měla za tento projekt zodpovědnost³² a zároveň na něj měla dohlížet. Hlavním výsledkem projektu je Registr vládních dokumentů Spojených států³³, který v současné době funguje v beta verzi³⁴ (UNIVERSITY OF MICHIGAN, 2016i).

Cílem *Registru vládních dokumentů Spojených států* je postihnout pokud možno všechny vládní dokumenty publikované od roku 1789 formou metadatových záznamů. Dalším krokem po vytvoření této databáze by byla postupná digitalizace těch dokumentů, které ještě v digitální podobě neexistují. Vývoj tohoto registru začal v roce 2013 a provázelo ho a stále provází mnoho problémů. Největším z nich je, že počet vládních dokumentů Spojených států prakticky není znám, existují odhady, které se pohybují mezi 1,5-3 miliony dokumentů. Z tohoto důvodu je velmi těžké vytvořit kompletní sbírku těchto dokumentů. V roce 2013 byla zveřejněna výzva knihovnám pro zasílání informací o vládních dokumentech z jejich fondů formou metadatových záznamů. Tímto způsobem bylo shromážděno kolem 26 milionů záznamů ze 43 institucí. Z tohoto obrovského objemu dat bylo nakonec vybráno cca 5,5 milionů záznamů (FURLOUGH, 2016a, s. 2-6). Prvním problémem, který v tomto procesu nastal, bylo odstranění duplicitních záznamů. K jejich identifikaci bylo použito identifikátoru SuDoc, kterým jsou označovány dokumenty publikované Vládní tiskovou kanceláří³⁵ (dále GPO). Všechny záznamy však tento identifikátor neměly, poněvadž ne všechny vládní dokumenty jsou vydávány touto institucí. Byly tedy použity také identifikátory OCLC z katalogu Worldcat, ISSN a ISBN. U duplicit bylo také nutné zkontrolovat, zda jde skutečně o stejný dokument. Bylo možné, že šlo o stejný dokument, ale v jiném formátu. V takových případech nebyl jeden ze záznamů odstraněn, ale oba byly sloučeny pod jeden záznam. Další skupinou záznamů, která byla z budoucí databáze vyloučena, byly záznamy, které obsahovaly málo informací, nebylo podle nich možné dokument jednoznačně identifikovat. Z tohoto důvodu bylo vyřazeno přes 5 milionů záznamů (GLENN, 2016).

Rozhraní *Registru vládních dokumentů Spojených států* je umístěno v rámci webu HathiTrust. Jde o beta verzi, na hlavní stránce je tedy umístěn formulář pro zpětnou vazbu od uživatelů. Databázi lze prohledávat pomocí jednoduchého vyhledávacího pole. Existuje zde možnost vyhledávat podle těchto údajů: název, autor, identifikátor SuDoc, identifikátor

³² HathiTrust Government Documents Planning and Advisory Group

³³ HathiTrust U.S. Government Documents Registry

³⁴ Registr je dostupný na této adrese: https://www.hathitrust.org/usdocs_registry/

³⁵ Government Printing Office

OCLC a vydavatel. Záznamy lze také procházet pomocí filtrů umístěných po levé straně obrazovky. Je možné vybrat si pouze záznamy s dostupným plným textem (dokumenty z HTDL), záznamy s omezeným zobrazením dokumentu (dokumenty z HTDL) nebo záznamy dokumentů z fondů institucí mimo partnerství HathiTrust. Další možnosti filtrování jsou podle formátu dokumentu, data vydání, autora a předmětu. Výsledky vyhledávání lze řadit podle relevance, roku vydání (sestupně či vzestupně) a abecedně podle autora nebo názvu dokumentu. Také lze nastavit počet záznamů zobrazených na jedné straně (10, 20, 50 nebo 100). Stručný záznam, který je zobrazen ve výsledcích vyhledávání obsahuje údaje o názvu, vydání (rok a vydavatel) a jménu autora. Dále je u něj uveden odkaz na kompletní záznam v Registru a případně také odkaz na plný text v digitální knihovně HathiTrust. Kompletní záznam dokumentu není příliš obsáhlý, obsahuje skutečně základní informace potřebné k identifikaci dokumentu (UNIVERSITY OF MICHIGAN, 2016s).

An Act to Designate the Facility of the United States Postal Service Located at 201 B Street in Perryville, Arkansas, as the "Harold George Bennett Post Office."		Tools
Title:	An Act to Designate the Facility of the United States Postal Service Located at 201 B Street in Perryville, Arkansas, as the "Harold George Bennett Post Office."	JSON
Author:	United States	Export to Refworks
Format:	Book, Online, Print	Export to EndNote
Publisher:	[Washington, D.C.] : [U.S. Government Publishing Office]	Email
Published:	2016	SMS This
SuDoc Call Number:	AE 2.110:114-170	Cite
OCLC #:	954462029 956312574	MARC View
Physical Description:	1 online resource ([1] leaf). [1] leaf ; 24 cm	
Electronic Resource:	http://purl.fdlp.gov/GPO/gpo70571	
Electronic Resource:	http://purl.fdlp.gov/GPO/gpo70572	
Feedback		

Obrázek 3: Záznam dokumentu v Registru vládních dokumentů Spojených států³⁶

Záznam (viz [Obrázek 3](#)) obsahuje název dokumentu, jméno autora, nakladatelské údaje, formát dokumentu, identifikátory (SuDoc, OCLC, ISSN, ISBN, apod.), fyzický popis a případně odkaz na umístění plného textu. V záznamu může být také informace o sérii nebo edici, ve které příslušný dokument vyšel. Stejně jako na hlavní straně je i pod každým záznamem formulář pro zpětnou vazbu. Na pravé straně obrazovky se nacházejí nástroje pro

³⁶ Záznam lze zobrazit na této adrese: https://www.hathitrust.org/usdocs_registry/catalog/2700eb53-31ae-4507-8eda-08ed1312e019

práci se záznamem. Lze jej zobrazit ve formátu JSON³⁷, exportovat do dvou různých citačních manažerů, odeslat prostřednictvím e-mailu nebo SMS zprávy, zobrazit pro záznam citaci (ve třech citačních stylech) a také zobrazit záznam ve formátu MARC. V případě příkladu (viz [Obrázek 3](#)) se zobrazí tři záznamy ve formátu MARC, ze kterých byl vytvořen záznam v Registru (UNIVERSITY OF MICHIGAN, 2016s).

Vývoj Registru samozřejmě stále probíhá, další aktivity v rámci Programu federálních vládních dokumentů Spojených států spočívají v identifikaci vládních dokumentů, které je třeba digitalizovat a následně v samotné digitalizaci a zpřístupnění. Registr je každý den aktualizován synchronizováním s katalogem HTDL a také s Katalogem vládních publikací GPO³⁸ (FURLOUGH, 2016a).

System pro správu a revizi autorsko-právních metadat

Účelem Systému pro správu a revizi autorsko-právních metadat je ověření, zda je daný dokument (z HTDL) stále chráněn autorským zákonem. Tento projekt byl spuštěn v roce 2008 a je financován z grantů, které poskytuje IMLS. Vývoj CRMS od jeho počátku do současnosti by se dal rozdělit do tří fází, z nichž každá měla nějaké specifické zaměření:

- 2008-2011: Monografie vydané ve Spojených státech mezi lety 1923-1963
- 2011-2014: Dokončení z předchozí fáze, monografie vydané ve Velké Británii před rokem 1943, monografie vydané v Kanadě a Austrálii před rokem 1963
- 2014-2016: Dokončení z předchozí fáze, vytvoření publikace nebo jiného nástroje, který vysvětlí principy CRMS a umožní ho použít také jiným institucím

Výsledkem zatím poslední fáze je volně dostupná publikace s názvem *Finding of public domain: Copyright Review Management System Toolkit*³⁹. V prvním roce spolupracovalo na projektu 22 institucí, na rok 2017 se jich přihlásilo 18 (UNIVERSITY OF MICHIGAN, 2016e).

Velký problém dokumentů, které již nejsou chráněny autorským zákonem (tzv. public domain dokumentů), spočívá v tom, že není úplně jasné, jak takový dokument spolehlivě identifikovat. Kdyby pro tuto činnost existoval nějaký jednoduchý nástroj nebo návod, mohly

³⁷ Jde o formát pro výměnu dat, který je založen na programovacím jazyku JavaScript, je univerzální a zcela nezávislý na jazyku. Více informací např. zde: <http://www.json.org/json-cz.html>

³⁸ GPO Catalog of Government Publications, katalog je dostupný zde: <https://catalog.gpo.gov>

³⁹ LEVINE, Melissa a et al. *Finding the public domain: Copyright Review Management System Toolkit* [online]. Ann Arbor: Board of Regents, University of Michigan, 2016, 304 s. [cit. 2016-11-29]. ISBN 978-1-60785-374-9. Pod licencí CC-by dostupné z: <http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001>.

by se tím značně rozšířit možnosti digitálních knihoven při zpřístupňování plných textů dokumentů. Výše uvedená publikace představuje v tomto ohledu zatím jediný takový prostředek. Je zaměřena primárně na knihy, tedy monografické publikace. Obsah této příručky by se dal rozdělit do tří částí. První část provede uživatele činnostmi spojenými se spuštěním nějakého projektu pro revizi autorsko-právních metadat. Konkrétně jsou zde uvedeny tipy, jak sestavit projektový tým (hlavně co se týče profesních specializací) a dále jak určit soubor dokumentů, jejichž status se bude prověřovat. Také lze s pomocí první části definovat metody, postupy a požadované cíle projektu. Druhá část knihy je zaměřena spíše prakticky a skutečně vychází z praktických zkušeností, získaných v průběhu projektu CRMS. Zahrnuje možnosti a postupy vedení projektu, technické údaje, základní autorsko-právní informace a také informace týkající se personálního zabezpečení. Třetí část obsahuje zprávy o různých pilotních projektech a podprojektech, na kterých se tým CRMS podílel (např. projekty, které se týkají [vládních dokumentů](#)⁴⁰, dokumentů vydaných ve Španělsku⁴¹, v Německu⁴², atd.). Přílohy této příručky obsahují různé modely postupů a kódovníky, slovník pojmů a také dva testy, ve kterých si lze ověřit získané znalosti (LEVINE, 2016, s. 3-4).

Ostatní projekty a programy

Dalšími projekty a programy, které v rámci HathiTrust vznikají, jsou například Program sdílení tištěných dokumentů⁴³, dále projekt, který se týká publikování časopisů s názvem mPach. Jeden z dalších projektů se zabývá dlouhodobou archivací obrazových dokumentů v rámci HTDL včetně doprovázejících metadat⁴⁴. Program sdílení tištěných dokumentů má za cíl zjednodušit a zlevnit management tištěných fondů členských institucí HathiTrust. Tohoto cíle chce dosáhnout vytvořením sdíleného souboru tištěných dokumentů, jejichž digitální kopie jsou obsaženy v HTDL. K těmto dokumentům by měly přístup všechny členské instituce a mohly by je poskytovat svým uživatelům (UNIVERSITY OF MICHIGAN, 2016p). Projekt s názvem mPach vytváří a spravuje Michiganská univerzita. Jeho cílem je vytvořit prostředí pro přímé publikování elektronických verzí časopisů v režimu Open Access. Tento soubor nástrojů bude propojen s digitální knihovnou HathiTrust, aby bylo možné tyto dokumenty bezprostředně po jejich vydání dlouhodobě archivovat

⁴⁰ US State Government Documents

⁴¹ Reviewing Works Published in Spain

⁴² Humboldt University of Berlin: Rights Research Project for German Books

⁴³ Shared Print Program

⁴⁴ Minnesota Digital Library Image Preservation Prototype Project

a zpřístupnit. Zde je důležitá problematika dlouhodobého archivování dokumentů ve formátu PDF, což zatím v HTDL není zajištěno. Výše zmíněný projekt Minnesotské univerzity na archivaci obrazových dokumentů probíhal v roce 2010 a bylo v rámci něj vytvořeno a popsáno několik typů obsahu, včetně jejich vzájemného propojení. Další projekty a programy, které doprovázely a doprovázejí vývoj HathiTrust jsou realizovány pomocí grantů často poskytovaných IMLS (UNIVERSITY OF MICHIGAN, 2016k).

1.3.2 Výzkumné centrum HathiTrust⁴⁵

Výzkumné centrum HathiTrust (HTRC) bylo spuštěno v roce 2011, kořeny jeho vzniku však sahají až do roku 2008. V tomto roce oznámila společnost Google, že by v rámci svého projektu Google Books chtěla podobný nástroj vytvořit. Měl sloužit pro analýzu velkých objemů textu jak dokumentů, které již nejsou chráněny autorským zákonem, tak těch, které pod ni stále spadají. Zároveň mělo být zajištěno, aby při analýze obsahu těchto dokumentů nedocházelo k porušení autorských práv. Tento záměr Google uvedl v Dohodě o vypořádání Google Books⁴⁶ v rámci soudního sporu s autory a organizacemi, které je zastupují (např. Authors Guild apod.). V roce 2011 soud tuto dohodu zamítl a Google tedy k vytvoření podobného nástroje nikdy nepřistoupil. Správní rada HathiTrust si však tohoto záměru v původní dohodě všimla a rozhodla se ho realizovat v podobě HTRC. V roce 2009 byla mezi členy HathiTrust vznesena žádost o návrhy, jak by mohlo takové centrum vypadat a jaké by mělo mít funkce (DOWNIE, 2016). Vítězný návrh vyšel ze spolupráce Knihoven Indianské univerzity⁴⁷ a Illinoiské univerzity v Urbana-Champaign, které HTRC spravují a nejvíce se podílejí na jeho vývoji (BONN, 2016).

Poněvadž jde o velmi rozsáhlý projekt, má v rámci HathiTrust svou vlastní malou organizační strukturu. Skládá se ze dvou částí, Týmu výkonného managementu⁴⁸ a Poradního sboru⁴⁹. Tým výkonného managementu má pět členů, po dvou z Indianské a Illinoiské univerzity a jednoho z Virginské univerzity. Tento tým podává zprávy o stavu a vývoji projektu Správní radě HathiTrust. Poradní sbor je složen z deseti členů, kterými jsou výzkumníci, vědci a administrátoři z širokého spektra oborů. Jejich úkolem je poskytovat poradenství Týmu výkonného managementu (UNIVERSITY OF MICHIGAN, 2016j).

⁴⁵ HathiTrust Research Center

⁴⁶ Google Books Settlement Agreement

⁴⁷ Indiana University Libraries

⁴⁸ Executive Management Team

⁴⁹ Advisory Board

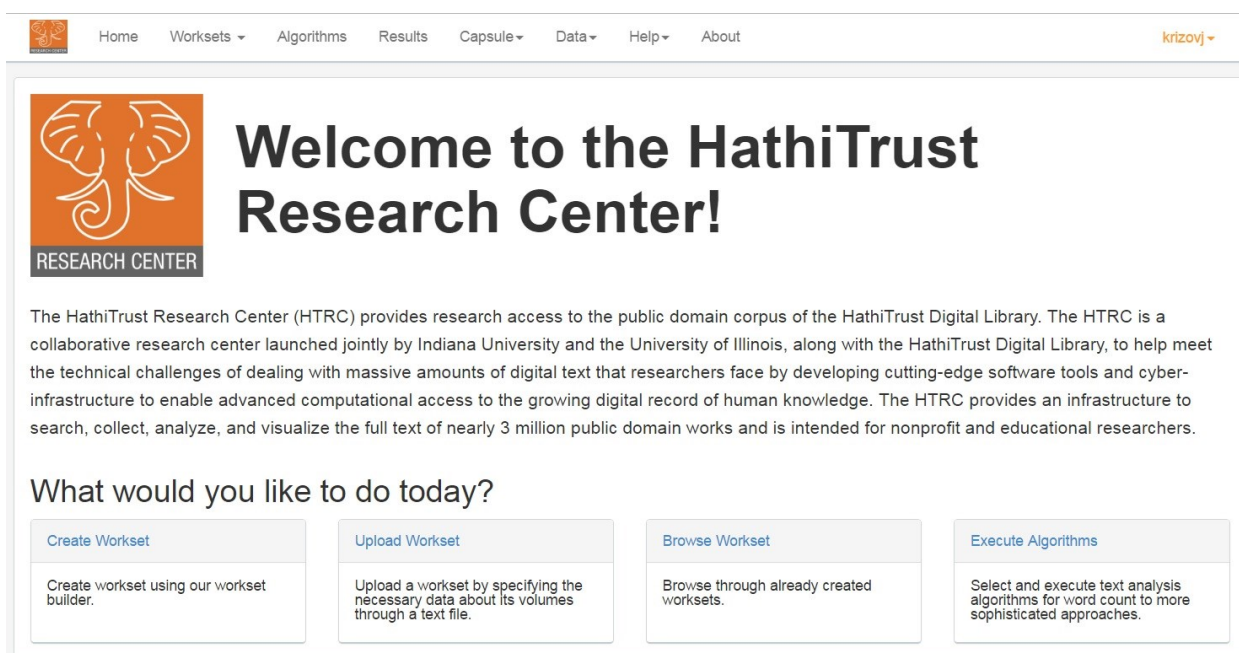
Obsah HTRC není totožný s obsahem HTDL, zatím je zde umístěno přes tři miliony děl z HTDL. Pomocí nástrojů v HTRC je možné analyzovat obsah těchto dokumentů na úrovni textu. Pro tyto nástroje je vytvořena podrobná a rozsáhlá dokumentace, navíc je možné navštívit jeden z workshopů, které HTRC pro své uživatele pořádá. V současné době poskytuje HTRC uživatelům přístup k obsahu dokumentů, které již nejsou pod ochranou autorského zákona ve formě OCR a také k jejich záznamům ve formátu MARC a METS (INDIANA UNIVERSITY LIBRARIES, 2016). Pro používání HTRC je nutná registrace, která je umožněna, stejně jako v případě digitální knihovny HathiTrust, i nečlenům tohoto partnerství. Protože je tento nástroj určen vědcům, akademickým pracovníkům a studentům, je nutné mít institucionální emailovou adresu univerzity nebo výzkumné instituce.

Uživatelské rozhraní HTRC⁵⁰

Uživatelské rozhraní HTRC je v současnosti dostupné ve verzi 3.2.0. Z hlavní stránky (viz [Obrázek 4](#)) jsou dostupné všechny nástroje, které HTRC nabízí. Nejprve je nutné definovat sadu dokumentů (tzv. workset), nad kterými bude probíhat následná analýza. Tuto sadu lze vytvořit přímo z webu HTRC pomocí vyhledávacího nástroje s názvem HTRC Workset Builder (funkce *Create Workset*), nahrát ji ve formátu CSV nebo TXT (funkce *Upload Workset*) nebo použít některou z těch, které vytvořili ostatní uživatelé (funkce *Browse Workset*). Pomocí funkce *Execute Algorithms* je možné použít na tyto sady dokumentů jeden z předem připravených algoritmů. V současné době lze použít jedenáct různých algoritmů, vytvořených celkem čtyřmi pracovníky HTRC. Jedním z algoritmů je možné například určit nejfrekventovanější slova v určené sadě dokumentů a počet jejich výskytů. Jiný je schopen určit témata v určité sadě, na základě blízkosti a pravděpodobnosti výskytu určitých slov. V záložce *Results* lze prohlížet výsledky všech aktuálně probíhajících i ukončených zkoumání, která se vážou k danému uživatelskému účtu. Funkce *Capsule* má umožňovat analýzu plných textů dokumentů, které jsou stále pod ochranou autorského zákona, aniž by toto právo bylo porušováno. Spuštění této služby se plánuje na rok 2017, zatím je možné si ji vyzkoušet na dokumentech, které již pod ochranou autorského zákona nejsou. Pro tuto funkci

⁵⁰ Uživatelské rozhraní HTRC je dostupné na adrese: <https://analytics.hathitrust.org/>

(stejně jako pro téměř všechny ostatní) je nutné se v HTRC zaregistrovat. Jde v podstatě o to, že se uživatel připojí k virtuálnímu počítači, je tedy dále potřeba nainstalovat nějaký VNC (Virtual Network Computing) program. Dalším krokem je vytvoření samotného vzdáleného počítače (capsule). Práce v tomto prostředí je možná ve dvou typech přístupu, v obslužném a chráněném. V obslužném módu je možné dostat se prostřednictvím internetu kamkoliv s výjimkou HTRC, naopak v chráněném rozhraní je možné prohlížet obsah HTRC a několik málo dalších stránek, ale již se nelze připojit k jiným webům. Záložka *Data* umožňuje uživateli prohlédnout či stáhnout dva již připravené datasety. Dále je pod touto funkcí možné vizualizovat frekvenci použití určitých slov pomocí aplikace Bookworm. Jsou zde zahrnuty texty z Digitální knihovny HathiTrust, které již nejsou pod ochranou autorského zákona. Tato služba je dostupná bez nutnosti registrace. Záložka *Help* obsahuje velmi podrobnou



Home Worksets Algorithms Results Capsule Data Help About krizovj

Welcome to the HathiTrust Research Center!

The HathiTrust Research Center (HTRC) provides research access to the public domain corpus of the HathiTrust Digital Library. The HTRC is a collaborative research center launched jointly by Indiana University and the University of Illinois, along with the HathiTrust Digital Library, to help meet the technical challenges of dealing with massive amounts of digital text that researchers face by developing cutting-edge software tools and cyber-infrastructure to enable advanced computational access to the growing digital record of human knowledge. The HTRC provides an infrastructure to search, collect, analyze, and visualize the full text of nearly 3 million public domain works and is intended for nonprofit and educational researchers.

What would you like to do today?

- Create Workset**
Create workset using our workset builder.
- Upload Workset**
Upload a workset by specifying the necessary data about its volumes through a text file.
- Browse Workset**
Browse through already created worksets.
- Execute Algorithms**
Select and execute text analysis algorithms for word count to more sophisticated approaches.

Obrázek 4: Hlavní stránka HTRC

a rozsáhlou nápovědu a dokumentaci k projektu HTRC (INDIANA UNIVERSITY LIBRARIES, 2016a).

1.4 Členství v HathiTrust

V současné době má partnerství HathiTrust pět institucionálních členů (jde o čtyři státní systémy vzdělávacích institucí a CIC) a sto čtrnáct (informace k 17. 10. 2016) členů

individuálních⁵¹. Ve všech případech jde o univerzity nebo univerzitní knihovny, a jejich uskupení a systémy (UNIVERSITY OF MICHIGAN, 2016m).

1.4.1 Podmínky přijetí

V případě zájmu o členství v partnerství HathiTrust je nutné splnit řadu vstupních požadavků, na jejichž základě je instituce přijata či odmítnuta. Některé tyto informace se také používají pro výpočet pravidelného ročního příspěvku, který bude příslušná instituce platit. Prvním krokem je samozřejmě vyplnění přihlášky (UNIVERSITY OF MICHIGAN, 2016l). Dalším požadavkem je zaslání informací o dokumentech ve fondu příslušné knihovny. Povinnou součástí těchto informací tvoří identifikátor OCLC kvůli předávání záznamů mezi OCLC a HathiTrust. Dále je nutné poskytnout informaci o interním systému identifikátorů (signatury a čárové kódy) v konkrétní knihovně. Ostatní údaje jsou vyžadovány, pokud jsou dostupné a ne vždy u všech typů dokumentů. Jde o status jednotky (ztraceno, na místě apod.), který je vhodné uvést u jedno a vícesvazkových monografií. U těchto typů dokumentů lze také uvést (pokud je k dispozici) popis fyzického stavu. Pokud existuje u vícesvazkových monografií nějaké speciální číslování, je vhodné ho v následujícím bodě také uvést. V případě seriálových publikací je vyžadován identifikátor ISSN (pokud je k dispozici). Poslední položkou této části žádosti o zařazení do HathiTrust, je informace, zda jde o vládní dokument či nikoliv. Všechny požadované údaje musí být následně doplněny souhlasem s tím, že tyto údaje budou poskytnuty OCLC (UNIVERSITY OF MICHIGAN, 2016n). Je také nutné vybrat zástupce žádající organizace, který ji bude v rámci HathiTrust reprezentovat (např. na různých setkáních apod.). Kvůli autentizaci a chráněné komunikaci mezi institucemi musí mít potenciální člen přístup k autentizační službě Shibboleth. Pokud ho žádající organizace nemá k dispozici, Michiganská univerzita může poskytnout pomoc při jeho instalaci i správě. Samozřejmostí každé žádosti je také fungující kontaktní e-mail, který je zařazen do seznamu adres, kam jsou rozesílána oficiální vyjádření a informace. Žádající organizace by také měla o svém zařazení do HathiTrust vydat tiskovou zprávu, pro kterou je dokonce dostupná šablona⁵². Součástí této tiskové zprávy je také krátký popis přístupující knihovny, ovšem HathiTrust vyžaduje tento popis zaslat také zvlášť. Pro případy, kdy by nefungoval primární kontakt, je nutné uvést ještě nějaký náhradní. HathiTrust sleduje přístup a různé aspekty služby pomocí nástroje Google Analytics. Pokud by měla knihovna zájem se k těmto statistickým informacím dostat, je potřeba uvést kontakt na Gmail. Členské instituce také mají

⁵¹ Seznam všech aktuálních členů HathiTrust: <https://www.hathitrust.org/community>

⁵² Šablona pro tiskovou zprávu je dostupná z webu HathiTrust: <http://bit.ly/eGDYNO>

možnost zapojit se do speciální skupiny v rámci HathiTrust, která má na starosti použitelnost HTDL z hlediska uživatelské zkušenosti. Pro identifikaci, zda uživatel přistupuje do HTDL z členské instituce, je potřeba poskytnout informace o IP adresách dané organizace (UNIVERSITY OF MICHIGAN, 2016l).

Od roku 2013 funguje nový systém výpočtu členských příspěvků. Za tímto účelem byly vytvořeny dva vzorce, které určují výši ročního poplatku pro každého člena. Ten je složen ze dvou částí. První je příspěvek na podporu správy dokumentů, které nespádají pod ochranu autorského zákona (tzv. public domain dokumenty). Druhou částí poplatku je příspěvek na podporu správy dokumentů, které jsou stále chráněny autorským zákonem. U těchto dokumentů však záleží na jejich počtu v HathiTrust a ve fondech příslušné instituce. Do vzorce se uvádí počet dokumentů, které se v obou úložištích překrývají. Pro výpočet první části příspěvku je používán tento vzorec:

$$(PD * X * C) / N$$

Člen	Vysvětlení
PD	Počet dokumentů v HathiTrust, které nespádají pod ochranu podle autorského zákona
X	Jde o flexibilní hodnotu, která je určována Správní radou HathiTrust, jejímž účelem je podpořit celkový rozvoj HathiTrust (zavádění nových služeb, zajištění funkcionality celého systému apod.)
C	Náklady na dlouhodobé uložení dokumentů, je určován podle počtu dokumentů, které bude chtít daná instituce uložit v nadcházejícím roce
N	Celkový počet členských institucí v HathiTrust

Tabulka 1: Výpočet členského příspěvku v HathiTrust, první část

Druhá část příspěvku se odvozuje z následujícího vzorce:

$$IC = (C * X) / H$$

Člen	Vysvětlení
IC	Počet dokumentů chráněných autorským zákonem, které se v HathiTrust a ve fondu příslušné instituce překrývají (spadají sem i dokumenty, které instituce měla ve fondech dříve)

H	Počet členů HathiTrust, kteří vlastní stejný dokument jako daná instituce
Ostatní členy (C a X) jsou stejné jako u předchozího vzorce.	

Tabulka 2: Výpočet členského příspěvku v HathiTrust, druhá část

Členské příspěvky jsou vypočítávány každý rok v lednu na rok následující (UNIVERSITY OF MICHIGAN, 2016f).

1.4.2 Výhody členství

Členství v partnerství HathiTrust s sebou pro instituci nese řadu výhod a benefitů. Jednou z hlavních výhod je možnost dlouhodobé archivace digitálních dokumentů dané instituce, aniž by musela vytvářet příslušnou infrastrukturu, technická řešení apod. Zároveň partnerství HathiTrust poskytuje uživatelské prostředí pro přístup k těmto dokumentům. Co se týče přístupu k plným textům, není možné ani členským institucím poskytnout přístup k plným textům, které jsou stále pod ochranou autorského zákona. Výjimku tvoří uživatelé s tzv. print disabilities, jde o uživatele, který trpí handicapem, který mu neumožňuje používat tištěnou verzi dokumentu. V takových případech může HathiTrust těmto uživatelům členských institucí s určitým omezením zpřístupnit plné texty dokumentů, které jsou stále pod ochranou autorského zákona (viz podkapitola s názvem [Cíle a funkce](#)). Dalším přínosem členství je možnost využívat a pomáhat s rozvojem nástrojů, projektů a systémů (viz výše podkapitola o [projektech](#)), ke kterým by se jednotlivé samotné instituce nemuseli dostat (YORK, 2012).

1.5 Technické parametry a standardy HTDL

Digitální knihovna HathiTrust nefunguje na základě jednoho konkrétního softwaru, jde o několik aplikací různých verzí, z nichž byla vytvořena platforma přímo pro tuto digitální knihovnu. Většinou jde o Open source software, což znamená, že je jiní uživatelé mohou využívat a případně tento zdrojový kód měnit podle svých potřeb. Patří mezi ně například programovací jazyky Perl a Java, program pro tvorbu databází MySQL, pro zajištění možnosti vyhledávat z textu je použit software Apache Solr a další. Výsledná skupina aplikací a programů, které tvoří HTDL je také open source, ovšem s tím rozdílem, že tento zdrojový kód je dostupný pouze členským institucím (UNIVERSITY OF MICHIGAN, 2016r).

Digitální knihovna HathiTrust je držitelem certifikátu důvěryhodného repozitáře s názvem TRAC⁵³, který jí byl udělen Centrem výzkumných knihoven (CRL)⁵⁴. Certifikát byl udělen na základě auditu, který probíhal v období listopad 2009 – prosinec 2010. Při auditu jsou sledovány tři základní kategorie: Organizační infrastruktura, Management digitálních objektů a Technologie, technická infrastruktura a zabezpečení. V každé kategorii je repozitáři udělen určitý počet bodů (1 nejméně až 5 nejvíce). HTDL dosáhla v první kategorii dvou bodů, ve druhé tří a v poslední kategorii čtyř bodů. Konkrétní sledovaná kritéria a hodnoty jsou určeny Kontrolním seznamem pro auditu a certifikace důvěryhodných repozitářů⁵⁵ (CENTER FOR RESEARCH LIBRARIES, 2011).

⁵³ Trustworthy Repositories Audit and Certification

⁵⁴ Center for Research Libraries

⁵⁵ Trustworthy Repositories Audit and Certification Checklist

2. Možnosti přístupu k obsahu HTDL

Hlavním uživatelským rozhraním digitální knihovny HathiTrust je rozhraní dostupné z webových stránek <https://www.hathitrust.org/> (viz [Obrázek 5](#)). K tomuto rozhraní je také k dispozici jeho mobilní verze. Dále od roku 2011 funguje rozhraní, vyvinuté ve spolupráci s OCLC s názvem HathiTrust Prototype Catalog, které je vyvinuto na základě systému Worldcat Local⁵⁶. Jde zatím o prototyp, který bude ještě dále vyvíjen. Záznamy dokumentů z HTDL jsou také dostupné všem členským institucím partnerství HathiTrust v jejich vlastních katalozích a discovery systémech, pokud o to mají zájem. Tímto způsobem jsou záznamy z digitální knihovny HathiTrust dostupné například v katalogu Knihovny Michiganské univerzity, který se jmenuje Mirlyn⁵⁷. Dalším příkladem je také discovery systém Lens⁵⁸, který vytváří a spravuje Chicagská univerzita, a kde jsou dostupné záznamy dokumentů z digitální knihovny HathiTrust, které již nejsou pod ochranou autorského zákona. Obsah HTDL lze také prohlížet prostřednictvím online katalogů Knihovny Univerzity Johna Hopkinse⁵⁹, Australské národní knihovny⁶⁰, Systému knihoven Indické univerzity⁶¹ nebo například pomocí discovery systému Bostonské univerzity s názvem Holmes⁶² a dalších (UNIVERSITY OF MICHIGAN, 2016o).

HTDL má dvě úrovně bibliografického popisu. Jednou z nich je dokument. Těmito dokumenty jsou jednosvazkové monografie, ale také časopisy a vícesvazkové monografie, například slovníky, atlasy apod. Každý z nich má svůj hlavní záznam. Druhou úroveň popisu představuje jednotka. Každý dokument má k sobě připojenou minimálně jednu jednotku, může jich však být mnohem více. Například záznam periodika má k sobě připojeny záznamy jednotek, jimiž jsou obvykle jednotlivé ročníky. Stejným způsobem jsou popsány vícesvazkové monografie. Výše uvedené způsobuje, že HTDL obsahuje odlišný počet jednotek a dokumentů.

V následující kapitole je popsáno originální rozhraní HTDL, registrace uživatelů, možnosti vyhledávání, obsah záznamů dokumentů, rozhraní pro prohlížení plných textů a funkce vytváření sbírek. Dále je stručně představena také mobilní verze HTDL prostřednictvím malé analýzy jejich funkcí v různých internetových prohlížečích na několika

⁵⁶ Dostupné zde: <http://hathitrust.worldcat.org/>

⁵⁷ <https://mirlyn.lib.umich.edu/>

⁵⁸ <https://catalog.lib.uchicago.edu/vufind/>

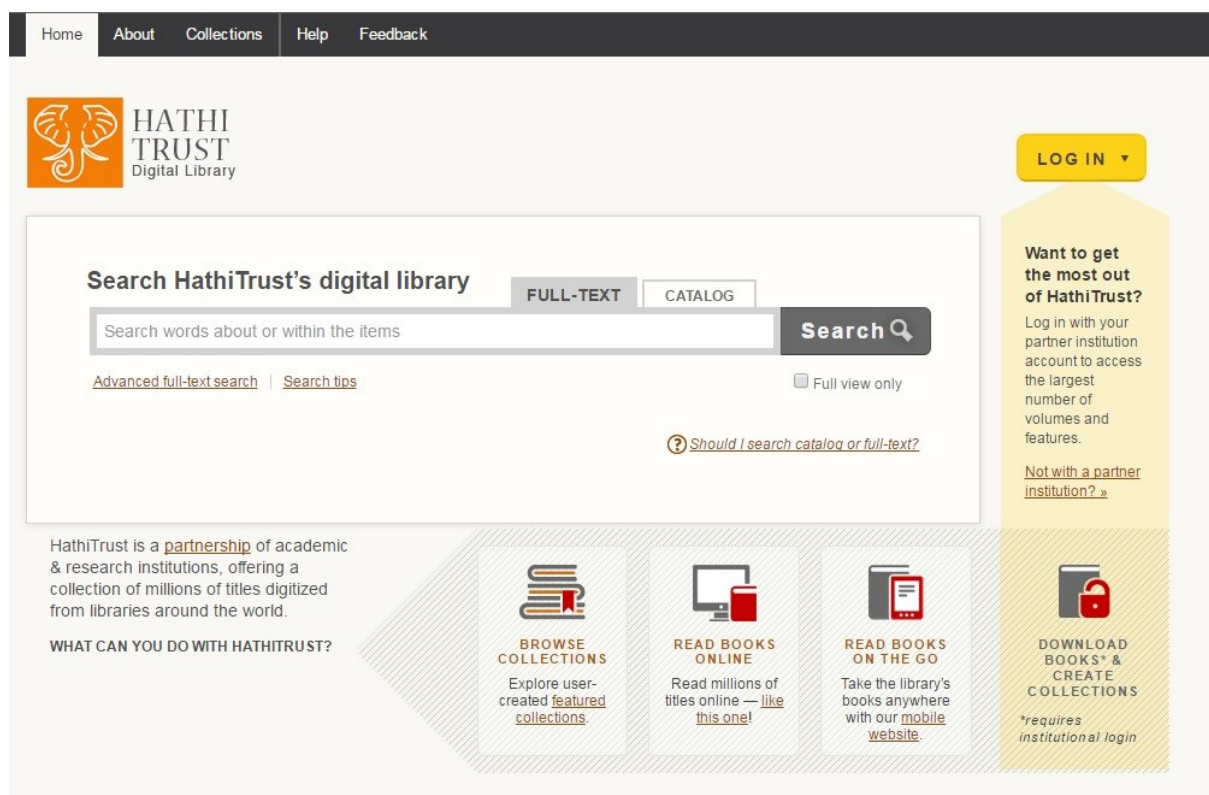
⁵⁹ <https://catalog.library.jhu.edu/>

⁶⁰ <http://catalogue.nla.gov.au/>

⁶¹ <http://www.iucat.iu.edu/>

⁶² http://bc-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vid=bclib

zařizování. Další část kapitoly popisuje metadata v HTDL, jejich pohyby, formáty a standardy. Poslední část je věnována ukládání a ochraně dokumentů a problematice autorských práv včetně přístupu k plným textům dokumentů.



Obrázek 5: Hlavní strana webových stránek Digitální knihovny HathiTrust

2.1 Analýza uživatelského rozhraní HTDL - Online katalog

Analýza uživatelského rozhraní HTDL byla vytvořena na základě praktických zkušeností a všechny popsané funkce byly v rozhraní skutečně vyzkoušeny.

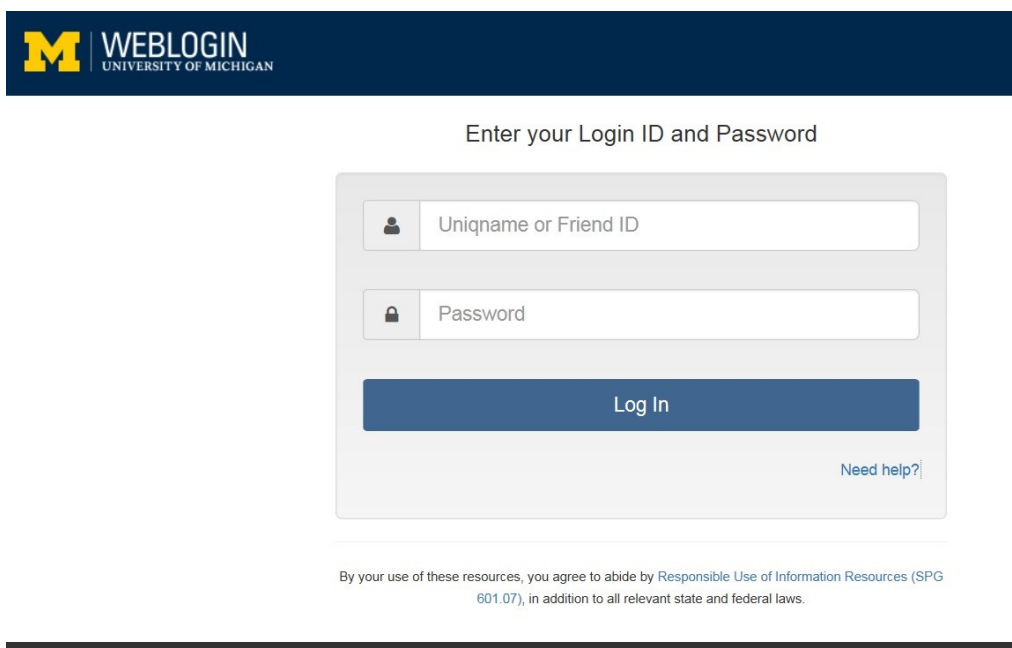
2.1.1 Registrace uživatelů

Zaregistrovat se v Digitální knihovně HathiTrust může každý, není nutné patřit pod některou z členských institucí. Hlavní výhodou registrace je možnost vytvářet sbírky dokumentů například podle tématu. Při registraci lze vybrat členskou instituci, pod kterou uživatel patří, nebo je možné získat spřátelený účet⁶³. Rozdíl spočívá v možnosti stahovat celé dokumenty ve formátu PDF, což je umožněno pouze uživatelům, kteří patří pod nějakou členskou instituci (a samozřejmě pouze pokud jde o volně dostupný dokument).

Registrace i přihlášení probíhá přes počítačovou síť Michiganské univerzity, která používá autentizační službu Shibboleth. Při přihlášení či registraci je tedy uživatel

⁶³ Friend account

přesměrován na webové stránky Michiganské univerzity (viz [Obrázek 6](#)) a po přihlášení zpět na web HathiTrust. Po vyplnění a odeslání registračního formuláře je na emailovou adresu uživatele zaslána automatická zpráva s odkazem pro aktivaci účtu. Po aktivování je odeslán ještě email s potvrzením vzniku uživatelského účtu s informacemi o změně či resetování přístupového hesla. Poté už je možné se kdykoliv přihlásit. Odhlášení uživatele z HTDL proběhne až po zavření internetového prohlížeče.



Obrázek 6: Přihlašovací stránka HTDL

2.1.2 Možnosti vyhledávání v HTDL

Hlavní vyhledávací rozhraní digitální knihovny HathiTrust poskytuje dvě možnosti vyhledávání. Lze vyhledávat podle údajů metadatových záznamů dokumentů (funkce *Catalog Search*) nebo z údajů metadatových záznamů a zároveň z plných textů dokumentů (funkce *Full-text Search*). Uživatel tedy vyhledává buď podle údajů o dokumentu, nebo podle slov z obsahu samotného dokumentu. Oba tyto vyhledávací způsoby jsou k dispozici ve formě jednoduchého i pokročilého vyhledávání. U obou typů vyhledávání lze použít Booleovské operátory (AND, OR), prostředky pro nahrazení znaků (*, ?) a uvozovky pro vyhledání přesné fráze. Pokud chce uživatel obsah digitální knihovny pouze procházet (tj. nehledá nic konkrétního), stačí do vyhledávacího pole zadat znak * (UNIVERSITY OF MICHIGAN, 2016h).

Vyhledávání bibliografických záznamů

V rámci jednoduchého vyhledávání je možné vyhledávat podle následujících údajů: název, autor, předmět, ISBN/ISSN, nakladatel, název edice. Vyhledávat lze i podle všech údajů najednou. V rámci pokročilého vyhledávání lze vyhledávat bibliografické záznamy podle čtyř polí naráz. Nelze přidat další vyhledávací pole jako v jiných vyhledávacích systémech⁶⁴. Tyto (maximálně čtyři) výrazy či fráze je možné spojit pomocí logických operátorů AND a OR. Dále lze vyhledávání limitovat pouze na dokumenty volně dostupné v plném textu, na dokumenty v určitém jazyce či podle formátu originálu. Samozřejmě zde nechybí možnost vyhledat dokumenty podle data vydání. Lze vyhledat záznamy dokumentů vydané před nebo během určitého roku, po nebo během určitého roku, mezi dvěma letopočty a také v rámci jednoho roku. Výsledkem tohoto typu vyhledávání jsou záznamy dokumentů (UNIVERSITY OF MICHIGAN, 2016).

Vyhledávání v plných textech

Jednoduché plnotextové vyhledávání sestává pouze z jednoho pole, kam je zapsán vyhledávací termín či fráze. Rozšířené vyhledávání v plných textech umožňuje vyhledávat podle stejných polí jako pokročilé vyhledávání bibliografických záznamů dokumentů, ke kterým přibývá navíc možnost vyhledávat podle slov v plných textech. Jednotlivá vyhledávací pole je opět možné kombinovat pomocí operátorů AND a OR, je zde navíc možnost zvolit si, podle kterých ze zadaných slov se má vyhledávat. Je možné vyhledávat podle všech zadaných slov, alespoň podle jednoho z nich, podle většiny těchto slov, nebo lze vyhledávat přesnou frázi. Dotaz je možné rozšířit pomocí další dvojice vyhledávacích polí. Obě tyto dvojice lze opět kombinovat pomocí logických operátorů AND a OR. Na rozdíl od způsobu vyhledávání bibliografických záznamů, je zde navíc také možnost vymazat všechny údaje zadané do vyhledávacího formuláře. Výsledkem tohoto typu vyhledávání jsou nikoliv záznamy dokumentů, ale záznamy jednotek. Na první pohled to tedy může působit, že se výsledky vyhledávání opakují, jde však o rozdílné jednotky stejných dokumentů (UNIVERSITY OF MICHIGAN, 2016a).

2.1.3 Výsledky vyhledávání

Jako výsledky vyhledávání jsou primárně zobrazeny zkrácené záznamy dokumentů, které obsahují název dokumentu, jméno autora a rok vydání. Součástí každého záznamu je

⁶⁴ Tuto funkci nabízí například rozšířené vyhledávání discovery systému Univerzity Karlovy s názvem [Ukaž](#), který je vytvářen společností EBSCO

také malý náhled titulní strany dokumentu. Z tohoto místa je také odkazováno na kompletní záznam (*Catalog Record*), druhý odkaz vede na plný text dokumentu (*Full View*), pokud je k dispozici. V případě že není, vede odkaz na funkci vyhledávání textů dokumentů s omezeným zobrazením výsledků (*Limited (search-only)*). V těchto výsledcích je dostupný pouze údaj o počtu nálezů hledaného termínu a také o jeho lokalizaci (na kterých stranách je uveden). Zobrazení výsledků plnotextového vyhledávání se mírně liší od zobrazení výsledků v případě vyhledávání bibliografických záznamů. V případě plnotextového vyhledávání nelze výsledky vyhledávání nijak seřadit (podle relevance, data, abecedy, apod.), lze si pouze zvolit, zda se mají zobrazit jen dokumenty s dostupným plným textem nebo všechny, u obou případů je samozřejmě uveden jejich počet. Při tomto způsobu vyhledávání je také možné měnit počet výsledků na stránku daného rozhraní a také je možné přidávat záznamy hromadně do sbírek. Naproti tomu výsledky vyhledávání v záznamech umožňují pouze řazení dle relevance, data a názvu, a zobrazení pouze těch dokumentů, které jsou dostupné v plném textu. V obou případech jsou v levém sloupci umístěny filtry, pomocí nichž lze dále přidávat vyhledávací kritéria. Výsledky je tedy možné dále filtrovat podle předmětu, jména autora, jazyka, místa vydání, data vydání, formátu a také podle umístění originálního dokumentu (ve fondu kterého člena HathiTrust se nachází). V obou typech vyhledávání je počet výsledků, které lze procházet, omezen na 20 000. Je tedy zobrazen počet výsledků (např. 65 000 záznamů), ale uživatel může postupně prohlédnout pouze prvních 20 000. K ostatním záznamům se lze dostat pomocí filtrů umístěných po levé straně výsledků vyhledávání.

2.1.4 Záznam dokumentu v HTDL

Záznamy dokumentů v HTDL mají dvě části. První je klasický bibliografický záznam a druhá by se dala charakterizovat jako informace o plném textu dokumentu. V druhé části záznamu může být buď plný text dokumentu, nebo pokud není tento uživateli z nějakého důvodu dostupný, alespoň možnost vyhledávání v jeho obsahu.

The screenshot shows the HathiTrust Digital Library interface. At the top, there are navigation links: Home, About, Collections, Help, Feedback. Below this is the HathiTrust logo and a search bar with 'FULL-TEXT' and 'CATALOG' buttons. The search bar contains the text 'Search words about or within the items'. To the right of the search bar is a 'LOG IN' button. Below the search bar are links for 'Advanced full-text search' and 'Search tips', and a checkbox for 'Full view only'.

The main content area displays a bibliographic record for 'Czech and Slovak literature in English : a bibliography / by George J. Kovtun.' The record includes the following fields:

- Main Author:** [Kovtun, George J.](#)
- Language(s):** English
- Published:** Washington : European Division, Library of Congress : 1988.
- Edition:** 2nd ed.
- Subjects:** [Slovak literature](#) > [Translations into English](#) > [Bibliography](#). [Czech literature](#) > [Translations into English](#) > [Bibliography](#).
- Note:** Includes indexes.
- Physical Description:** viii, 152 p. ; 23 cm.
- ISBN:** 0844405787
- Locate a Print Version:** [Find in a library](#)

Below the record is a 'Viewability' section with three links:

- [Full view](#) (original from Purdue University)
- [Full view](#) (original from University of Illinois at Urbana-Champaign)
- [Full view](#) (original from University of Michigan)

On the left side of the record, there is a 'Similar Items' section with several links to related bibliographies, such as 'Czech and Slovak literature in English : a bibliography / By: Kovtun, George J. Published: (1984)' and 'Tomáš G. Masaryk, 1850-1937 : a selective list of reading materials in English / By: Kovtun, George J. Published: (1981)'.

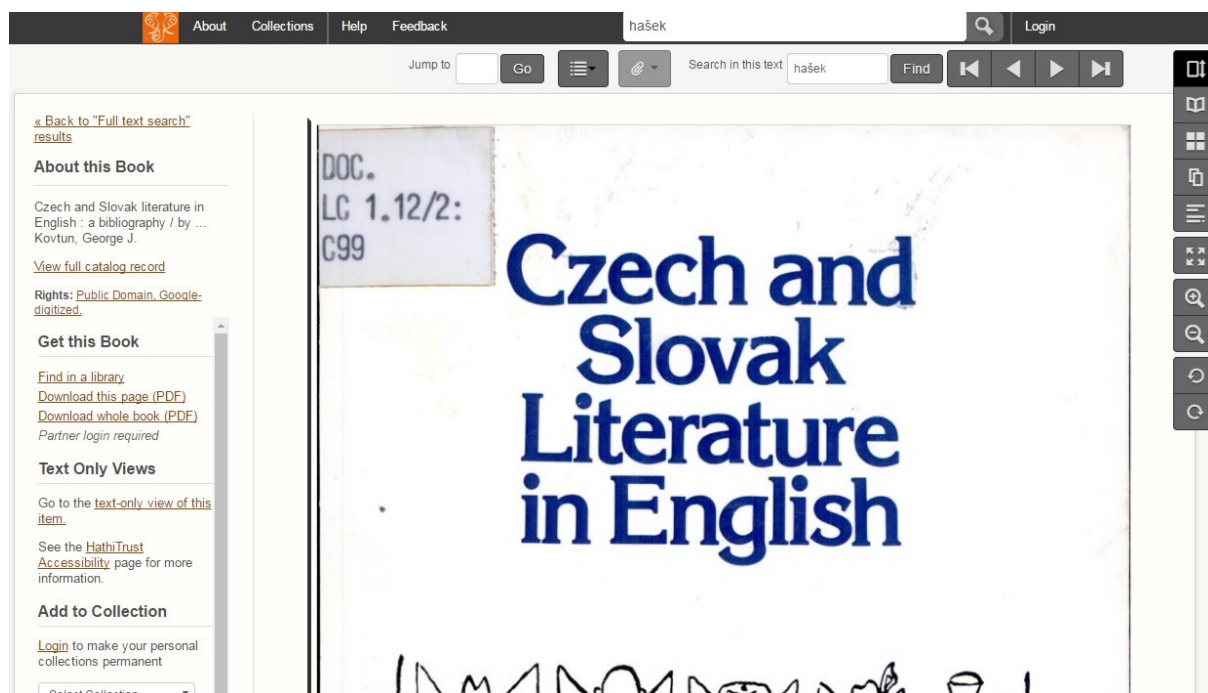
Obrázek 7: Bibliografický záznam dokumentu v online katalogu HathiTrust

Bibliografický záznam obsahuje následující údaje: název dokumentu, jméno autora, nakladatelské údaje, údaj o vydání, předmět (podle třídění Kongresové knihovny), poznámky, fyzický popis, identifikátor a odkaz na záznam do katalogu Worldcat, kde lze zjistit vlastníky dokumentu. Jméno autora funguje jako odkaz na seznam záznamů všech jím vytvořených dokumentů. Stejným způsobem fungují také názvy předmětových hesel. Pro každý záznam lze vygenerovat citační záznam ve dvou různých citačních stylech (APA a MLA) a dále je možné tento záznam stáhnout ve formátu ris. Bibliografický záznam dokumentu obsahuje také informace o jeho dostupnosti.

Pokud je dokument volně dostupný, je v záznamu umístěn odkaz na prohlížení plného textu a také informace o vlastníkově digitalizovaného dokumentu. Může dojít k situaci, kdy jsou v záznamu odkazy na více exemplářů stejného dokumentu. Příčiny těchto duplicit (multiplicit) je třeba hledat v počátcích digitalizačních projektů, kdy ještě nebyly zavedeny systémy jejich kontroly. Rozhraní pro prohlížení plných textů je rozděleno do dvou částí. V levém sloupci jsou uvedeny stručné údaje o zobrazovaném dokumentu, dále odkaz na kompletní bibliografický záznam a informace o právech. Další část levého sloupce zaujímají informace o získání dokumentu. Je zde umístěn odkaz do katalogu Worldcat a několik

možností stažení ve formátu PDF. Lze stáhnout aktuálně prohlíženou stranu, několik stran nebo celý dokument. Následující část levého sloupce umožňuje zobrazení dokumentu ve formě prostého textu, který je vhodný zejména pro uživatele se zrakovým postižením. Dále lze dokument přidat do vlastní sbírky, kterou si mohou vytvořit pouze přihlášení uživatelé. Sbírký a jejich funkcionality jsou popsány v následující části práce. Poslední oddíl levého sloupce tvoří možnosti sdílení dokumentu. Sdílet lze přes sedm různých sociálních sítí⁶⁵, pomocí permanentního odkazu buď na aktuálně prohlíženou stranu, či na celý dokument. Další možností jak dokument sdílet, je umístit jej přímo do HTML kódu jakékoliv webové stránky či blogu. Zde si uživatel musí zvolit, jak má prohlížeč rozhraní na jeho webu vypadat, může být buď scrollovací, nebo s otáčením stránek. V pravé části obrazovky je umístěno okno s náhledem dokumentu. Z rozložení obrazovky je vidět, že náhled dokumentu je považován za její nejdůležitější součást a všechny další funkcionality zabírají co nejméně místa. Náhledové okno má dvě lišty s různými funkcemi. Jedna je umístěna horizontálně v horní části okna a slouží k navigaci a orientaci v rámci dokumentu. Umožňuje přejít na konkrétní stránku, kapitolu, vyhledávat v plném textu, a přejít na následující / předchozí nebo první / poslední stránku. Druhá lišta leží na pravém okraji okna a její poloha je vertikální. Tento panel umožňuje pracovat s možnostmi zobrazení dokumentu. Dokument lze prohlížet scrollováním, otáčením listů, pomocí thumbnailů (malých náhledů), stránku po stránce či ve formě prostého textu. Samozřejmou funkcí je roztažení obrazu na celou obrazovku, přiblížení, oddálení a otáčení. Toto prohlížečské rozhraní je velmi intuitivní a uživatelsky přívětivé.

⁶⁵ Facebook, Twitter, Google+, reddit, Tumblr, VK a Pinterest



Obrázek 8: Rozhraní pro prohlížení plných textů dokumentů v HTDL

V případě, že dokument není volně dostupný, je v místě prohlížečského rozhraní pro plný text dokumentu vyhledávací okno, kam lze zadat vyhledávaný termín. Následně se objeví seznam čísel stran, na kterých se daný termín nachází. U jednotlivých čísel stran je uveden také údaj o počtu jeho výskytu na této stránce. Levý sloupec obsahuje stejné informace, jak v případě plného textu.

2.1.5 Funkce vytváření sbírek

Sbírka v HTDL by se dala definovat jako soubor dokumentů shromážděných uživatelem do jednoho celku a vybraných na základě nějakého kritéria. Kritérium lze uvést buď v názvu sbírky, nebo v jejím popisu. Svou vlastní sbírku si může v rámci HTDL vytvořit každý uživatel, dokonce i bez nutnosti přihlášení. V takovém případě je však možné vytvořit pouze soukromou sbírku pro účely vyhledávání, která není nikde uložena a po vypnutí prohlížeče již není dostupná. Naopak přihlášení uživatelé mohou vytvářet jak soukromé sbírky uložené pouze v uživatelském účtu, tak sbírky veřejně dostupné.

Většinou uživatelé vytvářejí sbírky na určité téma, což může být užitečné mnoha studentům i vědeckým pracovníkům či jiným uživatelům při odborných rešerších. V současné době je zveřejněno 2 855 sbírek⁶⁶ (UNIVERSITY OF MICHIGAN, 2017c). Velkou výhodou je možnost vyhledávat v plných textech dokumentů, které jsou v dané sbírce uloženy. Tato

⁶⁶ Údaj k 28. 5. 2017

funkce byla také částečně použita pro analýzu bohemikálních dokumentů v HTDL (viz [třetí kapitola](#)). V rámci veřejně dostupných sbírek je k dispozici také devatenáct sbírek vytvořených týmem HTDL.

Seznam sbírek lze seřadit abecedně podle názvu, abecedně podle vlastníka, podle data poslední změny (sestupně) a dále podle počtu dokumentů ve sbírce (od nejnižšího počtu po nevyšší a opačně). Lze také zobrazit pouze sbírky podle určitého počtu v nich obsažených záznamů, tedy od nejméně 25 záznamů až po nejméně 1000 záznamů. Podle aktuálního stavu mají sbírky 1 až 97 757 dokumentů⁶⁷. Po přidání tisíce položky se objeví upozornění, že od tohoto počtu dokumentů bude možné ve sbírce vyhledávat až nejméně po 48 hodinách. Toto zpoždění je způsobeno tím, že k takto objemné sbírce je nejprve potřeba vytvořit index (UNIVERSITY OF MICHIGAN, 2017c).

Samotné vytvoření sbírky je velmi jednoduché. Po přihlášení je možné ji vytvořit hned na několika místech. Jedním z nich je rozhraní pro prohlížení všech sbírek, kde se nachází funkce *Create new collection*. Vytvořit sbírku je také možné v záznamu každého dokumentu v možnostech přidání dokumentu do sbírky. Při zakládání sbírky uživatel pouze vyplní její název a popis, případně může nastavit její dostupnost, kterou je možné dodatečně kdykoliv změnit. Při fulltextovém vyhledávání lze do sbírky přidávat záznamy hromadně, v případě vyhledávání v katalogu to není možné a záznamy se musí přidávat po jednom. Rozhraní pro práci se sbírkou má dvě části. V levém sloupci se nacházejí popisné informace, tedy vlastník sbírky, její popis a status (soukromá nebo veřejná). Sbírkou lze sdílet prostřednictvím šesti různých sociálních sítí a permanentního odkazu. Také lze stáhnout metadata o dokumentech, které jsou v ní uloženy ve formátu .txt. Obsah takového souboru je popsán ve [třetí části práce](#), jelikož byl použit pro obsahovou analýzu HTDL.

2.1.6 Mobilní verze HTDL

Pro účely této analýzy byla mobilní verze HTDL vyzkoušena celkem na čtyřech zařízeních. Jedním byl tablet a ostatní tři byly různé mobilní telefony. Původně byl vybrán pouze tablet a mobilní telefon, ale vzhledem k časté nefunkčnosti některých prvků, byly přidány ještě dva různé telefony. Nebylo použito žádné zařízení od společnosti Apple, jelikož autorka práce neměla žádné takové k dispozici.

Tablet:

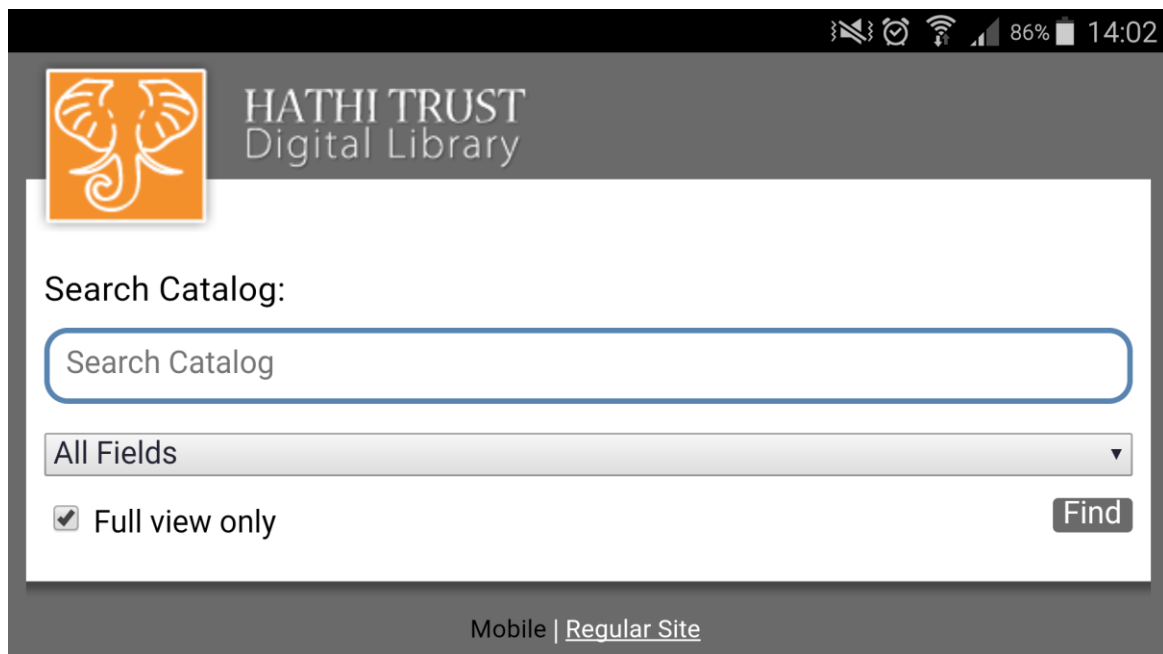
⁶⁷ Jde o sbírku, kde jsou umístěny všechny vládní dokumenty v HTDL

- Samsung SM-T230, Android 4.4.2

Mobilní telefony:

- Lenovo A536, Android 4.4.2
- Samsung S5, Android 6.0.1
- Huawei P8, Android 6.0.0

Mobilní rozhraní je oproti klasickému samozřejmě jednodušší. Sestává v podstatě z jednoho vyhledávacího okna pro katalogové vyhledávání, full-textové vyhledávání se zde nenabízí. Je možné vyhledávat podle názvu dokumentu, jména autora, předmětu, identifikátoru (ISBN nebo ISSN), nakladatele, názvu edice (série) a data publikování. Lze vyhledat pouze volně přístupné dokumenty zaškrtnutím možnosti „Full view only“.



Obrázek 9: Hlavní stránka mobilní verze HTDL

Výsledky vyhledávání jsou zobrazeny v podobě číslovaného seznamu stručných záznamů dokumentů. Tyto záznamy obsahují pouze informaci o názvu dokumentu, jméno autora, nakladatelské údaje a informaci o tom, zda je u dokumentu dostupný plný text nebo ne. Název dokumentu je odkazem na plný záznam. Ve výsledcích vyhledávání jsou k dispozici stejné filtry jako v klasickém rozhraní.

Bibliografický záznam dokumentu není příliš obsáhlý, sestává pouze z názvu dokumentu, jména autora a nakladatelských údajů. V záznamu je možnost zobrazit více informací (tlačítko *More Details*), kterými jsou údaj o jazyku, fyzický popis a odkaz do katalogu Worldcat pro zjištění lokalizace tištěného dokumentu. Záznam je možné odeslat emailem a zobrazit v klasickém rozhraní. Pokud je plný text dokumentu volně dostupný, lze

se k němu dostat pomocí tlačítka s názvem *Go To Full View*. Pokud není, lze v něm alespoň vyhledávat po kliknutí na tlačítko *Go To Limited View*.

Rozhraní pro prohlížení plných textů je opět řešeno jednodušeji než ve verzi klasické. Dokument se automaticky zobrazí ve formě, kdy prohlížení probíhá pomocí scrollování a nelze ji změnit. Je možné dokument prohlížet jako prostý text, zobrazit podrobnější informace o něm, dále je možné dokumenty stahovat a vyhledávat v nich. Rozhraní také umožňuje náhled dokumentu zvětšit či zmenšit.

Mobilní rozhraní HTDL ještě zřejmě nefunguje zcela tak, jak by mělo. Některé funkce nejsou dostupné, přestože tlačítka jsou v rozhraní zobrazena. Záleží na konkrétním zařízení, ale hlavně na internetovém prohlížeči. V následující tabulce jsou popsány nejčastější chyby a v kterých prohlížečích a na kterých zařízeních se objevují. Byly zahrnuty tyto chyby⁶⁸:

- **Filtry** – Jde o funkci pro rozbalení filtrů, která je umístěna ve výsledcích vyhledávání. Projevuje se tím, že tlačítko *Filter Results* vůbec nereaguje.
- **Následující** – Funkce přechodu na další stranu výsledků vyhledávání. Po použití tlačítka se objeví upozornění, že uživatel nemá oprávnění tuto stránku navštívit. V některých prohlížečích je tlačítko nazváno *Next* (např. v prohlížeči Chrome) a v jiných *Load More* (Opera mini). Jde o poměrně zásadní nedostatek, protože toto tlačítko nefungovalo v žádném testovaném prohlížeči ani na jednom zařízení. Uživatel se tak nemůže dostat k většímu množství výsledků, pouze k těm, které jsou na první stránce.
- **Výsledky v klasickém rozhraní** – Tlačítko, které by mělo uživatele odkázat na výsledky vyhledávání v klasickém rozhraní. Chyba spočívá v tom, že po kliknutí na tlačítko se zobrazí chybová hláška, která se liší podle použitého prohlížeče a v podstatě oznamuje, že k danému serveru není možné se připojit.
- **Více informací** – Funkce, umístěná v záznamu každého dokumentu, která zobrazuje další údaje o něm. Tlačítko opět ve většině případů nereaguje.
- **Odeslat emailem** – Jde rovněž o funkci záznamu a chyba se projevuje stejným způsobem jako ta předchozí.
- **Záznam v klasickém rozhraní** – Funkce, která má uživatele odkázat na záznam dokumentu v klasickém rozhraní HTDL. Chyba se projevuje tím, že odkaz vede pouze na hlavní stránku HTDL v klasickém rozhraní, ne na konkrétní záznam.

⁶⁸ Analýza byla provedena v období 10. - 13. 6. 2017

- **Full-text** – Takto byla nazvána chyba, která se projevuje v rozhraní pro prohlížení plných textů, a to tím, že nástrojová lišta je umístěna až na konci dokumentu. Uživatel se k ní tedy nedostane například z poloviny dokumentu, ale musí sjet až na jeho konec.

Zařízení	Prohlížeč	Filtry	Následující	Výsledky v klasickém rozhraní	Více informací	Odeslat emailem	Záznam v klasickém rozhraní	Full-text
Samsung SM-T230 (tablet)	Chrome	X	X	X	X	X	X	OK
	Samsung internet browser	X	X	X	X	X	X	OK
Lenovo A536	Vestavěný prohlížeč	OK	X	X	OK	OK	OK	OK
	Chrome	X	X	X	X	X	X	OK
Samsung S5	Chrome	X	X	X	X	X	X	OK
	Samsung internet browser	X	X	X	X	X	X	OK
	Mozilla	X	X	X	X	X	X	OK
	Opera mini	OK	X	X	OK	OK	OK	X
Huawei P8	Chrome	X	X	X	X	X	X	OK
	Vestavěný prohlížeč	X	X	X	X	X	X	OK

Tabulka 3: Chyby v mobilním rozhraní HTDL

Z uvedené tabulky vyplývá, že z testovaných prohlížečů je pro používání mobilní verze HTDL neoptimálnější vestavěný prohlížeč mobilního telefonu Lenovo A536. Jako druhý nejlepší vyšel prohlížeč Opera mini, který byl testován na mobilním telefonu Samsung S5. Ostatní prohlížeče dopadly stejně, objevily se v nich v podstatě všechny uvedené chyby kromě jedné. O všech těchto chybách byl tým pro vývoj HTDL informován pomocí formuláře pro zpětnou vazbu, na nějž je odkazováno z každé jednotlivé stránky webu HathiTrust. Oprava všech nahlášených chyb již byla v době odevzdání práce provedena.

2.2 Metadata v HTDL

V tak obrovském fondu, který představuje HTDL, jsou metadatové záznamy a jejich kvalita velice důležité. Jelikož tato digitální knihovna funguje již poměrně dlouhou dobu a metadatové záznamy do ní přicházejí z různých zdrojů, je jejich kvalita velmi různorodá. Je to způsobeno také tím, že v počátcích HTDL nebyl k dispozici tak propracovaný systém pro správu metadat, jaký funguje dnes. V následující části práce jsou popsána metadata v HTDL, jejich příjem, standard a možnosti získání. Na závěr je stručně shrnuta spolupráce HTDL a OCLC, která spočívá zejména ve výměně metadatových záznamů.

Metadatové záznamy dokumentů nejsou pro HTDL centrálně vytvářeny, ale jsou přijímány od přispěvatelů z řad členů partnerství HathiTrust. Instituce zasílají buď své vlastní záznamy, nebo záznamy svých dokumentů z databáze Worldcat. V HTDL jsou používány dva formáty metadatových záznamů, MARC a MARCXML. V katalogu HTDL jsou uživatelům k dispozici nekompletní záznamy. Kompletní záznamy v obou formátech lze zobrazit zapsáním „marc“ nebo „xml“ do adresního řádku na konec adresy záznamu v HTDL, viz [Obrázek 10](#) (ZAYTSEV, 2017).

Před samotným odesláním metadat i samotného obsahu je třeba vyplnit formulář s údaji o přispěvateli a dokumentech (metadatech), které budou odesílány⁶⁹.

The screenshot shows a web browser window with the URL <https://catalog.hathitrust.org/Record/011635618.marc>. The page displays a MARC record for the document 'Il marchese Salvatore Pes di Villamarina Memorie e documenti inediti'. The record is presented as a table with two columns: the MARC field number and its corresponding value.

LDR	00606nam a2200193L 4500
001	011635618
003	MiAaHDL
005	20120529000000.0
006	m d
007	cr bn ---auaua
008	890419s1863 ctu 00000bita d
035	\$a sdr-hvd001814096
035	\$a (OCoLC)19698598
035	\$a (OCoLC)4636869
040	\$a YUS \$c YUS \$d CRL \$d MH
100	1 \$a Bosio, Ferdinando, \$d 1829?-1881.
245	0 3 \$a Il marchese Salvatore Pes di Villamarina \$b Memorie e documenti inediti.
260	0 \$a Torino, \$b F. Franchini, \$c 1863.
300	\$a 257 p. \$c 19 cm.
538	\$a Mode of access: Internet.
600	1 4 \$a Pes, Salvatore, \$c marchese di Villamarina.
600	2 0 \$a Pes di Villamarina, Salvatore, \$c marchese, \$d 1808-1877.
CID	\$a 011635618
DAT	0 \$a 20120523123010.0 \$b 20120529000000.0
DAT	1 \$a 20120814092455.0 \$b 2016-11-04T18:08:44Z
CAT	\$a SDR-HVD \$d UNKNOWN \$i loader.pl-001-001
FMT	\$a BK
HOL	\$0 sdr-hvd001814096 \$a hvd \$b SDR \$c HVD \$p hvd.hnrmzd \$s HVD \$1 001814096
974	\$b HVD \$c HVD \$d 20161104 \$s google \$u hvd.hnrmzd \$y 1863 \$r pd \$q bib

Obrázek 10: Záznam dokumentu v HTDL ve formátu MARC

2.2.1 Příjem metadatových záznamů do HTDL

Metadata jsou do HTDL přijímána odděleně od samotného digitálního obsahu (o přijímání digitálních objektů [viz níže](#)). V současné době přispívá do HTDL svými záznamy a dokumenty 47⁷⁰ členů partnerství HathiTrust (STINE, 2016, s. 4-6). Na webových stránkách

⁶⁹ Formulář je dostupný na úložišti Google:

<https://docs.google.com/document/d/1Ko4kqzmzTPv12093uyxtUd11qoX61wWxV9kO64JJmw0k/edit>

⁷⁰ Informace k 10. 6. 2017

HathiTrust je pro ně připraven podrobný návod, jakým způsobem jsou metadata přijímána⁷¹. Příjem metadatových záznamů probíhá prostřednictvím protokolu FTPS⁷² skrze systém pro správu metadat Zephir. Metadata jsou tedy odesílána na Kalifornskou univerzitu, která tento systém spravuje a vyvíjí. Po vyplnění formuláře (viz výše) pošle přispěvatel pomocí FTPS vzorek svých metadat a počká na jejich předběžné schválení. Následně mohou být stejnou cestou odeslány všechny zamýšlené záznamy. Každá dávka souborů by měla obsahovat maximálně 100 000 záznamů ve formátu MARCXML. Zároveň by v každé dávce měly být záznamy dokumentů, které byly digitalizovány jedním subjektem. Pokud by tedy měla některá knihovna dokumenty, digitalizované společností Google a zároveň dokumenty, digitalizované například prostřednictvím nějakého lokálního digitalizačního projektu, musí záznamy těchto dokumentů rozdělit do dvou skupin podle zdroje digitalizace (UNIVERSITY OF MICHIGAN, 2017a). Názvy odesílaných souborů by měly mít standardizovanou strojově čitelnou podobu, podle následujícího vzoru:

<metadata source code>_<configuration code>_<date>_<digitizing agent>_<other distinguishing data>.xml

- Metadata source code – jde o identifikátor přispívající instituce. Lze podle něj zjistit, odkud konkrétní záznam pochází. Je to dvou až šesti písmenný kód, jejichž seznam je uveden ve výše zmíněném dokumentu s názvem *Submitting Metadata to Zephir*. V tomto dokumentu jsou uvedeny pouze kódy v současnosti přispívajících institucí, kompletní seznam marcovských kódů pro organizace je k dispozici zde: <http://www.loc.gov/marc/organizations/org-search.php>.
- Configuration code – jako v předchozím případě jde o kódové označení instituce, který označuje přispívající organizaci, ovšem s tím rozdílem, že je podrobnější. Například v rámci celé univerzity, identifikuje jednu konkrétní oborovou knihovnu. Pokud je tento kód totožný s předchozím, není nutné ho do názvu souboru uvádět. Seznam těchto kódů lze získat na stejných místech jako Metadata source code.
- Date – jde o datum vytvoření souboru ve formátu YYYYMMDD.
- Digitizing agent – označuje subjekt, který provedl digitalizaci dokumentu. Opět se jedná o písmenný kód, který v tomto případě poskytuje HathiTrust. Jejich seznam je k dispozici ve výše uvedeném dokumentu. Pokud by v něm požadovaný subjekt nebyl uveden, záleží jeho podoba na dohodě mezi přispěvatelem a HathiTrust. V současné době je v tomto seznamu celkem 38 zdrojů digitalizace.
- Other distinguishing data – jakékoliv dodatečné rozlišující informace, které chce přispěvatel uvést, například k odlišení souborů odesílaných v jeden den (STINE, 2016).

⁷¹ Dokument s názvem *Submitting Metadata to Zephir* je dostupný zde:

<https://docs.google.com/document/d/1ILpVfk5y3auLpBicflpGIJbf9bnNZ9qeJwkW48CZ1Xo/edit>

⁷² Adresa FTPS: ftps.cdlib.org (128.48.120.48)

Po odeslání záznamů je ještě nutné zaslat na emailovou adresu cdl-zphr-l@ucop.edu následující údaje v přesně strukturované formě, aby byly strojově čitelné:

- Název souboru=<název souboru>
- Velikost souboru=<velikost souboru v bytech>
- Počet záznamů=<počet záznamů>
- Kontaktní email=<emailová adresa, na kterou bude zaslán souhrn o přijetí souborů>

Po přijetí záznamů je každému záznamu pomocí systému pro správu metadat Zephir přiděleno skóre, které je zjištěno porovnáním se standardem pro metadatové záznamy, který je HTDL používán (více [viz níže](#)). Výše tohoto skóre následně určuje, který záznam bude do HTDL zařazen, pokud dorazí stejný záznam ze dvou různých zdrojů. Následně je na emailovou adresu příspěvatele odeslán podrobný popis celého procesu. Pokud jsou v záznamu nalezeny chyby, jsou v tomto emailu přesně popsány a příspěvateľ je vyzván, aby je opravil a záznamy opětovně odeslal přes FTPS (UNIVERSITY OF MICHIGAN, 2017a).

Po přijetí, kontrole a případných opravách, jsou záznamy se všemi verzemi uloženy do úložiště. Vybrané údaje ze záznamů jsou uloženy a indexovány v databázi. Následně Zephir exportuje nové záznamy do katalogu, databáze autorsko-právních metadat a dalších přístupových bodů. Zároveň přiděluje a denně odesílá seznam identifikátorů nových dokumentů do Michiganské univerzity, kde jsou uloženy do repozitáře, odkud se dají vyhledat pomocí katalogu. Zephir tedy přidává do katalogu metadatové záznamy a Michiganská univerzita k nim připojuje digitální objekty (UNIVERSITY OF MICHIGAN, 2016u).

2.2.2 Standard pro metadata v HTDL

Kromě níže popsaného standardu existují pro přijímané metadatové záznamy ještě obecnější principy a pravidla. Záznamy by měly:

- Být tak kompletní, jak je to jen možné
- Měl by existovat záznam pro každou jednotku
- Být kompatibilní s MARC21
- Být ve formátu MARCXML
- Být kódovány v utf8
- Měly by obsahovat identifikátor OCLC
- Měly by zahrnovat údaje, které jsou nutné pro automatické zjištění autorských práv (viz [kapitola o dostupnosti dokumentů](#) v HTDL) (UNIVERSITY OF MICHIGAN, 2017b)

Při kontrole, která následuje po přijetí záznamů, jsou záznamy porovnány se standardem a ohodnoceny přidělením skóre⁷³. Pro monografie se hodnotí přítomnost či nepřítomnost údajů celkem v jedenácti polích, pro periodika v devíti. Jde o pole, která jsou velmi důležitá pro určení přístupnosti dokumentu nebo pro vyhledávání, v případě identifikátoru OCLC pro synchronizaci s databází Worldcat. Jelikož jsou na základě absence některých údajů přidělovány také minusové body, je možné, že konečná hodnota přiděleného skóre, bude záporná. Podle výše záporných bodů by se dala pole v záznamech monografií seřadit podle důležitosti následovně:

Název pole	Hodnota v případě nepřítomnosti údaje
Údaje o názvu - 245 \$a	-50
Číslo OCLC - 035 \$a	-50
Místo vydání - 260 \$a nebo 264 \$a	-25
Datum vydání - 008 (pozice 7-10)	-20
Číslo v LCC - 010 \$a	-15
ISBN - 020 \$a	-15
ISSN - 490 \$x	0
Země vydání - 008 (pozice 15-17)	0
Počet stran - 300 \$a	0
Hlavní záhlaví - 100, 110, 111, 130	0
Údaje o edici - 490 \$a	0

Tabulka 4: Minusové body pro záznamy monografií

V případě periodických publikací tabulka vypadá takto:

Název pole	Hodnota v případě nepřítomnosti údaje
Místo vydání - 260 \$a nebo 264 \$a	-50
Číslo OCLC - 035 \$a	-50
Údaje o názvu - 245 \$a	-25
Země vydání - 008 (pozice 15-17)	-20
Číslo v LCC - 010 \$a	-15
ISSN - 020 \$a	-15

⁷³ Tabulka s rozpisem přidělování bodů je dostupná zde: <https://docs.google.com/document/d/1A6lq4bQBvHmqdz3Ab-CiqPpNqcXILe1bW5JCeNuw4x0/edit>

Datum vydání - 008 (pozice 7-10)	-15
Počet stran - 300 \$a	-15
Hlavní záhlaví - 100, 110, 111, 130	0

Tabulka 5: Minusové body pro záznamy periodických publikací

Z obou tabulek lze vyvodit, že nejdůležitějším údajem záznamu je identifikátor OCLC. Dále údaj o názvu a místo vydání. Následují země a datum vydání a identifikátory. Jako nejméně důležité se jeví hlavní záhlaví (UNIVERSITY OF MICHIGAN, 2013b). V případě, že v záznamu chybí některé z následujících polí, odešle systém pro správu metadat Zephir v emailu se souhrnem chybu a příspěvatel je vyzván k opravě záznamů.

- LDR
- 008
- Číslo OCLC
- Číslo záznamu v lokálním katalogu
- 245\$a
- Identifikátor jednotky (čárový kód nebo jiný identifikátor) (UNIVERSITY OF MICHIGAN, 2017b)

Standard pro metadatové záznamy přijímané HTDL je volně dostupný na webu HathiTrust. Jde o tabulku, která obsahuje název pole, jeho kódové označení ve formátu MARC, dále informace o tom, zda jde o vyžadovaný či silně preferovaný údaj a v některých případech také popis pole. Vyžadováno je celkem pět polí: LDR, 001, 008, 245 a 955. Všechna ostatní pole jsou silně doporučena. V přehledu jsou tučně zvýrazněna ta pole, která představují minimální bibliografický záznam (UNIVERSITY OF MICHIGAN, 2017a).

2.2.3 Možnosti získání záznamů z HTDL

Existuje několik možností, jak získat z HTDL metadata. První z nich jsou tzv. datasety, prostřednictvím nichž jsou k dispozici plné texty volně dostupných dokumentů vytvořené pomocí OCR. Platí zde odlišné podmínky pro dokumenty digitalizované společností Google a jinými subjekty. V obou případech platí, že dokumenty volně dostupné na území USA jsou k dispozici opět jen uživatelům nacházejícím se v USA. Ostatní uživatelé mohou získat pouze dokumenty, které jsou volně dostupné po celém světě. Pokud jde o dokumenty digitalizované jinak než společností Google, nevztahují se na ně žádná omezení pro jejich použití. Uživatel musí pouze podepsat prohlášení o tom, že pokud bude zjištěno, že se na některý dokument, jehož plný text uživatel prostřednictvím HTDL získá, stále vztahují autorská práva, HTDL o tom uživatele informuje a ten následně všechny kopie těchto souborů

odstraní. S dokumenty, které jsou digitalizovány společností Google je to o poznání obtížnější. Vztahují se na ně různá omezení, například nemohou být využívány komerčně, ale pouze pro účely vědeckého výzkumu. Instituce, které je žadatel o tyto dokumenty členem musí se společností Google podepsat zvláštní dohodu o poskytování dokumentů pro nekomerční účely⁷⁴ (UNIVERSITY OF MICHIGAN, 2017f).

Další možností, jak získat data z HTDL je použití jednoho ze dvou API rozhraní, které byly vyvinuty přímo pro HTDL. První z nich je určeno pro získání bibliografických metadat k menšímu počtu dokumentů ve formátu JSON. Pro získání záznamů je potřeba znát jeho identifikátor a vědět o jaký typ identifikátoru jde (např. číslo OCLC). Druhé umožňuje získat přímo skeny dokumentů ve formě obrazových souborů, obsah dokumentů ve formě OCR a další související metadata. Tato druhá možnost je však dostupná pouze pro přihlášené uživatele, kteří patří pod nějakou členskou instituci partnerství HathiTrust (UNIVERSITY OF MICHIGAN, 2017h).

Metadatové záznamy lze z HTDL získat také prostřednictvím protokolu OAI-PMH. Tuto funkci poskytuje a spravuje Michiganská univerzita pro více svých digitálních repozitářů najednou. Z tohoto důvodu je v některých příkazech nutné přidat informaci pro odlišení dílčí sbírky⁷⁵. Touto cestou lze stáhnout záznamy ve formátech MARC21 a nekvalifikovaném DC. Protokol OAI-PMH je jednou z cest, jak dostat obsah HTDL do lokálních katalogů členských institucí, které o tuto možnost mají zájem (UNIVERSITY OF MICHIGAN, 2017e).

2.2.4 Spolupráce s OCLC

Od roku 2010 spolupracuje partnerství HathiTrust s OCLC zejména tím způsobem, že mu dodává záznamy dokumentů obsažených v digitální knihovně HathiTrust, které jsou následně umístěny v katalogu Worldcat. V případě, že jde o záznam s dostupným plným textem, je součástí záznamu také odkaz na něj do digitální knihovny HathiTrust. Zároveň na základě této spolupráce vzniklo nové vyhledávací rozhraní pro digitální knihovnu HathiTrust. Jde o uživatelské rozhraní vytvořené na základě rozhraní katalogu Worldcat s názvem Worlcat Local (OCLC adds Google Books, HathiTrust records to Worldcat, 2010). Katalog Worldcat je s HathiTrust denně (každou noc) synchronizován a záznamy jsou aktualizovány. Kvůli tomuto propojení také HTDL požaduje od přispěvatelů metadatových záznamů údaj

⁷⁴ Smlouva je dostupná z webu HTDL:

https://www.hathitrust.org/documents/Google_PD_Distribution_Agreement_template.docx

⁷⁵ Do příkazu se na konec vloží následující údaj k identifikaci sbírky: set=hathitrust

o identifikátoru OCLC⁷⁶ a souhlas se sdílením těchto údajů. Všechny záznamy, které jsou v HTDL by tedy mělo být možné vyhledat také v katalogu Worldcat (ZAYTSEV, 2016⁷⁷).

2.3 Ukládání digitálních objektů do HTDL

Při přijímání digitálního obsahu do HTDL se postupuje podle toho, zda jsou již dokumenty digitalizovány či nikoliv. Pokud přispívající instituce digitalizaci teprve plánuje, nabízí Michiganská univerzita své vlastní digitalizační specifikace⁷⁸. Pokud se jimi instituce řídí, není pak s přijetím jejich obsahu do HTDL žádný problém (UNIVERSITY OF MICHIGAN, 2017g). V případě, že přispěvatel má již své dokumenty digitalizovány, nabízí HTDL nástroje pro validaci naskenovaných obrazů, metadat a také vstupních informačních balíčků (SIP). Pro nejnovější informace o těchto nástrojích se může přispěvatel přihlásit do speciální emailové konference. HTDL samozřejmě nabízí informační podporu pro práci s validačními nástroji (UNIVERSITY OF MICHIGAN, 2017k).

Samotný proces přenosu obsahu se liší podle zdroje jeho digitalizace. V případě, že je zdrojem společnost Google, informuje ji přispěvatel o svém záměru poskytnout digitalizovaný obsah HTDL. Následně si vybere formát/y, ve kterých se má obsah poskytnout. Většina institucí volí formát TIFF pro textové dokumenty a JPEG2000 pro obrazy. Jde o neefektivnější model, jelikož celková velikost souboru je o 25 % menší, než kdyby byly všechny objekty ve formátu JPEG2000. Po volbě formátu instituce požádá společnost Google, aby umožnila Michiganské univerzitě obsah stáhnout. Velmi podobný postup se používá v případě, že je zdrojem digitalizace nezisková organizace Internet Archive. Pokud přispěvatel zamýšlí zaslat obsah, který není digitalizován ani jednou z výše uvedených společností, je možné jej po domluvě s HTDL doručit nejrůznějšími způsoby. Například přes FTP server, nějaké HW médium (externí disk, DVD, BluRay disk atd.), dále prostřednictvím některé služby pro sdílení souborů (Box.com apod.) nebo jiným, předem domluveným způsobem (UNIVERSITY OF MICHIGAN, 2017j).

Jak již bylo uvedeno výše, obsah je uložen na Michiganské univerzitě, druhá kopie je umístěna na Indické univerzitě na zrcadleném serveru. Další dvě kopie jsou uloženy na zašifrovaných páskách, z nichž každá je uložena na jiném bezpečném místě v prostoru Michiganské univerzity (ZAYTSEV, 2017).

⁷⁶ Tento údaj se v záznamu MARC uvádí v poli 035

⁷⁷ Přepis emailu představuje přílohu č. 1 této práce

⁷⁸ Celý dokument je k dispozici zde:

<https://www.lib.umich.edu/sites/default/files/services/dlps/umdigitizationspecs20140825-by-cc-license.pdf>

V případě masových digitalizačních projektů (společnost Google a Internet Archive) může při procesu digitalizace samozřejmě docházet k chybám. Příklad chybně naskenované strany je vidět na [Obrázek 11](#)⁷⁹. Chyba byla nahlášena HTDL pomocí formuláře pro zpětnou vazbu. Společnost Google vyvinula proces kontroly a oprav těchto chyb. V případě, že je chyba již opravena, Michiganská univerzita si jednoduše stáhne opravenou verzi. V opačném případě se obě instituce domluví na nejrychlejším možném řešení (ZAYTSEV, Angelina, 2017).



Obrázek 11: Příklad chybně naskenované strany

2.4 Autorská práva

Jelikož autorské právo přímo souvisí možnostmi zpřístupňování plných textů, musí se jimi HTDL velmi intenzivně zabývat. V autorském právu Spojených států existuje výjimka pro knihovny a archivy (§ 108), které tento zákon neporušují, pokud si vytvoří kopii díla ze svého fondu. Je však omezena následujícími podmínkami:

- Kopie nesmí být vytvořena za jakýmkoliv komerčním účelem, přímým ani nepřímým.
- Knihovna (archiv), která kopii zamýšlí vytvořit, musí být přístupná veřejnosti.
- U kopie nebo přímo na kopii díla musí být uvedeno, že jde o dílo chráněné autorským zákonem.

⁷⁹ Odkaz na stranu: <https://babel.hathitrust.org/cgi/pt?id=mdp.39015058515449;view=1up;seq=3>

Výjimka se vztahuje na tři kopie a stejně jako v českém právu, také zde platí zákaz zpřístupňovat tyto dokumenty mimo prostory knihovny. A to ani v případě, že jde o poškozené či zničené dokumenty, jejichž náhradu nelze na trhu získat za přiměřenou cenu (Spojené státy americké, 2011, s. 19-22). V následující části práce jsou popsány možnosti zpřístupňování obsahu HTDL a autorsko-právní metadata, která je určují. Dále jsou stručně shrnuty základní informace o soudním sporu mezi partnerstvím HathiTrust a Authors Guild.

2.4.1 Dostupnost plných textů

Dostupnost plných textů v HTDL se řídí autorsko-právním statutem jednotky a typem uživatele. Autorsko-právní status jednotky je určen pomocí údajů v bibliografickém záznamu jednotky. Tento status determinují zejména datum a místo vydání dokumentu.

Dostupnost	Podmínky
Volně dostupné odkudkoliv	Dokumenty, vydané v USA před rokem 1923, v Kanadě a Austrálii před rokem 1897 a kdekoliv jinde před rokem 1877. Vládní dokumenty USA.
Volně dostupné v USA	Dokumenty, vydané mimo USA mezi lety 1877-1922.
Nedostupné (in-copyright)	Všechny ostatní dokumenty.

Tabulka 6: Podmínky dostupnosti plných textů v HTDL

V tabulce jsou uvedeny tři základní typy dostupnosti dokumentů v HTDL (UNIVERSITY OF MICHIGAN, 2017d). Existují možnosti širšího zpřístupnění obsahu, které jsou ovšem podmíněny typem uživatele, kterých je v současnosti celkem pět. Na webu HathiTrust je uveden jejich seznam a každý z nich má své kódové označení:

- Běžný uživatel (ORD) – nepřihlášený uživatel, nebo uživatel se spřáteleným účtem
- Znevýhodněný uživatel (SSD) – uživatel, který je identifikován jako osoba se znevýhodněním, které jí znemožňuje nebo výrazně ztěžuje percepci obsahu tištěného dokumentu
- Uživatel, který se nachází v budově knihovny Michiganské univerzity (LIB) – tento uživatel nemusí být přihlášený
- Uživatel, který je identifikován jako student nebo zaměstnanec Michiganské univerzity (UM) – tito uživatelé mohou využívat některých výjimek z autorského zákona v § 108
- Uživatel, který je identifikován jako člen některé partnerské instituce HathiTrust (HT)

Z výše uvedeného přehledu vyplývá, že nejomezenější přístup k obsahu HTDL má nepřihlášený uživatel či uživatel se spřáteleným účtem, který se nachází mimo území USA. Znevýhodnění uživatelé mohou získat přístup dokonce k plným textům dokumentů, které jsou stále chráněny autorským právem. Vztahuje se to však pouze na dokumenty, jejichž tištěnou verzi vlastní instituce, jíž jsou členy. Uživatelé označení jako LIB a UM mají oproti ostatním tu výhodu, že mohou využívat plné texty dokumentů, které jsou poškozeny a nelze je již na trhu sehnat za přiměřenou cenu, takže spadají pod § 108 autorského zákona USA. Z hlediska přístupu k plným textům nemají uživatelé z členských institucí HathiTrust větší výhody než nepřihlášení uživatelé přistupující do HTDL z US IP adres (UNIVERSITY OF MICHIGAN, 2017). Omezení IP adresou lze obejít použitím amerického proxy serveru, který lze buď nainstalovat jako software nebo využít možnosti webové verze proxy serveru. Tímto způsobem je možné zobrazit plné texty dokumentů, které jsou dostupné pouze z amerických IP adres. Nevýhodou tohoto nástroje je, že poměrně dost zpomaluje rychlost internetu, což je u načítání velkých souborů omezující. Tato metoda byla vyzkoušena na webovém proxy serveru s adresou: <http://unblockyouku.org>.

Výhoda uživatelů z členských institucí HathiTrust spočívá v možnostech stahování plných textů ve formátu PDF. Možnost stáhnout plný text celého dokumentu ve formátu PDF souvisí se zdrojem digitalizace, přesněji řečeno, souvisí s dokumenty digitalizovanými společností Google. Což představuje téměř 95 % obsahu HTDL. Přibližně 3,7 % dokumentů bylo digitalizováno organizací Internet Archive a kolem 1,5 % obsahu pochází z menších lokálních digitalizačních projektů. Druhé dvě možnosti nepředstavují ohledně stahování ve formátu PDF žádný problém. Protože všechny dokumenty, digitalizované těmito subjekty, jsou volně dostupné, může si každý uživatel (přihlášený i nepřihlášený) stáhnout jejich plný text ve formátu PDF⁸⁰. Naproti tomu společnost Google požaduje, aby jí digitalizované dokumenty byly ve formátu PDF dostupné pouze přihlášeným uživatelům, kteří spadají pod některou z členských institucí HathiTrust. Ostatní uživatelé jsou oprávněni stáhnout pouze jednu stranu dokumentu (ZAYTSEV, 2017).

Autorsko-právní metadata jsou shromažďována ve speciální databázi autorsko-právních metadat. Celkem jich existuje 27 typů, kdy každý z nich má identifikační číslo (1-27), název a popis. V následující tabulce jsou uvedeny některé z nich (UNIVERSITY OF MICHIGAN, 2017i):

⁸⁰ Například tento dokument lze celý stáhnout ve formátu PDF: <https://catalog.hathitrust.org/Record/100967934>

Název	Popis
pd	volně dostupné odkudkoliv
ic	chráněno autorským zákonem
op	dokumenty nedostupně na trhu za přiměřenou cenu (out-of-print) (stejná práva jako dokumenty chráněné autorským zákonem)
orph	osiřelá díla (stejná práva jako dokumenty chráněné autorským zákonem)
und	neurčeno (stejná práva jako dokumenty chráněné autorským zákonem)
ic-world	chráněno autorským zákonem, autorem povoleno zpřístupnění odkudkoliv
nobody	nedostupné žádnému uživateli
pdus	volně dostupné z USA
cc	dokument pod licencí Creative Commons, celkem je možnost použít jeden ze 13 typů, každý má vlastní označení, které začíná: „cc“

Tabulka 7: Vybraná autorsko-právní metadata

Podle vyjádření současného výkonného ředitele HathiTrust je výhledově plánováno vyvinout a zařadit do vyhledávání filtr pro autorsko-právní metadata. Ten by mohl být užitečný zejména pro uživatele, kteří do HTDL přistupují ze zemí mimo USA. V současné době jsou ve volně dostupných dokumentech zařazeny také dokumenty, dostupné pouze z US IP adres, najít volně dostupný dokument je tedy pro uživatele mimo USA poměrně obtížné (FURLOUGH, 2016).

2.4.2 Soudní spor Authors Guild v. HathiTrust

Jak již bylo uvedeno výše, partnerství HathiTrust prošlo, podobně jako společnost Google⁸¹, poměrně zdlouhavým soudním sporem s držiteli autorských práv. Původní soudní spor byl zahájen v září roku 2011. Šlo v podstatě o to, zda HTDL porušuje autorská práva. První spor se projednával u US District Court, Southern District of New York a soudcem byl Harold Baer. Žalující strana se skládala z dvanácti individuálních osob a sedmi organizací:

- The Authors Guild, Inc. (USA)
- The Australian Society of Authors Limited (Austrálie)
- Authors' Licensing and Collecting Society (Velká Británie)
- Union des Écrivaines et des Écrivains Québécois (Kanada, Quebec)

⁸¹ Více o soudních sporech společnosti Google například v této diplomové práci (zejm. kapitola 4.3): FRČKA, Pavel. *Vyhledávání knih Google*. Praha, 2014. 151 s. Diplomová práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková, PhD. Dostupné z: <https://is.cuni.cz/webapps/zzp/detail/108359/>.

- Sveriges Författarförbund (Švédsko)
- Norsk Faglitterær Forfatter-og Oversetterforening (Norsko)
- The Writers' Union of Canada (Kanada)

Další organizace s názvem Authors League Fund (USA) se nechtěla s ostatními sloučit, ale chtěla uplatnit v soudním sporu svůj nárok na základě přímého vlastnictví autorských práv. Ve všech případech jde o organizace, které zastupují držitele autorských práv. Žalovaná strana sestávala z těchto členů:

- Mary Sue Coleman, prezidentka Michiganské univerzity
- Mark G. Yudof, prezident Kalifornské univerzity
- Kevin Reilly, prezident Systému Wisconsinské univerzity
- Michael McRobbie, prezident Indianské univerzity
- Cornellova univerzita (jméno zastupující osoby nebylo uvedeno)

Soudce přijal návrh, aby do procesu za tuto stranu vstoupila také Národní federace nevidomých, zastoupená třemi osobami: Georgina Kleege, Blair Seidlitz, a Courtney Wheeler (UNITED STATES DISTRICT COURT. SOUTHERN DISTRICT OF NEW YORK, 2012, s. 1-2).

Soud projednával, zda žalovaná strana porušuje autorská práva žalující strany a osob, které zastupuje. Předmětem byla autorská díla, která jsou stále chráněna autorským zákonem a byla digitalizována společností Google. Během celého procesu digitalizace, předání digitálního obsahu příslušné knihovně a poskytnutí tohoto obsahu do HTDL, vznikne celkem dvanáct kopií tohoto dokumentu. V HTDL jsou pak užity třemi způsoby: pro fulltextové vyhledávání, dlouhodobou archivaci a pro jejich zpřístupnění uživatelům s tzv. print disabilities. Soud rozhodl, že tyto činnosti spadají pod ustanovení tzv. „*fair use*“, u něhož je nutné zvážit celkem čtyři faktory:

- Účel a charakter použití díla, chráněného autorským právem, včetně toho, zda bylo použito za komerčním účelem či pro vzdělávací účely
- Povaha díla, chráněného autorským právem
- Velikost použité části díla vzhledem k jeho celku
- Vliv použití díla na jeho hodnotu a případně na trh

Dále žalující strana navrhovala, aby soud vydal prohlášení, že žalovaná strana poruší autorský zákon, pokud bude zpřístupňovat díla jím chráněná v rámci projektu zaměřeného na osiřelá díla s názvem OWP (podrobněji o projektu [viz výše](#)). S tímto prohlášením měl být vydán

soudní zákaz v pokračování tohoto projektu. Jelikož se projekt nikdy nedostal dále než do plánovací fáze, soud rozhodl, že takový zákaz nevydává. Bylo by k tomu potřeba mnohem více informací o tom, jak by celý proces fungoval a kdo přesně by jím byl zasažen. Soudce uvedl, že pokud bude projekt v budoucnu opět spuštěn, může žalující strana zkusit opakovat svůj požadavek (UNITED STATES DISTRICT COURT. SOUTHERN DISTRICT OF NEW YORK, 2012, s. 3-22).

Žalující strana se po vynesení rozsudku odvolala a další soudní proces probíhal od 30. října 2013 do 10. června 2014, kdy byl vynesena potvrzující rozsudek (UNITED STATES COURT OF APPEALS (2ND CIRCUIT), 2014, s. 34). Za oficiální datum ukončení soudního sporu je považován 6. leden 2015, kdy byla mezi oběma stranami uzavřena dohoda o ukončení sporu ve prospěch žalované strany (COX, 2015). Dohoda sestává ze tří bodů. V prvním žalovaná strana deklaruje, že při zhotovování kopií autorských děl postupuje v souladu s § 108 autorského zákona USA, tedy že originální dokument byl poškozen, zcizen, či zničen a digitalizací se nahrazuje. Zároveň bylo vynaloženo přiměřené úsilí, na jehož základě bylo zjištěno, že dokument není komerčně dostupný za odpovídající cenu. Druhým bodem žalovaná strana souhlasí s tím, že po dobu pěti let od uzavření dohody bude žalující straně posílat oznámení o svých činnostech, uvedených v prvním bodě. Třetí bod potvrzuje, že rozhodnutím soudu zanikají všechny požadavky na odškodnění a každá strana hradí své náklady na soudní řízení sama (UNITED STATES DISTRICT COURT. SOUTHERN DISTRICT OF NEW YORK, 2015).

Rozhodnutí tohoto soudního sporu poskytlo mnoha knihovnám jasné ujištění, že mohou digitalizovat obsah svých fondů, pokud budou poskytovat uživatelům možnost fulltextového vyhledávání a pokud budou poskytovat přístup k plným textům znevýhodněným uživatelům.

3. Analýza obsahu digitální knihovny HathiTrust

Hlavní část této práce představuje obsahová analýza HTDL se zřetelem na dokumenty v českém jazyce. V následující kapitole je popsána její metodika a cíle. Následuje shrnutí a interpretace dostupných statistik celkového obsahu HTDL z hlediska jazyka dokumentu a data vydání. Další částí je představení výsledků analýzy česky psaných dokumentů v HTDL. Sledovanými kritérii byly datum a místo vydání dokumentu, dále jeho typ a částečně také obsah. Vzhledem k výše popsaným možnostem vyhledávání a funkcím sbírek jsou v některých případech popisovány počty dokumentů a v některých počty jednotek. Tato informace je vždy uvedena na začátku příslušné části.

3.1 Cíle a metodika obsahové analýzy HTDL

Cílem analýzy je stručně popsat celý obsah HTDL prostřednictvím interpretace dostupných statistik. Hlavním cílem je však zjistit, kolik a jaké česky psané dokumenty HTDL obsahuje. Dále postihnout vlastnosti těchto dokumentů podle různých kritérií, aby tak vznikl alespoň přibližný obraz o bohemikálních dokumentech v této digitální knihovně. Dílčím cílem bylo zjistit, zda jsou mezi těmito dokumenty nějaké zdroje, které by mohly být užitečné pro obor Informační studia a knihovnictví.

3.1.1 Metody sběru dat

Data pro obsahovou analýzu HTDL byla získána několika způsoby. Pro popsání celkového obsahu HTDL byly použity statistiky na jejích webových stránkách. Jsou generovány automaticky prostřednictvím záznamů dokumentů z katalogu HTDL. Statistiky jsou vytvářeny pro celé dokumenty, nikoliv pro jednotky. Pro získání dalších dat bylo také použito filtrování výsledků katalogového vyhledávání HTDL.

Hlavním zdrojem dat pro analýzu česky psaných dokumentů v HTDL, byla sbírka s názvem *All in Czech*, která byla vytvořena 9. 6. 2017 pomocí k tomu určeného nástroje v rozhraní HTDL. Jde o sbírku jednotek, nikoliv dokumentů. Je to způsobeno možnostmi a funkcemi katalogu HTDL, kdy v katalogovém vyhledávání (záznamy celých dokumentů), je možné přidávat položky do sbírek pouze po jedné. To je pro vytváření takto velké sbírky velmi časově neefektivní. Ve fulltextovém vyhledávání, při němž jsou zobrazovány jednotky, lze přidat až 100 položek současně. Bylo tedy rozhodnuto, že budou analyzovány jednotky. Rozhraní HTDL nabízí poměrně málo možností práce s těmito sbírkami, například ve sbírkách nejsou k dispozici žádné filtry. Jednotky lze pouze řadit podle názvu, jména autora a data vydání, ve všech případech sestupně i vzestupně. Sbírkou dále umožňují fulltextové

vyhledávání. Tato funkce byla použita při zjišťování, zda jsou mezi těmito dokumenty nějaké zdroje použitelné pro obor Informační studia a knihovnictví. Pro analýzu nejdůležitější funkci sbírek představovala možnost stáhnout metadata všech jednotek, které jsou v ní shromážděny. Šlo o soubor ve formátu .txt, který byl následně otevřen v programu Excel, kde byl automaticky strukturován. Soubor těchto metadat je přílohou č. 1 této práce, na přiloženém CD. Jeho obsahem jsou následující informace o každé jednotce:

- Identifikátor htid
- Název
- Jméno autora
- Datum vydání
- Identifikátor OCLC
- Identifikátor LCC
- Identifikátor ISBN
- Trvalý odkaz na záznam v katalogu
- Trvalý odkaz na jednotku

Dalším zdrojem dat pro analýzu česky psaných dokumentů v HTDL bylo, podobně jako v analýze celkového obsahu, vyhledávací rozhraní HTDL. Pro zjištění dalších informací (typ dokumentu, jeho obsah, apod.), které nejsou v metadatovém souboru sbírky k dispozici, byly použity filtry ve fulltextovém vyhledávání.

3.1.2 Metody analýzy dat

Hlavním nástrojem pro analýzu sesbíraných dat byl software MS Excel. V něm bylo pro zpracování a zobrazení výsledků analýzy použito několika nástrojů.

Pro celkovou analýzu obsahu HTDL byla použita hlavně funkce vytváření grafů. Data ze statistik byla zkopírována a byly z nich vytvořeny tabulky a následně grafy. Pro zobrazení výsledků analýzy česky psaných dokumentů bylo rovněž použito grafů. Pro zpracování dat, získaných prostřednictvím vytvoření sbírky s názvem *All in Czech*, jsem používala zejména funkce Filtr, která umožňuje zobrazit pouze určité hodnoty, jejichž počet lze poté snadno zjistit. Pro označení duplicitních údajů bylo použito funkce Podmíněné formátování. Pomocí této funkce bylo možné lépe analyzovat skupinu jednotek, u nichž nebylo uvedeno datum jejich vydání (podrobněji popsáno [níže](#) v části o česky psaných dokumentech bez data vydání).

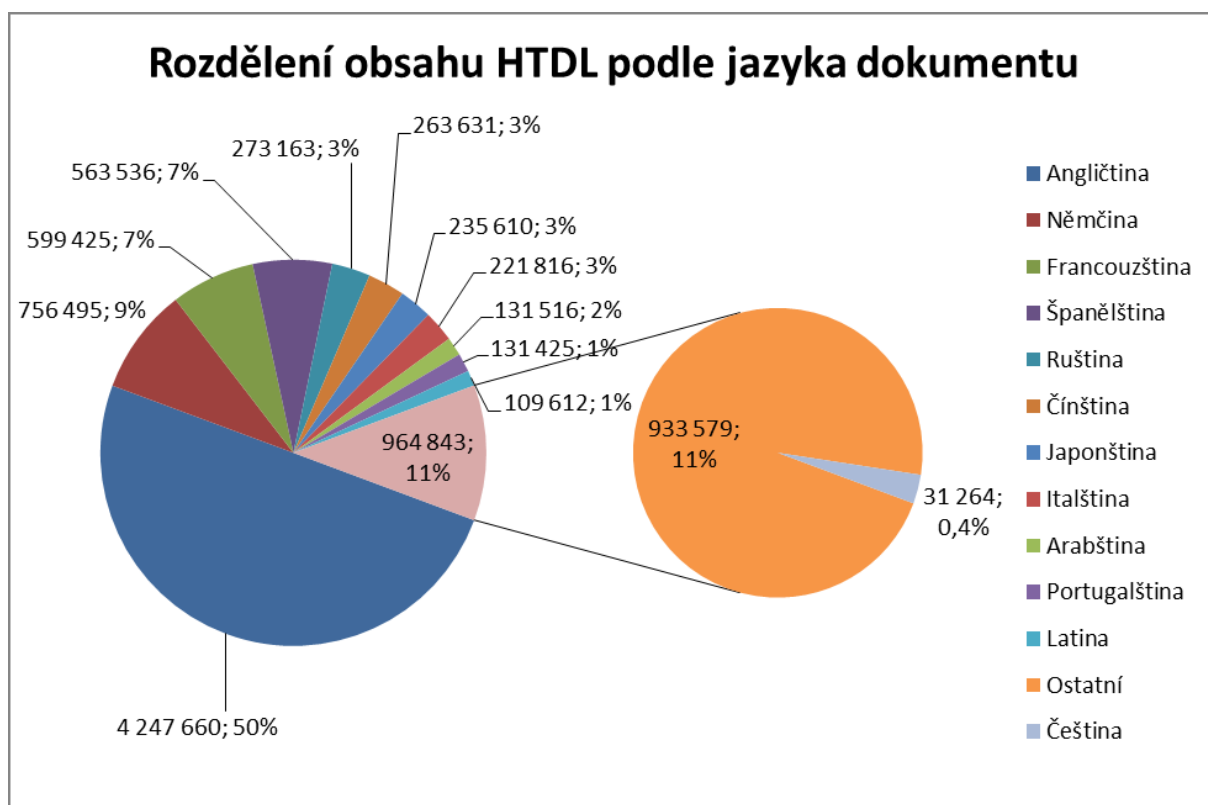
3.2 Dostupné statistiky obsahu HTDL

V následující části práce jsou stručně shrnuty celkové statistiky obsahu HTDL, které jsou umístěny na jejich webových stránkách. Tyto údaje jsou pravidelně aktualizovány, ale ne zcela synchronně s katalogem. To způsobuje, že ve statistikách a v katalogu se počty dokumentů mírně liší, v katalogu jich bývá o něco více. Statistiky jsou dostupné vždy pro všechny dokumenty nebo pouze pro dokumenty, které jsou volně dostupné. Kliknutím na položku ve statistice je uživatel odkázán do katalogu na seznam požadovaných dokumentů ve formě výsledků vyhledávání. Pokud tedy uživatel klikne ve statistice podle jazyka dokumentu na položku „German“, objeví se výsledky vyhledávání v katalogu pro všechny dokumenty vydané v německém jazyce, které lze dále filtrovat podle různých kritérií. Statistiky jsou vytvářeny pro dokumenty podle jazyka a data vydání. V následujícím textu je tedy stručně popsán obsah HTDL z hlediska těchto dvou údajů. Oba tyto údaje se pro statistiky stahují z MARC záznamu z pole 008 (v případě data vydání pozice 07-10 a v případě kódu pro jazyk pozice 35-37). Statistiky jsou vytvářeny pro úroveň dokumentu, nikoliv jednotky, jako je tomu v případě analýzy bohemik v další části kapitoly.

3.2.1 Jazyk dokumentu

V HTDL jsou v současnosti⁸² uloženy dokumenty v celkem 461 jazycích. V následujícím grafu je znázorněno zastoupení jazyků od 1 % v obsahu HTDL. Výsek jazyků, označených jako „Ostatní“ zahrnuje všechny jazyky, které nedosahují ani 1 %. V menším grafu je znázorněn počet dokumentů v češtině v množině „Ostatních“, málo početných jazyků.

⁸² Informace k 16. 6. 2017



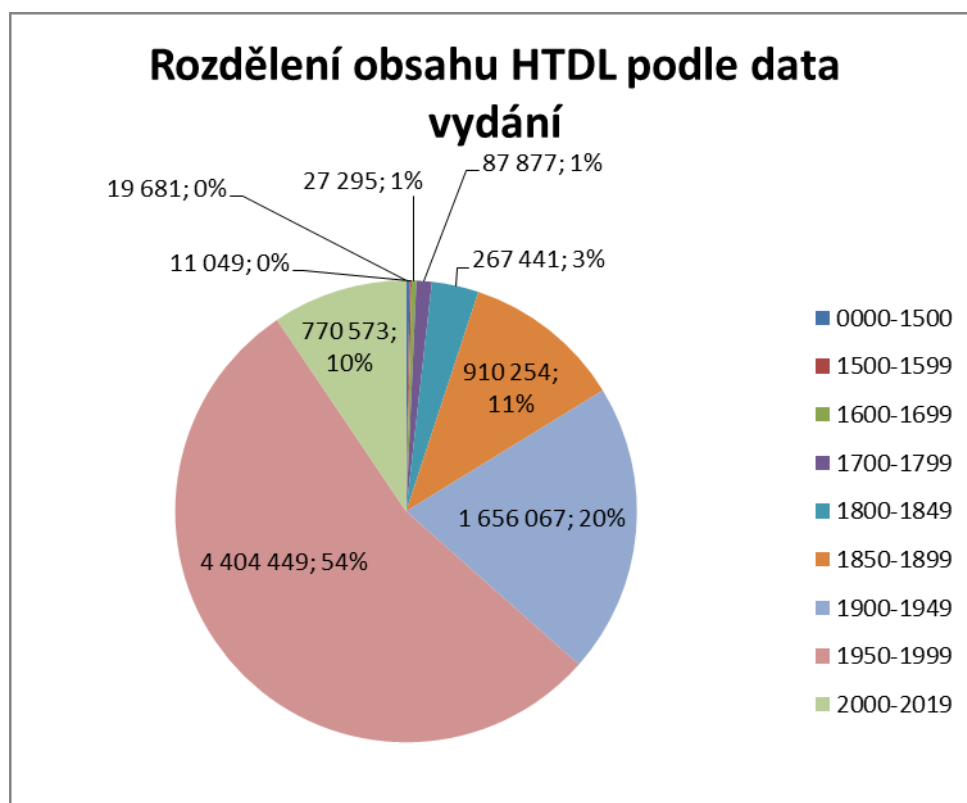
Graf 1: Rozdělení obsahu HTDL podle jazyka dokumentu

Z grafu vyplývá, že naprosto nepřekvapivě je v HTDL největší množství dokumentů psáno v angličtině. Je jich přesně polovina. Dalšími nejvíce zastoupenými jazyky jsou němčina, francouzština a španělština. Se třemi procenty následují ruština, čínština, japonština a italština. Dvěma procenty je zastoupena arabština a po jednom procentu portugalština a latina. Skupinu ostatních tvoří 450 dalších jazyků, které jsou celkově zastoupeny v 11 % dokumentů. Některé jazyky (např. francouzština, angličtina, němčina, řečtina) jsou ve statistikách rozděleny do skupin podle období, kdy byly používány. Angličtina je v nich například zastoupena hned třikrát, klasickou (dnešní) angličtinou (50 % obsahu HTDL), dále starou angličtinou (cca 5. - 12. století) a středověkou angličtinou (cca 12. – 16. století). Nejméně početnými skupinami je 21 jazyků, které jsou v HTDL zastoupeny pouze jedním dokumentem. Mezi těmito jazyky jsou například korsičtina, jazyk divehi, používaný na Maledivách a další velmi malé jazyky. Více než půl procenta (0,6 %) má skupina dokumentů, u nichž není jazyk v záznamu uvedený. Podle statistik obsahuje HTDL 0,2 % dokumentů, které nejsou jazykové povahy, takže v jejich záznamu jazyk také není uvedený. Poslední skupinou, u které není jazyk jasně deklarován, jsou vícejazyčné dokumenty, kterých však není příliš mnoho (0,1 %) (UNIVERSITY OF MICHIGAN, 2017).

Podle filtrů ve výsledcích vyhledávání jde v případě anglicky psaných dokumentů většinou o knihy a časopisy či jiné periodické publikace. Jsou zde však zastoupeny také slovníky, mapy, hudebniny a obrazové dokumenty a v menší míře audio a video nahrávky. Nejvíce anglicky psaných dokumentů (cca 2,5 mil.) vlastní Michiganská a Kalifornská univerzita, které do HTDL přispívají nejvíce. Zhruba stejný počet těchto dokumentů byl vydán na území USA, následuje Velká Británie a Anglie. U dalších více zastoupených jazyků (Němčina, Francouzština a Španělština) funguje podobná analogie, nejčastěji jde o knihy a seriály, které byly vydány v zemích, kde se danými jazyky hovoří. Největším množstvím dokumentů opět přispěly dvě výše uvedené univerzity. V případě ruštiny, čínštiny a japonštiny je situace ohledně země vydání a vlastníka dokumentu, obdobná. Změna nastává až u portugalštiny. Nejvíce dokumentů v tomto jazyce vlastní Texaská univerzita a opět jde ve většině případů o monografie a seriály. Dokumenty psané v latině přispěla v největší míře Univerzita Complutense v Madridu. Nejpočetnějším typem dokumentu jsou sice monografie, ale těsně za nimi nejsou jako obvykle seriálové publikace, ale hudebniny. Následuje početná skupina jazyků, které nedosáhly na 1 % obsahu HTDL. Je mezi nimi samozřejmě také čeština. Podobné procentuální zastoupení (0,4 %) jako čeština, mají ještě indonézština, dánština a korejština. Jen pro zajímavost, slovenština je v HTDL zastoupena čtyřikrát méně než čeština (0,1 %). Tato skutečnost však může být způsobena tím, že se při katalogizaci tyto jazyky často pletou, a navíc ani netřeba zmiňovat možnost spletení kódu jazyka, který je pro slovenštinu slo (např. se slovinštinou, která má kód slv).

3.2.2 Datum vydání

Dokumenty podle data vydání jsou ve statistikách na webu HTDL rozděleny do skupin po deseti letech. To platí o dokumentech vydaných od roku 1800. Dokumenty vydané před tímto datem jsou ve skupinách po sto letech (UNIVERSITY OF MICHIGAN, 2017I). V následujícím grafu jsou dokumenty, vydané od roku 1800, kvůli přehlednosti rozděleny po padesáti letech.



Graf 2: Rozdělení obsahu HTDL podle data vydání

Co se týče dat vydání, nejpočetnější skupinu tvoří dokumenty, vydané mezi lety 1950-1999, konkrétně v 80. letech 20. století. Následují dokumenty, publikované v první polovině 20. století. Třetí nejpočetnější skupinou jsou dokumenty, vydané v rozmezí let 1850-1899. Následuje skupina nejnovějších dokumentů, vydaných v letech 2000-2019. Dokumenty, vydané mezi lety 0000-1849, jsou v HTDL zastoupeny nejméně. To je nejspíše způsobeno tím, že v průběhu doby se počet produkováných dokumentů neustále zvyšuje.

Nejnovějšími dokumenty jsou podle filtrů ve vyhledávání čtyři knihy, které mají být vydány v roce 2019⁸³. S největší pravděpodobností jde o chyby v katalogizačních záznamech, nejnovější jsou tedy dokumenty, vydané v roce 2016. Nejstarší dokument má v záznamu uvedeno jako rok vydání „0018“⁸⁴. Až v pořadí patnáctý nejstarší dokument má v záznamu u data vydání uveden důvěryhodný údaj. Jde o rok 500, v podrobnějším záznamu je uvedeno šesté století. Jde o rukopis v kopštině⁸⁵, který je samozřejmě dostupný v plném textu a byl digitalizován Michiganskou univerzitou. Hlavně z tohoto důvodu jej lze bez problémů stáhnout v celém rozsahu v PDF.

⁸³ Například: <https://catalog.hathitrust.org/Record/000988286>

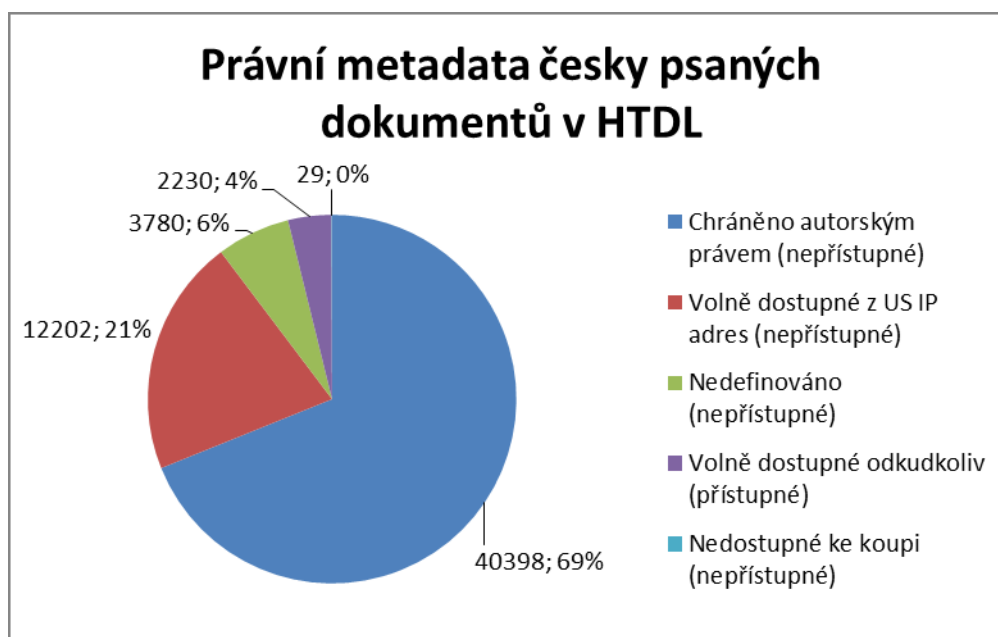
⁸⁴ <https://catalog.hathitrust.org/Record/102032364>

⁸⁵ Odkaz na záznam rukopisu: <https://catalog.hathitrust.org/Record/101703234>

3.3 Výsledky analýzy bohemikálních dokumentů v HTDL

Jak již bylo uvedeno výše, hlavním zdrojem dat pro analýzu česky psaných dokumentů v HTDL byla funkce vytváření sbírek. Dalším důležitým nástrojem analýzy byl soubor s metadaty o těchto jednotkách, který je možné ke každé sbírce stáhnout. Soubor byl nejprve otevřen v jednoduchém textovém dokumentu, jehož obsah byl následně zkopírován do tabulkového editoru MS Excel, kde byla data automaticky strukturována. Pomocí tohoto programu byla data také analyzována a to zejména prostřednictvím funkce *Filtry*.

Sbírka s názvem All in Czech byla vytvořena dne 9. 6. 2017 a obsahuje celkem 58 639 záznamů jednotek. Z tohoto množství jsou jako volně dostupná označena 4 % (2 230). Dostupnost těchto dokumentů je určována na základě autorsko-právních metadat, která determinují možnosti jejich zpřístupnění. Následující graf znázorňuje podíl různých autorsko-právních statutů, které jsou uvedeny v souboru metadat o této sbírce. Kompletní přehled autorsko-právních statutů a jejich kódových označení je dostupný na webu HathiTrust⁸⁶.



Graf 3: Autorsko-právní metadata česky psaných dokumentů v HTDL

Z grafu je zřejmé, že v dokumentech, které nejsou označeny jako „chráněno autorským právem“, jsou ještě možnosti, jak některé z nich zpřístupnit. Nejslibnější možnost představuje skupina dokumentů, u kterých nejsou autorsko-právní metadata definována. Zde je vidět, že projekty jako například CRMS, mají rozhodně smysl. Je pomocí nich možné

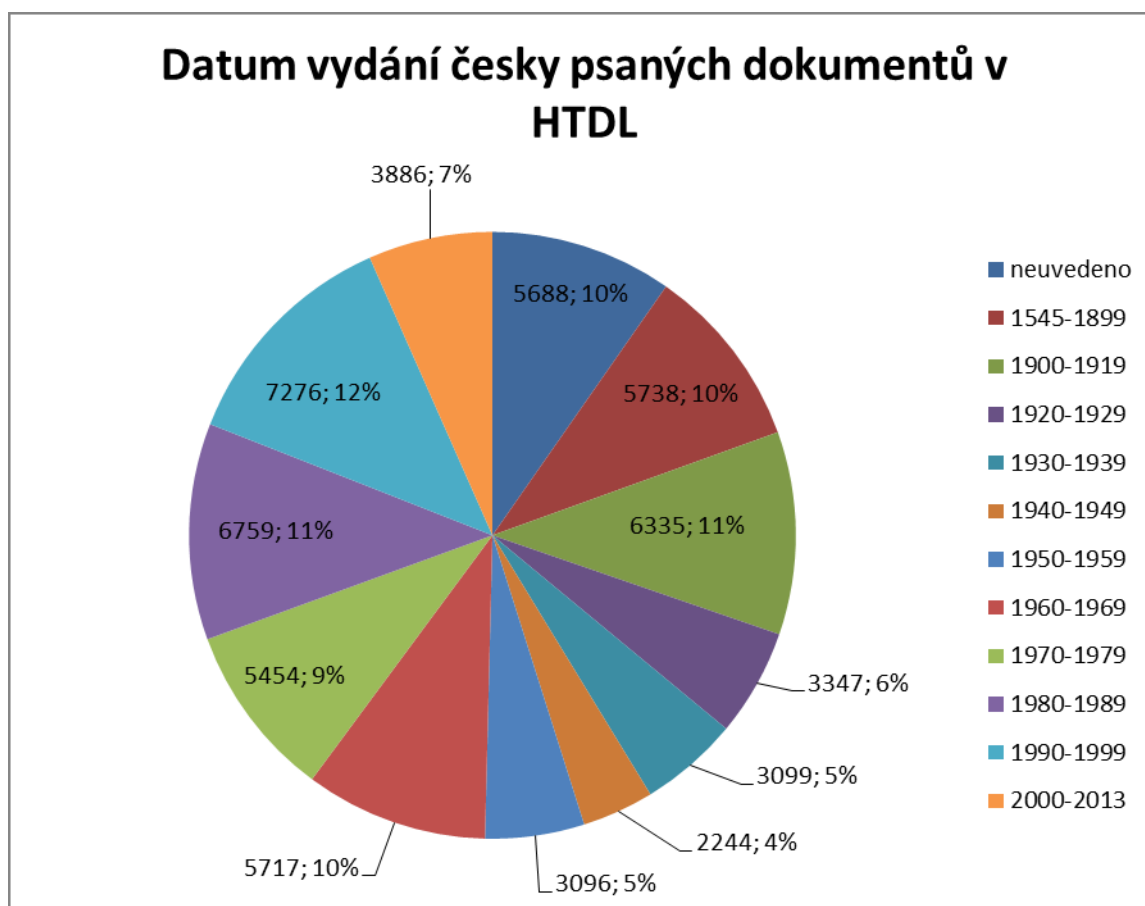
⁸⁶ Informace a rozdělení autorsko-právních metadat je dostupné zde: https://www.hathitrust.org/rights_database

rozšířit přístup k dokumentům, které do té doby nebyly dostupné, jelikož prostě nebyla známa autorsko-právní metadata. Problémem je objem HTDL a časová náročnost procesu zjišťování těchto údajů. Další možností, jak se dostat k plným textům z České republiky nedostupných dokumentů (volně dostupné z US IP adres), je již [výše popsaná](#) možnost použití amerického proxy serveru.

3.3.1 Datum vydání dokumentu

Dalším sledovaným údajem bylo datum vydání. Tyto údaje byly rozděleny celkem do dvanácti kategorií. První z nich je situace, kdy není datum vydání v záznamu uvedeno nebo je uvedeno v chybné formě. Tento údaj se z metadatového záznamu stahuje z pole 008 a pokud je uveden například jako rozsah dat, datum vydání se v metadatach objeví takto: „0000-00-00“⁸⁷. Další skupinu tvoří dokumenty, vydané v rozmezí let 1545-1899. Tento rozsah byl vybrán na základě předpokladu, že v něm bude v HTDL poměrně málo dokumentů. Stejným předpokladem se řídilo vytvoření skupiny dokumentů, které byly vydány mezi lety 1900-1919. Následující skupiny, kromě poslední, jsou rozděleny po desetiletích. S datem vydání velmi souvisejí možnosti zpřístupnění plných textů. Na základě tohoto údaje se totiž také určují autorsko-právní metadata, která pak umožňují dokumenty zpřístupnit. Jelikož u česky psaných dokumentů tento údaj v poměrně velkém množství záznamů chybí, existuje zde možnost, že by v případě jejich uvedení (či správného zápisu), bylo volně dostupné větší množství těchto dokumentů, než je tomu nyní. Následující graf znázorňuje podíl dokumentů vydaných v jednotlivých rozsazích let.

⁸⁷ Tuto situaci lze vidět například u tohoto záznamu: <https://catalog.hathitrust.org/Record/000622907>



Graf 4: Datum vydání česky psaných dokumentů v HTDL

Nejmenší skupinu představují dokumenty vydané ve čtyřicátých letech 20. století (4 %), následují padesátá a třicátá léta (5 %), čtvrtou nejmenší kategorií jsou dokumenty vydané ve dvacátých letech (6 %). 7 % dokumentů bylo vydáno mezi lety 2000-2013 a 9% v sedmdesátých letech 20. století. Po 10 % obsahují tyto skupiny dokumentů: dokumenty vydané mezi lety 1545-1899, dokumenty vydané v šedesátých letech a nakonec dokumenty, v jejichž záznamech není uvedeno datum (nebo je uvedeno chybně). 11 % dokumentů bylo vydáno mezi lety 1900-1919 a stejné procento dokumentů bylo vydáno také v osmdesátých letech. Nejvíce dokumentů (12 %) bylo vydáno v devadesátých letech 20. století. Na základě těchto informací lze uvést, že více než polovina česky psaných dokumentů v HTDL byla vydána mezi lety 1950-2013. Podobně je na tom vlastně obsah HTDL celkově ([viz výše](#)).

Podle metadat jednotek dokumentů, jejichž pomocí byla analýza vytvořena, byly dva nejstarší dokumenty v češtině (z těch dokumentů, které měly v záznamu uvedeno datum vydání), vydány v letech 1545 a 1580. Ovšem po prohlédnutí kompletního záznamu ve

formátu MARC⁸⁸ vyšlo najevo, že jde o vícesvazkový dokument, který vycházel v letech 1877-1900. Údaje o jednotlivých svazcích (jednotkách) jsou v záznamu rozepsány v opakovatelném poli 974. Z nich je zřejmé, že osm svazků této publikace bylo (samozřejmě spolu s metadatovými záznamy) přijato z Pensylvánské státní univerzity. Zbývající dva svazky byly přijaty z Chicagské univerzity a jejich záznamy byly zjevně špatně zpracovány. V podpoli pro datum vydání je totiž uvedeno datum (rok), který souvisí s obsahem dokumentu, nikoliv s datem jeho vydání. Název dokumentu totiž zní: *Sněmy české od leta 1526 až po naši dobu. c.1 v.1 1526-45*. Druhý svazek (který měl být vydán v roce 1580) zahrnoval informace o sněmech českých v letech 1577-1580 a byl vydán v roce 1887⁸⁹. Tato chyba v popisu způsobila, že jsou oba tyto svazky volně dostupné v plném textu, i když by být neměly⁹⁰.

Na základě výše uvedené chyby jsem se rozhodla zjistit, zda je situace obdobná také u dokumentů vydaných v pozdější době (do konce 18. století) nebo zda jde pouze o ojedinělou chybu. V případě dokumentů, vydaných v průběhu 17. století, jde v podstatě o tentýž problém. V tomto období bylo podle dostupných dat vydáno sedm dokumentů. Při zobrazení záznamu bylo zjištěno, že ve skutečnosti vyšly v padesátých letech 20. století⁹¹. Opět se jedná o jeden vícesvazkový dokument, tentokrát s názvem: *Prameny k československým dějinám vojenským*. Jako u předchozího dokumentu, i zde byl na základě dat vydání špatně určen přístup k plným textům. U šesti svazků, kde je chybně uvedeno datum, je k dispozici plný text dokumentu. Poslední dokument, který měl být podle metadat vydán v roce 1666, byl popsán správně.

Dokumenty, vydané v 18. století (celkem šlo o čtrnáct jednotek) byly v tomto období skutečně vydány, pouze u dvou dokumentů bylo chybně uvedeno, že jsou psány v češtině. V obou případech šlo o latinu⁹². Tento údaj byl ověřen jednak rychlým prohlédnutím plného textu dokumentu v HTDL, jednak ověřením v Souborném katalogu ČR. Tuto informaci jsem se pokusila ověřit také v databázi Worldcat pomocí čísel OCLC. U dokumentu s názvem *Tentamen physico-experimentale...* bylo skutečně uvedeno, že je psán v latině, ovšem

⁸⁸ Odkaz na kompletní záznam: <https://catalog.hathitrust.org/Record/012289018.marc>. Jde o jeden dokument s více jednotkami, kdy jedna měla být vydána v roce 1545 a druhá v roce 1580.

⁸⁹ Data byla ověřena na plných textech obou svazků.

⁹⁰ Ostatní svazky se správnými metadaty volně dostupné nejsou

⁹¹ Odkaz na kompletní záznam: <https://catalog.hathitrust.org/Record/000677903.marc>

⁹² Jde o tyto dva dokumenty: <http://catalog.hathitrust.org/Record/011561575> a <http://catalog.hathitrust.org/Record/001985660>

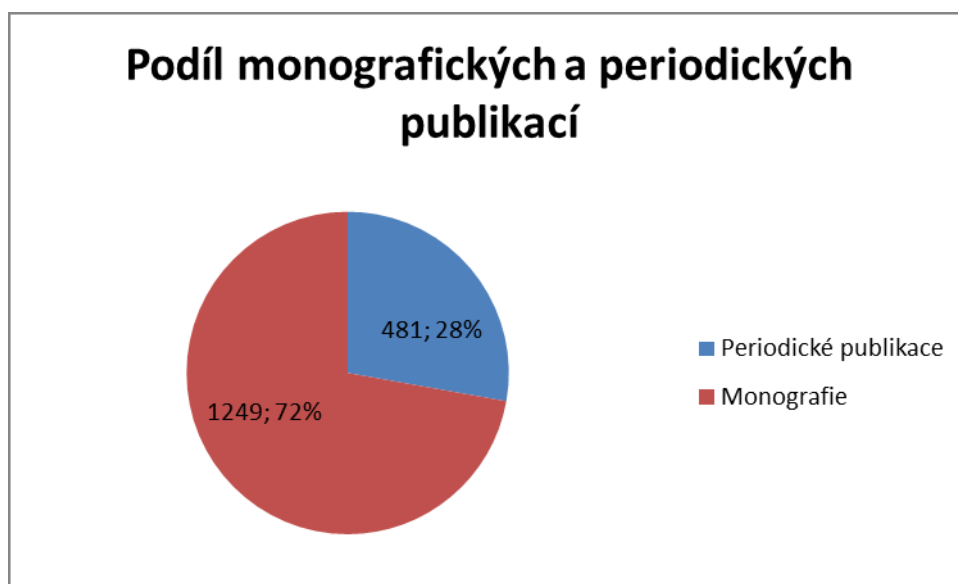
u druhého dokumentu *Scriptorum rerum bohemicarum*⁹³, byla jako jazyk dokumentu chybně uvedena čeština.

Dokumenty, u nichž není k dispozici údaj o datu vydání

Další součást analýzy je věnována těm dokumentům, u nichž nebylo uvedeno datum vydání. Jak je vidět v Grafu 4 (viz [výše](#)), bylo jich celkem 5 688. Protože jde o poměrně velké množství dat, rozhodla jsem se analyzovat záznamy celých dokumentů, ne všech jednotek. Ty jsem získala odstraněním duplicitních jednotek podle údaje o trvalém odkazu na záznam v katalogu, který obsahuje jeho jednoznačný identifikátor. Více jednotek mělo tedy totožný identifikátor záznamu. Tyto duplicity byly odstraněny ve dvou krocích opět pomocí Excelu. Nejprve byly duplicity označeny pomocí funkce *Podmíněné formátování*. Záznamy s opakujícím se identifikátorem v trvalém odkazu byly následně ručně odstraněny. Z původních 5 688 záznamů jednotek vzniklo 1 730 záznamů dokumentů.

Záměrem bylo zjistit, z jakého důvodu u těchto dokumentů není uvedeno datum. Na základě předpokladu, že u periodik je často datum v záznamu uvedeno chybně nebo vůbec, jsem tyto dokumenty rozdělila na monografie a periodika. Nejprve bylo z 1 730 záznamů odstraněno 232 záznamů, v nichž byl uveden identifikátor ISBN. U těchto dokumentů se dá předpokládat, že jde o monografie. K ostatním záznamům (1 498) byl ručně doplněn údaj, zda jde o periodikum či monografii. Tato informace byla odvozena z údajů o názvu dokumentu a jménu autora. V případě sporných situací byla ověřována v návěští záznamu v HTDL v pozici 07.

⁹³ Záznam dokumentu v databázi Worldcat: <http://www.worldcat.org/oclc/4254142>



Graf 5: Podíl monografických a periodických publikací v česky psaných dokumentech v HTDL, u nichž není uvedeno datum vydání

Na základě výše popsané analýzy bylo zjištěno, že téměř třetinu dokumentů, v jejichž záznamu nebylo uvedeno datum, tvoří periodika. Podle mého názoru to může být způsobeno tím, že u záznamů periodických publikací je často jako datum vydání používáno rozsahu dat (např. 1941- nebo 1941-9999 apod.), která jsou systémem pro správu metadat Zephir analyzována jako neplatná (tedy vznikne výše uvedený zápis 0000-00-00). Další možností je, že přispívající instituce do metadat jednotky datum vydání vůbec neuvedla. Mezi periodickými dokumenty nebyly pouze časopisy, ale také ročenky, zprávy, sborníky apod. Co se týče monografií, šlo ve většině případů o vícesvazkové publikace, jako například spisy určitých autorů, slovníky, atlasy a samozřejmě šlo také o samostatné publikace, u kterých nebylo uvedeno datum vydání.

3.3.2 Místo vydání

Dalším údajem, sledovaným u česky psaných dokumentů (jednotek) v HTLD, byla země vydání. Protože se tento údaj v metadatech o jednotkách ve sbírce nenachází, bylo pro sledování této informace použito funkce filtrování výsledků ve fulltextovém vyhledávání v rozhraní HTDL. Údaj o zemi vydání se z MARC záznamu přebírá z pole 008, z pozic 15-17, kde je tato informace uvedena ve formě kódu. Pro ověření kódu byl použit jejich oficiální seznam⁹⁴. Pro zobrazení výsledků nebyl použit graf, jako u ostatních údajů, ale tabulka. Tato možnost byla vybrána kvůli vysokému počtu různých míst vydání, které by

⁹⁴ Kompletní seznam kódů je dostupný na webu Kongresové knihovny: https://www.loc.gov/marc/countries/countries_code.html#c

v grafu nebyly tak zřetelné. Následující tabulka obsahuje údaj o názvu země, kde byl dokument vydán, počet jednotek, které zde byly vydány a také jejich procentuální zastoupení v celkovém počtu.

Země vydání	Počet jednotek	Procentuální zastoupení
Československo a ČR	53 380	91%
Neuvedeno	2 259	4%
Panamské průplavové pásmo (Canal Zone)	716	1,22%
USA	577	0,98%
Německo	315	0,54%
Velká Británie	301	0,51%
Slovensko	283	0,48%
Panama	189	0,32%
Kanada	186	0,32%
Švýcarsko	84	0,14%
Rakousko	67	0,11%
Francie	57	0,10%
Srbsko a Černá Hora	39	0,07%
Itálie	34	0,06%
Polsko	20	0,03%
Švédsko	18	0,03%
Německá demokratická republika	16	0,03%
Západní Sahara	15	0,03%
Maďarsko	14	0,02%
El Salvador	13	0,02%
Ruská federace	12	0,02%
Ukrajina	10	0,02%
Středoafriická republika	8	0,01%
SSSR	7	0,01%
Bulharsko	6	0,01%
Nizozemí	6	0,01%
Slovinsko	5	0,01%
Lichtenštejnsko	4	0,01%

Tabulka 8: Počty česky psaných dokumentů v HTDL podle místa vydání

Z tabulky vyplývá, že převážná většina jednotek dokumentů v češtině byla vydána v Československu nebo v České republice. Tyto dva údaje o zemi vydání byly ve filtrech odděleny, ale pro účely této analýzy byly kvůli přehlednosti sloučeny. V Československu bylo vydáno 34 117 jednotek a v České republice 19 263. Následující nejpočetnější položkou (4 %) je možnost, že údaj o zemi vydání v záznamu úplně chyběl.

Následuje překvapivě velmi početná (716) skupina dokumentů vydaných v Panamském průplavovém pásmu (Canal Zone). Jelikož jde na tuto oblast o podezřele vysoký počet dokumentů, třicet z těchto záznamů bylo náhodně vybráno a prozkoumáno. Seznam takto prověřených záznamů představuje [přílohu č. 3](#) této práce. V seznamu je vždy uveden název dokumentu, skutečné místo vydání (podle pole 260 v MARC záznamu) a odkaz do katalogu HTDL. Jde opět o jednotky, počet záznamů celých dokumentů je tedy samozřejmě nižší než 716. Všechny prohlédnuté dokumenty kromě jednoho byly podle údajů v záznamech v poli 260, vydány v ČR (jeden z nich byl vydán na Slovensku). Z dokumentů vydaných v ČR byl jeden vydán v Písku, jeden v Brně a jeden v Plzni, všechny ostatní měly jako místo vydání uvedenu Prahu. Z těchto informací by se dalo předpokládat, že všechny dokumenty (nebo jejich většina), v jejichž záznamu je uvedeno, že byly vydány v oblasti Panamského průplavového pásma, byly ve skutečnosti vydány v České republice (nebo v Československu). K pochybení zřejmě došlo při vytváření katalogových záznamů, konkrétně při vyplňování pole 008. Nejspíše došlo k popletení kódů pro Českou republiku (xr) a pro Panamské průplavové pásmo (cz). Mohlo k tomu dojít chybným, i když pochopitelným předpokladem, že kód pro zemi by měl být podobný kódovému označení jazyka (pro češtinu jde o kód cze).

Další skupinou byly dokumenty, vydané v USA. Bylo jich zjištěno celkem 577. Jsou mezi nimi knihy, časopisy, noviny, hudebniny (bylo mezi nimi mimo jiné celkem 16 děl Antonína Dvořáka⁹⁵), slovníky a další. Při prohlížení záznamů těchto dokumentů byla náhodou nalezena podobná chyba, jako u předchozí skupiny dokumentů. V záznamu časopisu *Besedy času* je opět v poli 008 chybně uvedený údaj. Podle pole 260 byl vydán v Praze⁹⁶.

Dokumentů, vydaných v Německu, Velké Británii a na Slovensku byl zhruba stejný počet (kolem 0,5 % z celkového počtu). Dokumenty, vydané ve Velké Británii, jsou ve filtrech katalogu HTDL rozděleny na ty, publikované ve Velké Británii (152) a v Anglii (149). Podobně jako Československo a ČR byly sloučeny do jedné skupiny.

Skupina dokumentů, vydaných v Panamě je opět překvapivě početná (189). Jelikož však nejde o zas až tak vysoké číslo, byly prohlédnuty všechny záznamy v této skupině. Všechny dokumenty, kromě tří, byly vydány v ČR (nebo v Československu). Zmíněné tři

⁹⁵ Například dílo Sirotek: <https://catalog.hathitrust.org/Record/002594060>

⁹⁶ Záznam časopisu *Besedy času*: <https://catalog.hathitrust.org/Record/100541630>

dokumenty byly vydány na Slovensku⁹⁷. Opět zjevně došlo k chybě (k mnoha chybám) při vyplňování pole 008. Jelikož je kód pro označení Panamy (pn) od označení ČR velmi odlišný, nepodařilo se mi odhadnout příčiny této záměny.

Záznamy dokumentů, které byly vydány v Kanadě (186) nebyly procházeny, jelikož se dá předpokládat, že v tomto případě je údaj o zemi vydání uveden správně. Tento předpoklad částečně potvrdila informace z filtrů v katalogu HTDL, že většina dokumentů vydaných v tomto státě byla vydána mezi lety 1970-1999. V tomto období na území Kanady působilo české exilové nakladatelství '68 *Publishers*. Na základě výše uvedených informací lze počtu těchto dokumentů důvěřovat.

Podle filtrování bylo 84 dokumentů vydáno ve Švýcarsku. Jelikož tento počet nebyl příliš vysoký, opět byly všechny záznamy prohlédnuty. Ukázalo se, že u převážné většiny dokumentů (77) byl tento údaj uveden správně. U jednoho dokumentu bylo jako správné místo vydání zjištěno město Beru ve Francii. Šest dokumentů bylo podle pole 260 vydáno v ČR. Vzhledem ke kódovému označení Švýcarska (sz) mohlo dojít k chybě v zápisu, kdy chtěl katalogizátor (také chybně) zapsat kód cz (Canal Zone). V záznamech dokumentů vydaných v Rakousku byl tento údaj ve všech záznamech uveden správně.

Další skupina dokumentů byla podle filtrů vydána v soustátí Srbsko a Černá Hora. Problém je nejspíše ve významu tohoto názvu v češtině a v angličtině. Podle údajů na české Wikipedii (Srbsko a Černá Hora) existoval tento státní útvar mezi lety 2003-2006, ale podle anglické verze této stránky (Serbia and Montenegro) vznikl již v roce 1992. Ovšem podle filtrů nebyl v těchto letech vydán ani jeden dokument. Správnější země původu než Srbsko a Černá Hora, by tedy bylo označení Jugoslávie, ovšem kódové označení tohoto útvaru v seznamu kódů pro země vydání není k dispozici. Nejvíce z těchto dokumentů (26) bylo vydáno na území dnešního Srbska (zejména v městech Novi Sad a Bělehrad). Druhou nejpočetnější skupinou byly dokumenty, vydané v Záhřebu (dnešní Chorvatsko), kterých bylo 16. Pět dokumentů bylo vydáno v Lublani, hlavním městě Slovinska. Ve dvou záznamech byla jako místo vydání uvedeno chorvatské město Daruvar. Po jednom dokumentu bylo vydáno v Postojně (Slovinsko), Sarajevě (Bosna a Hercegovina) a v Celje (Slovinsko).

⁹⁷ Odkazy na záznamy těchto dokumentů: <https://catalog.hathitrust.org/Record/008547127>, <https://catalog.hathitrust.org/Record/008851231> a <https://catalog.hathitrust.org/Record/009160673>

Většina dokumentů, vydaných v Itálii, vydalo nakladatelství Křesťanská akademie v Římě. Co se týče dokumentů, vydaných ve Francii, ve skutečnosti bylo pět z nich vydáno v ČR (Praha) a jeden ve Stockholmu (Švédsko). Celkem 7 z dokumentů vydaných v Polsku, má jako místo vydání v poli 260 uvedenu ČR. Dokumenty, deklarované jako vydané ve Švédsku, byly ve třech případech vydány ve Švýcarsku (Curych). Jeden dokument, který měl být vydán v Německé demokratické republice, byl publikován v ČR a jeden sice na území pozdější NDR (Lipsko), ale v roce 1915, kdy NDR ještě neexistovala.

Nepříliš početnou skupinou (15) jsou dokumenty, které byly údajně vydány v oblasti Západní Sahary. Podle pole 260 v MARC záznamu bylo 11 z nich vydáno v ČR a čtyři na Slovensku. Pravděpodobně došlo opět k chybě při zápisu kódového označení země, Západní Sahara má označení ss. Stejný případ představují skupiny jednotek, údajně vydané v El Salvadoru a ve Středoafrické republice. Ani jeden dokument v těchto zemích ve skutečnosti vydán nebyl. V ostatních skupinách dokumentů byly v údajích o místě vydání pouze jednotlivé chyby.

3.3.3 Typ dokumentu

Údaj o typu dokumentu nebyl v metadatech o sbírce uveden, tyto informace byly tedy zjišťovány pomocí filtrů ve fulltextovém vyhledávání. Po sečtení počtu dokumentů v jednotlivých typech, vyšlo najevo, že by dokumentů muselo být o mnoho více, než jich skutečně je. Mělo by jich být 58 639, ovšem po sečtení všech hodnot jich vyšlo 86 717. Toto je pravděpodobně způsobeno tím, že se údaj o typu dokumentu stahuje hned z několika míst záznamu. Nasvědčuje tomu i fakt, že se některé typy dokumentů (alespoň podle názvu) zřejmě překrývají. Pravděpodobně je použito návěští záznamu, kde je v pozici 07 zakódována bibliografická úroveň záznamu. Dalšími zdroji by také mohla být pole 006 a 008. Kombinací těchto údajů vzniklo ve filtrech k česky psaným dokumentům celkem 18 následujících typů:

- Kniha
- Seriál
- Časopis
- Biografie
- Slovníky
- Konference
- Statistiky
- Hudebniny

- Notové zápisy (Musical Score)
- Fotografie a obrazová díla
- Adresáře
- Encyklopedie
- Mikrofiše
- Noviny
- Archivy
- Rukopisy
- Mapy
- Atlasy map

Položky seznamu jsou seřazeny podle počtu v nich obsažených jednotek. Nejpočetnější skupinou jsou, podle očekávání, knihy, které tvoří 39 % celkového počtu. Následují dvě skupiny, které se velmi pravděpodobně z větší části překrývají, jde o seriály (28 %) a časopisy (26 %). Další v pořadí jsou biografie (3 %), které se nejspíše překrývají s kategorií knih. Po jednom procentu mají slovníky, konference a statistiky. Následujících 11 kategorií zaujímá celkem 4 % obsahu. Hudebniny a Notové zápisy mají totožný počet výskytu (351), je tedy docela možné, že jde o tytéž dokumenty. Nasvědčoval by tomu také fakt, že několik prvních dokumentů v obou těchto skupinách je totožných. Fotografií a obrazových děl je celkem 330. Dalšími, stejným počtem zastoupenými typy dokumentů (104), jsou adresáře a encyklopedie. V tomto případě jde však nejspíše o náhodu, po rychlém prohlédnutí několika prvních stran obou skupin, vyšlo najevo, že se tyto dokumenty nepřekrývají. Mikrofiše jsou zastoupeny v 57 případech. Následují archivy, kterých je celkem 6. Poslední tři typy dokumentů jsou zastoupeny pouze jednou, kdy v případě mapy a atlasu map jde o stejný dokument.

3.3.4 Obsah dokumentů ve vztahu k oboru Informační studia a knihovnictví

Pro zjištění údajů o obsahu česky psaných dokumentů v HTDL ve vztahu k oboru Informační studia a knihovnictví bylo opět použito sbírky *All in Czech*. Tentokrát šlo o funkci fulltextového vyhledávání v jejím obsahu. Pro vyhledávání byl použit vyhledávací termín „knihovnictví“. Toto slovo bylo vybráno také kvůli tomu, že jej není nutné doplňovat prostředky pro nahrazování znaků, protože má ve všech pádech stejný tvar. Počet jednotek česky psaných dokumentů, v nichž se výraz knihovnictví vyskytuje, je 3 657. Pouze 12 z nich je dostupných v plném textu (odkudkoliv). Výsledky vyhledávání byly nejprve seřazeny podle relevance, tedy podle počtu výskytů termínu v publikaci, ale také podle počtu nalezených

termínu na jednu stránku. Jako nejvíce relevantní se tedy jeví dokument s názvem *Bibliografie českého knihovnictví, bibliografie a VTI*, díl z roku 1989. Tento ročník obsahuje výraz knihovnictví celkem na 17 stranách. Počet výskytů na stránku se pohybuje od 1 do 33. Na sedmém místě v řazení podle relevance jsou *Dějiny knihoven a knihovnictví v českých zemích a vybrané kapitoly z obecných dějin* od Jiřího Cejпка a dalších autorů, vydané v roce 1998. V této publikaci se hledaný výraz objevuje na 66 stranách, od 12 výskytů na jednu stranu po jeden. V seznamu nejrelevantnějších dokumentů je velmi vysoko uvedena ještě jedna publikace Jiřího Cejпка. Jde o *Základy knihovnictví* z roku 1956. Mezi dalšími relevantními dokumenty byly dále například: *Adresář knihoven ČR* (1995 a 1997), další díly *Bibliografie českého knihovnictví*, různé ročníky *Ročenky Státní knihovny ČSR*, *Časopis československých knihovníků*, sborník s názvem *Knihovna*, časopis *Knihovník*, časopis *Čtenář*, dále *Přehled bází dat použitelných v knihovnách* od Aleše Brožka z roku 1997 atd. Přestože nebyly projity všechny nalezené záznamy, dalo by se uvést, že mezi nejrelevantnější česky psané dokumenty v HTDL, které se zabývají knihovnictvím, patří zejména periodika, bibliografie a několik monografií. Jako nejméně relevantní byly určeny dokumenty, ve kterých se termín knihovnictví vyskytoval pouze jednou. Šlo například o tyto dokumenty: časopis *Volná myšlenka* (1905), *Literární měsíčník* (1978) nebo časopis *Ekonom* (2007) atd. HTDL je tedy pro obor Knihovnictví v českém prostředí rozhodně relevantním zdrojem, bohužel je velmi málo těchto dokumentů dostupných v plném textu.

Aby bylo dokumenty možno dále nějakým způsobem rozdělit, byl do fulltextového vyhledávání v katalogu HTDL zadán termín „knihovnictví“ a následně byl použit filtr pro zobrazení dokumentů v českém jazyce. Tím byl získán stejný počet výsledků, jako ve sbírce, s tím rozdílem, že ve vyhledávání je prostřednictvím filtrů možné zjistit o dokumentech mnohem více informací. Typy dokumentů se, jak již bylo uvedeno výše, překrývají. Typ dokumentu seriál se v záznamech těchto jednotek objevil celkem v 3 055 případech, což je skoro 84 % ze všech jednotek, ve kterých byl nalezen termín knihovnictví. O knihu šlo v 11 % případů. Dalšími typy dokumentů, kde se hledaný termín vyskytl, byly slovníky, biografie, statistiky, adresáře, konferenční sborníky a encyklopedie, fotografie, mikrofiše a noviny (tisk). Nejvíce těchto dokumentů vlastní opět Michiganská a Kalifornská univerzita (obě dohromady mají ve svých fondech 2 293 těchto jednotek).

Závěr

Digitální knihovna HathiTrust funguje v současné době již devátým rokem. Vznikla z popudu stejnojmenného partnerství, do kterého je v současné době zapojeno více než sto, zatím stále převážně amerických univerzitních knihoven. Celé partnerství vzniklo v podstatě díky masovému digitalizačnímu projektu společnosti Google. Na rozdíl od této společnosti, jde však o partnerství čistě neziskové a až na příspěvky z grantů na dílčí projekty, v podstatě finančně soběstačné. Prostřednictvím těchto projektů vzniklo v rámci HTDL několik dílčích služeb, které mohou být a jistě již jsou, velmi přínosnými pro své uživatele, kterými jsou zejména uživatelé členských knihoven. Jde například o projekt Výzkumného centra HathiTrust, které umožňuje hloubkovou analýzu velkého množství textu. Dalším příkladem je Registr vládních dokumentů Spojených států, jehož cílem je v první fázi identifikovat a shromáždit záznamy všech vládních dokumentů USA. V další fázi půjde o jejich digitalizaci a zpřístupnění. HTDL je také držitelem certifikátu TRAC pro důvěryhodné repozitáře, členské instituce mohou tedy využít možnosti dlouhodobě archivovat dokumenty, kterými do HTDL přispěly.

Analýza uživatelského prostředí HTDL ukázala, že obsahuje mnoho užitečných funkcí, například tvorbu sbírek či podrobné filtrování. Zajímavé jsou také dva vyhledávací módy, katalogový a fulltextový. Jedinou větší slabinou v rozhraní HTDL je jeho mobilní verze, ve které bylo několik poměrně zásadních prvků zcela nefunkčních. Všechny chyby, které však byly nahlášeny, byly však velmi rychle opraveny.

Velkým problémem HTDL, a vlastně všech současných digitálních knihoven, jsou možnosti zpřístupnění plných textů dokumentů. Uživatel se většinou k plnému textu dokument nedostane, v HTDL je tato situace ještě zhoršena tím, že z českých IP adres lze zobrazit pouze dokumenty, které jsou volně dostupné po celém světě. Jde hlavně o dokumenty vydané před rokem 1877. Pro určení možnosti zpřístupnění dokumentu, jsou velmi důležitá autorsko-právní metadata, která jsou určována pomocí údajů z bibliografických záznamů.

Na základě analýzy celkového obsahu HTDL prostřednictvím interpretace dostupných statistik vyšlo najevo, že nejvíce dokumentů je psáno v angličtině. Ovšem alespoň jedním dokumentem je zde zastoupeno 460 dalších jazyků. Čeština se v HTDL vyskytuje celkem v 0,4 % dokumentů. V průběhu analýzy obsahu podle data vydání vyšly najevo první větší chyby v metadatových záznamech, kdy se objevovaly u data vydání takové údaje, jako například: 0018, 0263 apod.

Česky psané dokumenty v HTDL byly analyzovány na základě čtyř kritérií, kterými byly datum a místo vydání, typ dokumentu a jeho obsah se zaměřením na obor Informační věda a knihovnictví. Co se týče data vydání, poměrně vysoké procento česky psaných jednotek dokumentů (5 688, což je 10 %) nemá v záznamu uvedeno datum vydání. Nakonec se ukázalo, že v mnoha případech půjde zřejmě o chyby v zápisu dat vydání, kdy katalogizátoři uváděli do záznamu jednotky datum vydání celého dokumentu. Zejména u časopisů, kde se uvádí rozsah let, není bohužel systém pro správu metadat schopen tento údaj zpracovat. V některých případech byly v údajích o vydání objeveny chyby, díky nimž se změnil autorsko-právní status. Příslušné jednotky tedy byly zpřístupněny všem uživatelům, i když neměly být přístupné nikomu. Ohledně místa vydání česky psaných dokumentů, bylo v těchto údajích nalezeno poměrně velké množství chyb, které jsou ve většině případů zřejmě způsobeny chybným zápisem kódu pro označení země vydání v poli 008. Vznikla tak poměrně početná skupina jednotek česky psaných dokumentů (716), které byly údajně vydány v oblasti Panamského průplavového pásma či menší skupiny dokumentů, publikovaných v El Salvadoru, Středoafričské republice či v oblasti Západní Sahary. Nejčastěji zastoupeným typem dokumentu v češtině se ukázala být kniha, následují periodika a seriálové publikace (které se zřejmě z větší části překrývají) a další, již menším počtem zastoupené typy. Obsah HTDL, který se týká knihovnictví a je psán v češtině, se ukázal pro tento obor velmi relevantním. Problém opět nastává se zpřístupněním, kdy je z celkem 3 657 dokumentů, přístupných pouhých 12.

Jako velký problém se v analýze ukázala v některých případech častá chybovost při zpracovávání katalogizačních záznamů dokumentů (a také jejich jednotek). V takto obrovském množství dat, se dají chyby v záznamech pochopit, na druhou stranu mohou velmi mást uživatele. Například ve výše uvedených případech pouhým chybným zápisem údajů o místě vydání, vznikly ve vyhledávacím rozhraní zcela nové filtry. Uživatel tak může o tyto dokumenty zcela přijít, poněvadž ho ani nenapadne v takových kategoriích vyhledávat nebo je procházet. Všechny chyby, které byly v práci uvedeny, byly oznámeny HTDL prostřednictvím formuláře pro zpětnou vazbu. Přínosem práce by tedy časem mohlo být také zvýšení úrovně metadatových záznamů česky psaných dokumentů v HTDL.

Seznam použitých zdrojů

- BONN, Maria, 2016. Computation, corpus, community: The HathiTrust Research Center today. *College & Research Libraries News* [online]. 2016, 77(4), 194-197 [cit. 2016-10-29]. ISSN 2150-6698 (online). Dostupné z: <http://crln.acrl.org/content/77/4/194.full>.
- CENTER FOR RESEARCH LIBRARIES, 2011. *Certification Report on the HathiTrust Digital Repository*. Chicago: Center for Research Libraries, 2011, 6 s. Dostupné také z: <http://www.crl.edu/sites/default/files/reports/CRL%20HathiTrust%202011.pdf>.
- CHRISTENSON, Heather, 2011. HathiTrust: A Research Library at Web Scale. *Library Resources* [online]. 2011, 55(2), 93-102 [2016-05-18]. ISSN 0024-2527. Dostupné také z: <http://www.ebscohost.com>.
- CORNELL LAW SCHOOL, 2015. Partnership. In: *Wex* [online]. Ithaca: Legal Information Institute, 2015 [cit. 2016-12-18]. Dostupné z: <https://www.law.cornell.edu/wex/partnership>.
- COX, Krista, 2015. Authors Guild v. HathiTrust Litigation Ends in Victory for Fair Use. In: *ARL Policy Notes* [online]. 2015-01-08 [cit. 2017-06-18]. Dostupné z: <http://policynotes.arl.org/?p=837>.
- DOWNIE, J. Stephen a et al, 2016. The HathiTrust Research Center: Exploring the Full-Text Frontier. *EDUCAUSE Review* [online]. 2016, 51(3) [cit. 2016-10-29]. ISSN 1945-709X (online). Dostupné z: <http://er.educause.edu/articles/2016/5/the-hathitrust-research-center-exploring-the-full-text-frontier>.
- FRČKA, Pavel, 2014. *Vyhledávání knih Google*. Praha, 2014. 151 s. Diplomová práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková, PhD. Dostupné z: <https://is.cuni.cz/webapps/zzp/detail/108359/>.
- FURLOUGH, Mike, 2016. HathiTrust New Member Webcast 2015 [záznam webinaru]. In: *Youtube* [online]. 2017-02-22, 58 minut [cit. 2017-06-15]. Dostupné z: <https://www.youtube.com/watch?v=oDfMrIt70As>.
- FURLOUGH, Mike a Valerie GLENN, 2016a. Detecting US Federal Documents to Expand Access. In: *IFLA WLIC 2016* [online]. s. 1-10 [cit. 2016-11-26]. E-print dostupný z: <http://library.ifla.org/1380/1/099-furlough-en.pdf>.
- GLENN, Valerie, 2016. *Creating a Registry of U.S. Federal Documents* [online]. 24. 6. 2016. Prezentováno v rámci: Print Archive Network Forum, ALA Annual Meeting. [prezentace PPT]. 10 snímků. Dostupné z: https://www.crl.edu/sites/default/files/attachments/events/3_Hathitrust_FedDocs.pptx.
- HathiTrust releases Zephir, 2013. *Advanced Technology Libraries* [online]. 2013, 42(12), 5-6 [cit. 2016-11-08]. ISSN 0044-636X. Dostupné z: www.ebscohost.com.
- INDIANA UNIVERSITY LIBRARIES, 2016. HTRC User Getting Started FAQ. In: *HathiTrust Research Center Documentation* [online]. Indiana University Libraries, 2016 [cit. 2016-12-02]. Dostupné z: <https://wiki.htrc.illinois.edu/display/COM/HTRC+User+Getting+Started+FAQ>.
- INDIANA UNIVERSITY LIBRARIES, 2016a. HathiTrust Research Center Documentation [online]. Indiana University Libraries, 2016 [cit. 2016-12-10]. Dostupné z: <https://wiki.htrc.illinois.edu/display/COM/HathiTrust+Research+Center+Documentation>.
- JOHNSON, Richard K. Free (or fee) to all? *Library Journal* [online]. 2009, 134(1), 42-43 [cit. 2016-09-25]. ISSN 0363-0277. Dostupné z: www.ebscohost.com.
- LEVINE, Melissa a et al, 2016. *Finding the public domain: Copyright Review Management System Toolkit* [online]. Ann Arbor: Board of Regents, University of Michigan, 2016, 304 s. [cit.

2016-11-29]. ISBN 978-1-60785-374-9. Pod licencí CC-by dostupné z:
<http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001>.

OCLC adds Google Books, HathiTrust records to Worldcat, 2010. *Advanced Technology Libraries* [online]. 2010, 39(5), 4-5 [cit. 2016-11-08]. ISSN 0044-636X. Dostupné komerčně z:
www.ebscohost.com.

Serbia and Montenegro. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-06-19]. Dostupné z:
https://en.wikipedia.org/wiki/Serbia_and_Montenegro.

Spojené státy americké, 2011. *Copyright law of the United States and related laws contained in Title 17 of the United States code*. Washington: United States Copyright Office. Dostupné také z:
<http://www.copyright.gov/title17/circ92.pdf>.

Srbsko a Černá Hora. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-06-19]. Dostupné z:
https://cs.wikipedia.org/wiki/Srbsko_a_%C4%8Cern%C3%A1_Hora.

STINE, Kathryn, 2013. Introducing Zephir, HathiTrust's New Metadata Management System. In: *California Digital Library* [online]. California: CDL, 2013 [cit. 2016-11-11]. Dostupné z:
<http://www.cdlib.org/cdlibinfo/2013/11/07/introducing-zephir-hathitrusts-new-metadata-management-system/>.

STINE, Kathryn, 2016. *Submitting Metadata to Zephir: A Guide for HathiTrust Contributing Partner Institutions* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2017-06-13]. Dostupné z:
<https://docs.google.com/document/d/1ILpVfk5y3auLpBicflpGIJbf9bnNZ9qeJwkW48CZ1Xo/edit>.

UNITED STATES. COURT OF APPEALS (2ND CIRCUIT), 2014. *Author's guild, Inc. v. HathiTrust: No. 12-4547-cv: Decision* [online]. 2014-06-10 [cit. 2016-11-14]. Dostupné z:
<http://law.justia.com/cases/federal/appellate-courts/ca2/12-4547/12-4547-2014-06-10.html>.

UNITED STATES DISTRICT COURT. SOUTHERN DISTRICT OF NEW YORK, 2012. *Author's Guild, Inc. et al v. HathiTrust et al: No. 11 CV 6351 (HB) Opinion & Order* [online]. 2012-10-10 [cit. 2017-06-17]. Dostupné z: <https://cases.justia.com/federal/district-courts/new-york/nysdce/1:2011cv06351/384619/156/0.pdf>.

UNITED STATES DISTRICT COURT. SOUTHERN DISTRICT OF NEW YORK, 2015. *Case No. 11 Civ. 6351 (NRB) Stipulation* [online]. 2015-01-06 [cit. 2017-06-17]. Dostupné z:
https://www.hathitrust.org/documents/authors_guild_v_hathitrust_stipulation.pdf.

UNIVERSITY OF MICHIGAN, 2008. Update on August 2008 Activities. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2008 [cit. 2016-11-03]. Dostupné: https://www.hathitrust.org/updates_august2008.

UNIVERSITY OF MICHIGAN, 2008a. Update on June 2008 Activities. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2008 [cit. 2016-11-03]. Dostupné z: https://www.hathitrust.org/updates_june2008.

UNIVERSITY OF MICHIGAN, 2008b. Update on March 2008 Activities. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2008 [cit. 2016-11-03]. Dostupné z: https://www.hathitrust.org/updates_march2008.

UNIVERSITY OF MICHIGAN, 2013. Documentation. In: *U-M Library* [online]. Ann Arbor: University of Michigan, 2013 [cit. 2016-11-14]. Dostupné z: <http://www.lib.umich.edu/orphan-works/documentation>.

UNIVERSITY OF MICHIGAN, 2013a. Orphan Works Project – FAQ. In: *U-M Library* [online]. Ann Arbor: University of Michigan, 2013 [cit. 2016-11-14]. Dostupné z:
<http://www.lib.umich.edu/orphan-works/faq>.

UNIVERSITY OF MICHIGAN, 2013b. Zephir Record Score. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2013 [cit. 2017-06-12]. Dostupné z: <https://docs.google.com/document/d/1A6lq4bQBvHmqdz3Ab-CiqPpNqcXILe1bW5JCeNuW4x0/edit>.

UNIVERSITY OF MICHIGAN, 2014. Mike Furlough Selected To Be Executive Director of HathiTrust. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2014 [cit. 2016-11-20]. Dostupné z: https://www.hathitrust.org/mike_furlough_executive_director.

UNIVERSITY OF MICHIGAN, 2014a. Orphan Works Project. In: *U-M Library* [online]. Ann Arbor: University of Michigan, 2014 [cit. 2016-11-14]. Dostupné z: <http://www.lib.umich.edu/orphan-works>.

UNIVERSITY OF MICHIGAN, 2016. Advanced Catalog Search. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2017-01-06]. Dostupné z: <https://catalog.hathitrust.org/Search/Advanced>.

UNIVERSITY OF MICHIGAN, 2016a. Advanced Full-text Search. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2017-01-06]. Dostupné z: <https://babel.hathitrust.org/cgi/ls?a=page&page=advanced>.

UNIVERSITY OF MICHIGAN, 2016b. Board of Governors. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-08]. Dostupné z: https://www.hathitrust.org/board_of_governors.

UNIVERSITY OF MICHIGAN, 2016c. Bylaws. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-08]. Dostupné z: <https://www.hathitrust.org/bylaws>.

UNIVERSITY OF MICHIGAN, 2016d. Contact. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-05-18]. Dostupné z: <https://www.hathitrust.org/contact>.

UNIVERSITY OF MICHIGAN, 2016e. Copyright Review. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-29]. Dostupné z: <https://www.hathitrust.org/copyright-review>.

UNIVERSITY OF MICHIGAN, 2016f. Cost. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-30]. Dostupné z: <https://www.hathitrust.org/cost>.

UNIVERSITY OF MICHIGAN, 2016g. Governance. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-12-02]. Dostupné z: <https://www.hathitrust.org/governance>.

UNIVERSITY OF MICHIGAN, 2016h. Help - Using the Digital Library. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-12-02]. Dostupné z: <https://www.hathitrust.org/help>.

UNIVERSITY OF MICHIGAN, 2016i. HathiTrust U.S. Federal Government Documents Program. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-23]. Dostupné z: <https://www.hathitrust.org/usgovdocs>.

UNIVERSITY OF MICHIGAN, 2016j. HTRC Governance. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-12-10]. Dostupné z: https://www.hathitrust.org/htrc_governance.

UNIVERSITY OF MICHIGAN, 2016k. Other Projects and Grants. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-10-17]. Dostupné z: <https://www.hathitrust.org/other-projects-grants>.

UNIVERSITY OF MICHIGAN, 2016l. Partnership Checklist. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-10-17]. Dostupné z: https://www.hathitrust.org/partnership_checklist.

UNIVERSITY OF MICHIGAN, 2016m. Partnership Community. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-05-18]. Dostupné z: <https://www.hathitrust.org/community>.

UNIVERSITY OF MICHIGAN, 2016n. Print Holdings Information. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-10-18]. Dostupné z: https://www.hathitrust.org/print_holdings.

UNIVERSITY OF MICHIGAN, 2016o. Searching, Reading, Building Collections, and Accessibility. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2017-01-07]. Dostupné z: <https://www.hathitrust.org/access>.

UNIVERSITY OF MICHIGAN, 2016p. Shared Print Program. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-12-18]. Dostupné z: https://www.hathitrust.org/shared_print_program.

UNIVERSITY OF MICHIGAN, 2016q. Strategic Advisory Board. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-28]. Dostupné z: <https://www.hathitrust.org/sab>.

UNIVERSITY OF MICHIGAN, 2016r. Technological Profile. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2017-01-05]. Dostupné z: <https://www.hathitrust.org/technology>.

UNIVERSITY OF MICHIGAN, 2016s. United States Government Documents Registry. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-28]. Dostupné z: https://www.hathitrust.org/usdocs_registry/.

UNIVERSITY OF MICHIGAN, 2016t. Working Groups and Committees. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-12-10]. Dostupné z: https://www.hathitrust.org/working_groups.

UNIVERSITY OF MICHIGAN, 2016u. Zephir. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2016 [cit. 2016-11-08]. Dostupné z: <https://www.hathitrust.org/zephir>.

UNIVERSITY OF MICHIGAN, 2017. Access Determination for HathiTrust Objects. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-16]. Dostupné z: https://www.hathitrust.org/access_determination.

UNIVERSITY OF MICHIGAN, 2017a. Bibliographic Metadata Submission. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-07]. Dostupné z: https://www.hathitrust.org/bib_data_submission.

UNIVERSITY OF MICHIGAN, 2017b. Bibliographic Metadata Specification. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-08]. Dostupné z: https://www.hathitrust.org/bib_specifications.

UNIVERSITY OF MICHIGAN, 2017c. Collections. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-01-06]. Dostupné z: <https://babel.hathitrust.org/cgi/mb?colltype=updated>.

UNIVERSITY OF MICHIGAN, 2017d. Copyright. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-16]. Dostupné z: <https://www.hathitrust.org/copyright>.

- UNIVERSITY OF MICHIGAN, 2017e. Data Availability and APIs. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-14]. Dostupné z: <https://www.hathitrust.org/data>.
- UNIVERSITY OF MICHIGAN, 2017f. Datasets. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-14]. Dostupné z: <https://www.hathitrust.org/datasets>.
- UNIVERSITY OF MICHIGAN, 2017g. Getting Content Into HathiTrust. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-16]. Dostupné z: <https://www.hathitrust.org/ingest>.
- UNIVERSITY OF MICHIGAN, 2017h. HathiTrust Bibliographic API. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-14]. Dostupné z: https://www.hathitrust.org/bib_api.
- UNIVERSITY OF MICHIGAN, 2017i. HathiTrust Rights Database. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-08]. Dostupné z: https://www.hathitrust.org/rights_database.
- UNIVERSITY OF MICHIGAN, 2017j. Ingest Checklist. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-16]. Dostupné z: https://www.hathitrust.org/ingest_checklist.
- UNIVERSITY OF MICHIGAN, 2017k. Ingest Tools. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-06-16]. Dostupné z: https://www.hathitrust.org/ingest_tools.
- UNIVERSITY OF MICHIGAN, 2017l. Statistics and Visualizations. In: *HathiTrust Digital Library: Millions of books online* [online]. Ann Arbor: University of Michigan, 2017 [cit. 2017-01-06]. Dostupné z: https://www.hathitrust.org/statistics_visualizations.
- YORK, Jeremy, 2012. The Elephant in the Library. *Library Issues: Briefings for Faculty and Administrators* [online]. 2012, **32**(3), 1-4 [cit. 2016-05-18]. ISSN 0734-3035. Dostupné také z: <https://www.hathitrust.org/documents/HathiTrust-LI-201201.pdf>.
- ZAYTSEV, Angelina, 2015. HathiTrust and a Mission for Accessibility. *The Journal of Electronic Publishing* [online]. 2015, **18**(3) [cit. 2016-05-18]. ISSN 1080-2711. Dostupné z: <http://dx.doi.org/10.3998/3336451.0018.304>.
- ZAYTSEV, Angelina, 2016. (HTS-118991) Request for information [online]. 18. 10. 2016 16:21; [cit. 2016-10-18]. Osobní komunikace. Text emailu je [přílohou č. 1](#) této práce.
- ZAYTSEV, Angelina, 2017. Discovering the HathiTrust Digital Library Collection Webcast, 04/05/2017 [záznam webináře]. In: *Youtube* [online]. 2017-04-05, 61 minut [cit. 2017-05-31]. Dostupné z: <https://www.youtube.com/watch?v=P-S8BSuk2nI>.

Seznam obrázků

Obrázek 1: Schéma organizační struktury partnerství HathiTrust.....	15
Obrázek 2: Schéma ověřování osiřelých děl (UNIVERSITY OF MICHIGAN, 2013)	21
Obrázek 3: Záznam dokumentu v Registru vládních dokumentů Spojených států	24
Obrázek 4: Hlavní stránka HTRC.....	29
Obrázek 5: Hlavní strana webových stránek Digitální knihovny HathiTrust.....	35
Obrázek 6: Přihlašovací stránka HTDL	36
Obrázek 7: Bibliografický záznam dokumentu v online katalogu HathiTrust	39
Obrázek 8: Rozhraní pro prohlížení plných textů dokumentů v HTDL.....	41
Obrázek 9: Hlavní stránka mobilní verze HTDL.....	43
Obrázek 10: Záznam dokumentu v HTDL ve formátu MARC	46
Obrázek 11: Příklad chybně naskenované strany	53

Seznam tabulek

Tabulka 1: Výpočet členského příspěvku v HathiTrust, první část	31
Tabulka 2: Výpočet členského příspěvku v HathiTrust, druhá část	32
Tabulka 3: Chyby v mobilním rozhraní HTDL	45
Tabulka 4: Minusové body pro záznamy monografií	49
Tabulka 5: Minusové body pro záznamy periodických publikací.....	50
Tabulka 6: Podmínky dostupnosti plných textů v HTDL.....	54
Tabulka 7: Vybraná autorsko-právní metadata	56
Tabulka 8: Počty česky psaných dokumentů v HTDL podle místa vydání	71

Seznam grafů

Graf 1: Rozdělení obsahu HTDL podle jazyka dokumentu	62
Graf 2: Rozdělení obsahu HTDL podle data vydání.....	64
Graf 3: Autorsko-právní metadata česky psaných dokumentů v HTDL.....	65
Graf 4: Datum vydání česky psaných dokumentů v HTDL	67
Graf 5: Podíl monografických a periodických publikací v česky psaných dokumentech v HTDL, u nichž není uvedeno datum vydání.....	70

Seznam příloh na CD

Příloha č. 1: Metadata k dokumentům ve sbírce All in Czech, soubor s názvem „AllinCzech“ ve formátu .txt.

Přílohy

Příloha č. 2: Text emailu na dotazy ohledně fungování HTDL

ZAYTSEV, Angelina, 2016. (*HTS-118991*) *Request for information* [online]. 18. 10. 2016 16:21; [cit. 2016-10-18]. Osobní komunikace.

Hello Jana Křížová,

Thank you for contacting us with your request. I've provided answers to your questions below.

1. How HathiTrust cooperate with OCLC? How does it work?

- Records in HathiTrust may have originated as OCLC records. In most cases, there is an OCLC number in every record, as a requirement of deposit. HathiTrust syncs nightly with OCLC, and they create electronic versions of HT records to include in OCLC.

2. Are there really digital copies in HTDL (all on one place), or are they at member's servers and HTDL just links to them?

- The content listed in our catalog is stored on the HathiTrust servers and preserved by HathiTrust. This is different from services like Europeana or DPLA, which serve more as a centralized gateway to other websites.

3. How are records of document produced? Are they taken from OCLC or something? And it is possible to get a record in another format, for example in MODS?

- Records come originally from contributor's catalogs and may be original records or copy records from OCLC. Please see https://www.hathitrust.org/bib_metadata_correction for more information. We use Marc and MarcXML formats exclusively for bibliographic metadata.

Thank you,
Angelina Zaytsev
Collection Services Librarian, HathiTrust

Příloha č. 3: Dokumenty, údajně vydané v Panamském průplavovém pásmu a jejich skutečná místa vydání

Název dokumentu	Místo vydání	Odkaz do HTDL
Pražský pitaval	ČR (Praha)	https://catalog.hathitrust.org/Record/001744584
Všeobecný církevní dějepis	ČR (Praha)	https://catalog.hathitrust.org/Record/001931001
Pražský sborník historický	ČR (Praha)	https://catalog.hathitrust.org/Record/000518067
Slovenská vlastiveda : zemepis a dejepis pre slovenské ľudové školy	ČR (Praha)	https://catalog.hathitrust.org/Record/100683605
Obzor národohospodársky; časopis venovaný otázkám národohospodárskym a socialněpolitickým.	ČR (Praha)	https://catalog.hathitrust.org/Record/100156574
Plyn	ČR (Praha)	https://catalog.hathitrust.org/Record/011456884
Soud hraničný. Z dějin českého práva	ČR (Praha)	https://catalog.hathitrust.org/Record/000905398
Dvanáct hodin v pancéřové síni	ČR (Praha)	https://catalog.hathitrust.org/Record/012410821
Za vítězství socialismu; stati a projevy	ČR (Praha)	https://catalog.hathitrust.org/Record/001341327
Božena Němcová	ČR (Praha)	https://catalog.hathitrust.org/Record/007370256
Bulharsko na ceste k socialisticke revoluci	ČR (Praha)	https://catalog.hathitrust.org/Record/010143997
R. U. R. Rossum's Universal Robots; kolektivní drama o vstupní komedii a třech dějstvích	ČR (Praha)	https://catalog.hathitrust.org/Record/000886497
Urbánkův věstník bibliografický	ČR (Praha)	https://catalog.hathitrust.org/Record/008607542
Vypsání husitské války	ČR (Praha)	https://catalog.hathitrust.org/Record/006009575
Páv zpívá o štěstí	ČR (Praha)	https://catalog.hathitrust.org/Record/007278055
Památky archaeologické	ČR (Praha)	https://catalog.hathitrust.org/Record/000636517
Země zamyšlená	ČR (Praha)	https://catalog.hathitrust.org/Record/001965251
Acta Universitatis Carolinae. Iuridica	ČR (Praha)	https://catalog.hathitrust.org/Record/000060849
Klavír a housle : čtyři povídky	ČR (Praha)	https://catalog.hathitrust.org/Record/100640015
Politický katechism	ČR (Praha)	https://catalog.hathitrust.org/Record/001747726
Řepné virosy a jejich přenašeči (Monografie o virových chorobách řepy s ohledem na potřeby praxe i vědy)	ČR (Praha)	https://catalog.hathitrust.org/Record/009084744
Natašina vina	ČR (Praha)	https://catalog.hathitrust.org/Record/012203429
Dějiny královského krajského města Písku nad Otavou	ČR (Písek)	https://catalog.hathitrust.org/Record/007036527
Západní Slované a Germáni	ČR (Brno)	https://catalog.hathitrust.org/Record/008235867
Živena; národní almanach	Slovensko	https://catalog.hathitrust.org/Record/012407970
K podmínkam vedecke prace	ČR (Praha)	https://catalog.hathitrust.org/Record/010149411
Westöhmen in Goethes Leben, Werk und Wirken	ČR (Plzeň)	https://catalog.hathitrust.org/Record/009101727
Za sošialisticheskuiu sel'skokhoziâistvennuuiu nauku	ČR (Praha)	https://catalog.hathitrust.org/Record/007328904
Lásky hra osudná	ČR (Praha)	https://catalog.hathitrust.org/Record/100784057

