

KARLOVA UNIVERZITA V PRAZE  
FILOZOFICKÁ FAKULTA  
ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ

Informační věda



Disertační práce

Aplikace data miningu v oboru informačních studií:  
Analýza chování uživatelů knihovny

*Školitel:*

prof. RNDr. Jiří Ivánek, CSc.

*Autor:*

Mgr. Anna Janíková

2016

CHARLES UNIVERSITY IN PRAGUE  
FACULTY OF ARTS  
INSTITUTE OF INFORMATION STUDIES AND LIBRARIANSHIP

Information Science



## Disertation Thesis

Usage of Data Mining in Information Studies:  
Analysis of User's Behavior of a Library

*Supervisor:*

prof. RNDr. Jiří Ivánek, CSc.

*Author:*

Mgr. Anna Janíková

2016

## **Prohlášení**

Prohlašuji, že jsem disertační práci napsala samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury, a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne

Jméno a příjmení

## **Klíčová slova**

Data mining – knihovna – uživatel – knihovní software – data – datová analýza – jazyk R

## **Keywords**

Data mining – library – user – library management software – data – data analysis – R language

## **Abstrakt**

Práce analyzuje možnosti použití data miningu na datech knihovny, konkrétně na datech týkajících se uživatelů a jejich chování zaznamenaném v softwaru Clavius. Využívá k tomu data Univerzitní knihovny Slezské univerzity v Opavě. Práce nejdříve vysvětluje pojem data mining, věnuje se různým způsobům použití data miningu v knihovnách a poté navazuje rozbohem využitelnosti dostupných dat ze systému Clavius. Dále se zaměřuje na problém, jakými daty lze podpořit management knihovny a v konečné části práce po provedení data miningu rozebírá zjištěné skutečnosti.

## **Abstract**

This thesis deals with usage of data mining on data of a library. It specializes in data of users and their behaviour registered by library management system Clavius. For this purpose the data of University Library of Silesian University in Opava were used. Firstly the data mining term is explained and it's use in libraries is thoroughly described. Secondly the ways of usage of the Clavius data are discussed then the text follows data mining usage for management of the library in particular. Finally the results of data mining concerning user's behaviour is described in the rear of this thesis.

# Obsah

<b>Úvod</b>	<b>1</b>
<b>1 Data mining</b>	<b>3</b>
1.1 Metodiky data miningu . . . . .	5
1.2 Technologie a software . . . . .	7
1.3 Metody data miningu . . . . .	10
1.4 Data mining a právní otázky . . . . .	11
<b>2 Data mining v knihovnách</b>	<b>14</b>
2.1 Použití data miningu ve fyzických knihovnách . . . . .	15
2.1.1 Akvizice s využitím DM . . . . .	16
2.1.2 Analýzy záznamů v katalogu . . . . .	18
2.1.3 Systémy pro podporu rozhodování . . . . .	20
2.1.4 Rozdělování financí . . . . .	24
2.1.5 Služby a data mining . . . . .	25
2.1.6 Evaluace OPACu a webu knihovny . . . . .	27
2.1.7 Průzkum chování uživatelů pomocí DM . . . . .	29
2.1.8 Analýzy dat ze sociálních sítí . . . . .	36
2.2 Použití data miningu v digitálních knihovnách . . . . .	36
2.2.1 Teoretické práce . . . . .	36
2.2.2 Praktické práce . . . . .	37
2.2.3 Text mining . . . . .	38
2.3 Bibliomining . . . . .	38
2.4 Data mining a knihovny v Česku a na Slovensku . . . . .	39
<b>3 Analýza dat z knihovního softwaru Clavius</b>	<b>41</b>
3.1 Cíle praktické části . . . . .	43
3.2 Postup . . . . .	44
3.3 Současná situace UKSU . . . . .	45
3.4 Využití dat z knihovnického softwaru Clavius . . . . .	50
3.4.1 Akvizice a data z Clavia . . . . .	50
3.4.2 Analýza bibliografických záznamů v na datech z Clavia . . . . .	52
3.4.3 Systémy pro podporu rozhodování a Clavius . . . . .	54
3.4.4 Rozdělování financí a data z Clavia . . . . .	54
3.4.5 Služby a DM na datech z Clavia . . . . .	55

## Obsah

3.4.6	Evaluace OPACu či webu knihovny pomocí dat z Clavia . . .	56
3.4.7	Uživ. chování a vliv knihovny na retenci studentů . . . . .	57
3.4.8	Uživ. chování a využívání fondu . . . . .	58
3.4.9	Uživ. chování a organizace fondu . . . . .	59
3.4.10	Analýza dat ze sociálních sítí – vyhodnocení . . . . .	60
3.4.11	Sumarizace výsledků . . . . .	60
3.5	Podpora rozhodování v knihovně pomocí DM . . . . .	61
3.5.1	Vyhodnocení – okruhy problémů . . . . .	63
3.5.2	Vyhodnocení – Data mining nad daty z Clavia . . . . .	64
3.5.3	Problémy při počítání statistik . . . . .	67
3.5.4	Analýzy – rozložení uživatelů . . . . .	70
3.5.5	Analýzy – vzorce chování . . . . .	78
	<b>Závěr</b>	<b>95</b>
	<b>Příloha: R kódy použitých algoritmů</b>	<b>109</b>

## Seznam obrázků

1.1	Graf sousloví data mining v publikacích. . . . .	3
1.2	Metodika CRISP-DM. . . . .	6
3.1	Vzorová statistika výpůjček v měsíci. . . . .	65
3.2	Vzorová statistika výpůjček ve dnech v měsíci. . . . .	66
3.3	Vzorová statistika výpůjček MVS na měsíc. . . . .	66
3.4	Vzorová statistika návštěv ve dnech v týdnu za měsíc. . . . .	67
3.5	Vzorová věková statistika čtenářů. . . . .	68
3.6	Vzorová statistika návštěv. . . . .	68
3.7	Vzorová statistika sezn. čtenářů s vypůjčenými dokumenty. . . . .	69
3.8	Aktivní uživatelé na kategorii a fakultu. . . . .	71
3.9	Aktivní uživatelé k velikosti fakulty a ke kategorii. . . . .	72
3.10	Stálí uživatelé z fakult na kategorii. . . . .	72
3.11	Po tři roky aktivní uživatelé z fakult a v kategoriích. . . . .	73
3.12	V jednom roce aktivní uživatelé z fakult a v kategoriích. . . . .	74
3.13	Typy uživatelů v jednom roce, kteří si vypůjčili. . . . .	75
3.14	Typy uživatelů v jednom roce, kteří se přihlásili. . . . .	76
3.15	Stálí uživatelé v kategoriích ve výpůjčkách. . . . .	77
3.16	Stálí uživatelé v přihlašování do OPACu. . . . .	77
3.17	Rozložení žen a mužů na fakultách. . . . .	78
3.18	Aktivní ženy a muži k potenciálním uživatelům. . . . .	79
3.19	Vypůjčování publikací, přihlašování a zobrazování webu. . . . .	80
3.20	Přihlašování do OPACu v akademických letech. . . . .	82
3.21	Přihlašování do OPACu v týdnu v ak. letech. . . . .	82
3.22	Přihlašování v hodinách ve dnech v týdnu. . . . .	84
3.23	Přihlášení kategorií uživatelů v průběhu roku. . . . .	85
3.24	Prům. přihlášení na kategorii v ak. letech. . . . .	86
3.25	Histogram návratovosti v přihlašování do OPACu. . . . .	88
3.26	Poměr návratovosti v přihlašování do OPACu. . . . .	89
3.27	Návratovost v přihlašování do OPACu. . . . .	90
3.28	Půjčené dokum. k době výpůjčky a ke kategorii. . . . .	92
3.29	Histogram půjčování dokumentů. . . . .	92
3.30	Korelace přihlášení a akce uživatele v tabulce výpůjčky. . . . .	93

## Seznam tabulek

3.1	Počty studentů v ak. letech 2011/2012 až 2014/2015 . . . . .	47
3.2	Počty pracovníků v akademických letech 2011/2012 až 2014/2015. . .	48
3.3	Počty potenc. uživ. UKSU v ak. letech 2011/2012 až 2014/2015. . .	48
3.4	Statistiky UKSU v ak. letech 2011/2012 až 2014/2015. . . . .	49
3.5	Souhrn způsobů použití dat z Clavia . . . . .	62
3.6	Výpůjčky UKSU v ak. letech 2011/2012 až 2014/2015. . . . .	67
3.7	Potenciální uživatelé k počtu aktivních uživ. v ak. letech. . . . .	86
3.8	Rozložení kvantilů v délce výpůjční doby pro kategorie. . . . .	91



## Seznam příloh použitých algoritmů

1	Sledování aktivity přihlašování do OPACu. . . . .	110
2	Skript pro určení korelace mezi přihlašováním do OPACu a záznamu v tabulce kvypujcky. . . . .	118
3	Skript pro určení vztahu mezi přihlašováním uživatelů a jejich aktivitou v knihovně. . . . .	120

# Úvod

Data mining (dále jen DM) je známou a používanou metodou převážně v komerční sféře, kde pomáhá firmám s detekcí podvodů, zjišťováním nákupního chování, klasifikací a segmentací zákazníků, sumarizací údajů atd.; k tomu využívá techniky z informatiky a statistiky. Úkolem je objevit fakta, která jsou předem neznámá.

Na konci 90. let 20. století a v první dekádě 21. století se o toto téma začal zajímat také knihovnický svět; začaly se objevovat první články, které ukazovaly, že DM je metoda použitelná také v knihovnách. Celý koncept data miningu souvisí s principem, který říká, že když máme data a musíme je uchovávat, můžeme je také využít k získávání dalších znalostí. Také knihovny disponují nejen knihami, ale také záznamy o nich; mají informace o čtenářích, výpůjčkách, objednávkách knih a dalších záležitostech ze života knihovny. Nejen ve světě obchodu, ale také v knihovnách tedy může být DM použit k evaluaci činností, k lepšímu rozhodování a efektivnějšímu fungování.

Data mining je známějším tématem hlavně v zahraničních knihovnách, v Čechách mnoho článků s jeho aplikací nenajdeme. Ovšem také české knihovny disponují databázemi, na kterých lze tento druh analýz provádět.

Protože je téma data miningu v českých knihovnách tak málo známé, nejdříve se jej budeme snažit představit.

Toto představení bude zahrnovat vysvětlení pojmů, které se na něj váží; metodik, technologií a softwaru, které se k jeho provádění používají, dále metod, jejichž použití na data se za data mining označuje, a nakonec se podíváme na právní a etické otázky, kterými by se každý výzkumník měl zabývat.

Pak se dostaneme k prvnímu cíli této práce, analýze použití data miningu v knihovnách; zjištění, k čemu jej knihovny využívají. Metodou zpracování bude průzkum literatury. Zdrojem, ze kterého budeme čerpat, budou články v odborných databázích oboru Knihovnictví a informační věda (Library Information Science & Technology Abstracts, Library & Information Science Source, Library and Information Science Abstract), dále české a slovenské odborné časopisy (Knihovna, Knihovna Plus, ProInflow, Knížnica a ITlib).

Druhým cílem bude zjistit, jaké je použití data miningu na dostupných datech. Použijeme data z knihovního softwaru Clavius, využívaného Univerzitní knihovnou Slezské univerzity v Opavě, dostupná pro účely této práce.

Knihovnický software pro knihovnu reprezentuje klíčový zdroj dat, proto se budeme zabývat zpracováním dat z něj. Příklady pro použití data miningu bu-

deme čerpat z prostudovaných knihovnických článků, protože v jejich případě předpokládáme, že jde o ověřené způsoby použití.

Posledním cílem bude provedení data miningu na datech Univerzitní knihovny Slezské univerzity v Opavě, vycházející z jejích potřeb (a z možností softwaru Clavius). Protože má knihovna dlouhodobě problémy s ochotou zřizovatele jí naslouchat, budeme se primárně zaměřovat na analýzy, které by knihovně mohly pomoci v manažerských rozhodnutích a podpořily by její pozici u zřizovatele.

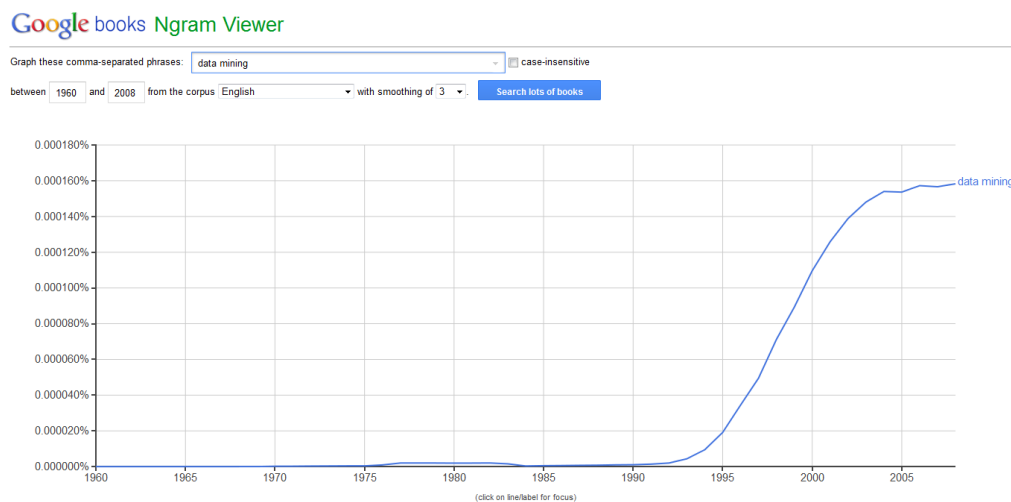
Tohoto nejlépe dosáhneme data miningem zaměřeným na uživatele a jejich chování (zaznamenané v knihovním softwaru), jelikož reflektuje míru využití knihovny jednotlivými skupinami uživatelů a může podpořit její argumenty při vyjednávání.

Protože software Clavius základními statistikami využití knihovny disponuje, bude v práci vymezeno, kde jsou hranice jeho statistik, za níž znalosti o uživateli v systému nejdou, a které mohou být zároveň data miningovým zpracováním dat posunuty.

# 1 Data mining

První zmínky o data miningu<sup>1</sup> pocházejí od statistiků ze 60. let, kdy se jako „data fishing“ nebo „data dredging“ označoval špatný zvyk hledat v datech závislosti bez předchozího stanovení hypotézy. Později (v devadesátých letech) se v komunitě databázových specialistů vžil termín „database mining“ a ještě později jeho zkrácená podoba data mining.<sup>2</sup>

Jako další termíny, označující totéž, se používaly: „data archaeology“, „information harvesting“, „information discovery“ atd.<sup>3</sup> Frekvenci použití termínu data mining ilustruje graf na Obrázku č. 1.1, který zobrazuje jeho uvedení v anglicky psaných monografiích.



Obrázek 1.1: Graf používání sousloví data mining v publikacích zhotovený přes Google Ngram Viewer.

S termínem data mining těsně souvisí také termín knowledge discovery in databases (dále jen KDD). Fayyad, Shapiro a Smyth vymezili ve svém článku

<sup>1</sup>Termín „data mining“ je do jisté míry nevhodným označením, protože se nejedná o vytěžování dat, ale o vytěžování znalostí z dat.

<sup>2</sup>PIATETSKY-SHAPIRO, Gregory. *Lesson: Data Mining, and Knowledge Discovery: An Introduction*. In: *KDnuggets* [online]. [cit. 2015-05-20]. Dostupný z: ([http://www.kdnuggets.com/data\\_mining\\_course/x1-intro-to-data-mining-notes.html](http://www.kdnuggets.com/data_mining_course/x1-intro-to-data-mining-notes.html)).

<sup>3</sup>Piatetsky-Shapiro, Gregory, 2015.

v roce 1997 KDD jako proces hledání dříve neznámých a užitečných znalostí v datech, k tomuto cíli měl KDD využívat metod data miningu.<sup>4</sup> Data mining byl tedy chápán jako užší termín, který zahrnuje iteraci procesů k objevení konkrétních závislostí v datech. Mnoho autorů<sup>5</sup> od té doby však pro stejný rozsah termínu KDD použilo termín data mining. Termíny jsou používány jako synonyma v takovém rozsahu, že je jako taková budeme uvažovat i pro účely této práce.

S tématem DM je spojena řada termínů, které v různých fázích vývoje zpracování dat přicházejí do módy, a za čas zase mizí. Do jisté míry se významy mohou kryt, někdy se specializují, jindy přibírají další hlediska. Mezi termíny, které se obsahem kryjí, mohou patřit výše zmíněné „data archaeology“,<sup>6</sup> KDD atd. Jaké je však spojení DM a dalších (nyní populárních) termínů jako „data science“, „data analysis“, „predictive analytics“, „web mining“ a „text mining“, „business intelligence“, „decision support system“ a také problematika „big data“?

Data analysis označuje datové analýzy obecně, nezáleží na využitím nástrojů (jazyk R, Weka, Excel). Patří sem jak vizualizace, tak data mining, statistické postupy atd.<sup>7</sup> Prvním z dalších dvou termínů je big data. Tak se označují data, která není možné pro jejich množství a multiplatformitu zpracovat běžným způsobem. Obsahují podrobné informace o nejrůznějších jevech. S těmito daty pak pracuje data science, snaží se je analyzovat, zpracovat do statistik a nacházet v nich nejrůznější korelace.<sup>8</sup> Data science k tomuto účelu využívá statistiku, vizualizace a samozřejmě DM. Protože jsou big data poměrně komplexním seskupením, je velmi těžké je smysluplně zpracovat bez využití DM.

Nástroj pro podporu rozhodování (decision support system) může znamenat nástroj pro analýzu dat, často s využitím DM (avšak ne nutně, viz technologie OLAP níže). Tento nástroj slouží jako zdroj informací pro lepší rozhodování managementu. Pokud k tomu přidáme ještě reporting (tj. sestavování zpráv) a možnost vizualizací, můžeme jej shrnout pod termín business intelligence. Pokud termín rozšíříme na různé úrovně zpracování informací z různých částí podniku, mluvíme již o informačním managementu.<sup>9</sup>

---

<sup>4</sup>FAYYAD, Usama et al. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. Fall 1996, roč. 17. ISSN 0738-4602.

<sup>5</sup>Včetně Piatetsky-Shapira a dalších. (Piatetsky-Shapiro, Gregory, 2015)

<sup>6</sup>Odvozeniny od tohoto termínu používá například Nicholson v podobě „internet archeology“, „digital library archeology“. (NICHOLSON, Scott. *Bibliomining.com* [online]. [cit. 2015-05-20]. Dostupný z: <http://www.bibliomining.com/>)

<sup>7</sup>Data Analysis. In: *BusinessDictionary.com* [online]. WebFinance, c2015, [cit. 2015-06-12]. Dostupný z: <http://www.businessdictionary.com/definition/data-analysis.html>.

<sup>8</sup>Jako příklad použití big data může sloužit formule F1 nebo Německá fotbalová reprezentace. (Co umějí Big Data? Třeba pošlou Formuli 1 do boxů nebo poradí fotbalistům. In: *Technet.cz*, 2015)

<sup>9</sup>Business Intelligence. In: *Technopedia* [online]. Janalta Interactive, c2015, [cit. 2015-06-13]. Dostupný z: <http://www.techopedia.com/definition/345/business-intelligence-bi>.

Web mining a text mining označují DM aplikovaný na určitý problém. Text mining, DM aplikovaný na text,<sup>10</sup> bývá používán v textových databázích například ke zjišťování o čem text je, nebo k analýze podobnosti textů mezi sebou.<sup>11</sup> Web mining je zase DM aplikovaný na data z webu, nebo o webu.<sup>12,13</sup> Kromě jiného slouží také k CRM (customer relationship management).

### 1.1 Metodiky data miningu

Pro správné provedení data miningu a zajištění možnosti převedení znalostí do jiných projektů se používá řada metodik. Některé jsou typické pouze pro konkrétní softwarové produkty, jiné jsou na platformě nezávislé. Uvedeme zde tři nejpoužívanější.<sup>14</sup>

Na platformě závislá metodika A5 (od firmy SPSS) má název odvozen od pěti kroků, které je nutné s daty provést. Jsou to: zhodnocení, kterými úkoly se budeme zabývat a jaká data budou potřeba (assess); zajištění přístupu k datům a také příprava dat (access); provedení potřebných akcí k vyhodnocení dat (analýze), tento krok je poměrně složitý a zahrnuje diskuzi nad výsledky a jejich přeměnu v doporučení. Ze znalostí, nalezených v datech, se stávají doporučení pro konání (act). Posledním krokem je převedení znalostí do praxe a zajištění opakovatelnosti procesu (automate).<sup>15</sup>

Další metodikou je SEMMA od firmy SAS, což je akronym slov sample – explore – modify – model – assess. V první fázi (sample) jsou vybírány vhodné objekty k modelování. Ve fázi druhé (explore) probíhá objevování vazeb a závislostí v datech, ve třetí (modify) selekce a úprava dat. Ve čtvrté (model) pak analýza s použitím různých metod a poslední krok (assess) zahrnuje zhodnocení modelu, jeho interpretaci a případné porovnání s dalšími výsledky.<sup>16</sup> Poslední krok metodiky SEMMA může být zároveň impulzem k průzkumu nového problému, který začíná zase od začátku.

Poslední metodikou, která je na rozdíl od předchozích dvou na platformě nezávislá, je CRISP-DM (Cross-Industry Standard Process for Data Mining). Vznikla na základě spolupráce několika velkých společností díky snaze vytvořit univerzální postup, který bude použitelný v nejrůznějších aplikacích. Sestává ze šesti kroků, jejichž posloupnost je dána pouze volně; mezi jednotlivými operacemi lze různě

<sup>10</sup>Podobně jako u DM se nejedná o vytěžování textu, ale vytěžování znalostí z textu.

<sup>11</sup>Text Mining. In: *Statistics – Textbook* [online]. Dell, c2015, [cit. 2015-06-13]. Dostupný z: <http://documents.software.dell.com/Statistics/Textbook/Text-Mining>.

<sup>12</sup>Web Mining. In: *Technopedia* [online]. Janalta Interactive, c2015, [cit. 2015-06-13]. Dostupný z: <http://www.techopedia.com/definition/15634/web-mining>.

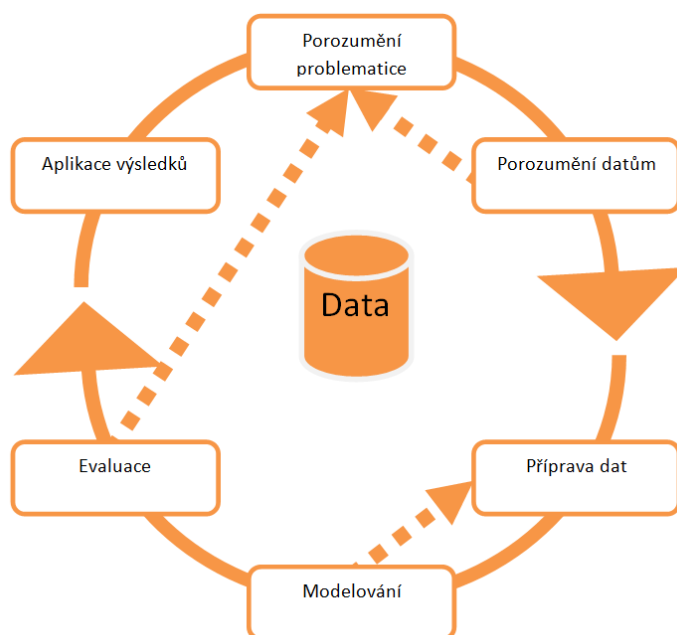
<sup>13</sup>Bývají pomocí něj zpracována data o užívání webu, struktuře nebo obsahu webu. (Web Mining. In: *Technopedia*, 2015)

<sup>14</sup>BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9.

<sup>15</sup>BERKA, Petr, 2003, s. 22–23.

<sup>16</sup>BERKA, Petr, 2003, s. 23–24.

přecházet.<sup>17</sup> Nejlépe tuto metodiku ilustruje Obrázek č.1.2. Jednotlivé fáze jsou charakterizovány takto:



Obrázek 1.2: Metodika CRISP-DM.

- Pochopení problematiky, zahrnující manažersky formulovaný úkol.
- Porozumění datům, které kromě sběru dat obsahuje také deskriptivní charakteristiku a vytipování zajímavých skupin.
- Dále příprava dat, jejich čištění, úpravy, převody, integrace a selekce.
- Fáze modelování zahrnuje výběr vhodných technik a volbu jejich nastavení, případně testování správnosti výsledků.
- Evaluace zahrnuje interpretaci a hodnocení výsledků.
- Aplikace výsledků skrývá transformaci zjištěných znalostí do vhodné formy (např. závěrečná zpráva).<sup>18</sup>

V rámci návaznosti na další kapitoly by bylo dobré zdůraznit několik myšlenek, které se nám zdají jako klíčové. Všechny zmíněné metodiky vyzdvihují kromě

<sup>17</sup>BERKA, Petr, 2003, s. 24–25.

<sup>18</sup>BERKA, Petr, 2003, s. 24–28.

aplikace sofistikovaných metod a zhodnocení výsledků také neméně důležitý krok pochopení kontextu. Pochopení kontextu je zásadní pro správný výběr techniky zpracování dat a také správnou interpretaci výsledků. V případě, že tento krok není správně proveden, není možné při nejlepší snaze z výsledků cokoli vyvodit.

Metodika CRISP-DM dokonce rozkládá tento důležitý krok do dvou částí, porozumění problému a porozumění datům. Obojí je velmi důležité, a proto CRISP-DM zároveň zajišťuje možnost cyklického pohybu mezi oběma fázemi. Umožňuje před využitím výsledků přechod na začátek procesu. Tento krok můžeme též považovat za klíčový, protože v průběhu data miningu se mohou objevit nové skutečnosti, které by aplikaci výsledků do praxe ovlivnily. Z tohoto důvodu je dobré mít možnost znovu proces opakovat a zpřesňovat výsledky, které se v datech projeví.

Obecně lze říci (také ze zkušeností při zpracovávání výsledků), že po provedení pouze jedné posloupnosti daných kroků metodiky lze vyvodit jen velmi málo. Kroky je někdy nutné provádět opakovaně, přičemž posloupnost mezi nimi není vždy formálně dodržena. Ve věku big data se dokonce ukazuje produktivnějším stanovení vlastní metodiky v závislosti na zpracovávaném úkolu.<sup>19</sup>

V průběhu této práce se budeme držet metodiky CRISP-DM, umožňující volnější pohyb mezi jednotlivými fázemi zpracování dat.

## 1.2 Technologie a software

Protože je proces vytěžování z dat velmi komplexní, je velmi těžké jej provádět bez vhodných nástrojů. Bylo by možné například databázi prozkoumávat pomocí SQL dotazů a udělat si obrázek o tom, jaké výsledky se v datech nacházejí,<sup>20</sup> ovšem byla by to práce úmorná a zároveň těžko zopakovatelná na nových datech. SQL databáze nejsou stavěny na to, aby prováděly statistické nebo pokročilejší operace s daty. Proto se často k analýzám používají zakoupené nástroje, například od firmy SPSS<sup>21</sup> nebo od firmy SAS.<sup>22</sup> Obě firmy jsou vedoucími hráči na trhu v této oblasti.<sup>23</sup>

---

<sup>19</sup>PIATETSKY, Gregory. *CRISP-DM, still the top methodology for analytics, data mining, or data science projects*. In: *KDnuggets* [online]. Oct 28, 2014, [cit. 2016-01-20]. Dostupný z: (<http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>).

<sup>20</sup>Příklad takového přístupu v českém prostředí knihoven představuje bakalářská práce Michaly Soškové. (SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného on-line katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)))

<sup>21</sup>Detaily viz webová stránka: <http://www-01.ibm.com/software/cz/analytics/spss/>

<sup>22</sup>Další informace o softwaru viz webová stránka: [http://www.sas.com/cs\\_cz/home.html](http://www.sas.com/cs_cz/home.html)

<sup>23</sup>Kromě těchto dvou jsou samozřejmě využívány také další programy STATISTICA Data Miner, RapidMiner, Stata a mnoho dalších.



Ačkoli si společnosti SPSS a SAS stále drží na trhu majoritní podíl, existuje veřejná diskuze o tom, jak dlouho tomu tak bude.<sup>24</sup>

Je pravda, že již od počátku se na scéně nacházely nástroje, které umožňovaly zpracování dat bez vazby na komerční software. Příkladem může být Weka (Waikato Environment for Knowledge Analysis), distribuovaný pod licencí GNU GPL, jehož první verze byla vypuštěna v roce 1997.<sup>25</sup> Kromě tohoto softwaru existuje samozřejmě také řada dalších nástrojů pro datovou analýzu a data mining, operujících např. pod licencí GNU, BSD, Apache nebo alespoň jako shareware. Jedním z nich je jazyk R, přímý potomek jazyka S,<sup>26</sup> distribuovaný též pod licencí GNU GPL.<sup>27</sup> Pro tento jazyk tvoří uživatelské prostředí například R Studio, vyvíjené jako Open Source.<sup>28</sup>

Jazyk R se používá pro úlohy, jako je lineární a nelineární modelování, analýza časových řad, běžné statistické testy, klasifikace, shlukování a vizualizace. Je možné jej rozšiřovat pomocí nainstalovaných balíčků.<sup>29</sup> Díky tomu, že se jedná o Open Source software, jsou rozšiřovací balíčky velmi často dostupné ihned po vydání nové metody zpracování dat.

Výhodou je tedy to, že je zadarmo, lze volně rozšiřovat jeho funkcionalitu, a dává uživateli plnou kontrolu nad prováděnými operacemi.<sup>30</sup> Komerční softwary naproti tomu umožňují pouze specifické operace v každém typu programu,<sup>31</sup> uživatel nemá kontrolu nad prováděnými operacemi, a každý program stojí velké množství peněz.

Hlavně pro vědce je důležité vědět, co se při vyhodnocování děje. Podobně v rámci týmové spolupráce je jednodušší, aby si všichni stáhli program, který je zadarmo, bez nutnosti zajišťovat nové licenční klíče. U všech GNU softwarů je samozřejmě má podrobná dokumentace ke každému z balíčků.

---

<sup>24</sup>The DataCamp Team. What is the best statistical programming language? Infograph. In: *DataCamp* [online]. June 3rd, 2014, [cit. 2016-01-20]. Dostupný z: (<https://www.datacamp.com/community/tutorials/statistical-language-wars-the-infograph>); STAMPER, Jason. Open source BI firms refute SAS CEO's criticism. In: *Computer Bussiness Review* [online]. Digital Media Group, 15:33, November 25, 2010, [cit. 2016-01-20]. Dostupný z: (<http://www.cbronline.com/news/open-source-bi-firms-refute-sas-ceos-criticism>).

<sup>25</sup>Machine Learning Group at the University of Waikato. *Weka 3: Data Mining Software in Java*, 2016.

<sup>26</sup>Jazyk S byl vyvinut v Bellových laboratořích statistikem Johnem Chambersem. (*The R Project for Statistical Computing* [online]. R Foundation, [cit. 2015-06-20]. Dostupný z: (<http://www.r-project.org/>))

<sup>27</sup>*The R Project for Statistical Computing*, 2015.

<sup>28</sup>*R Studio*, 2016.

<sup>29</sup>*The R Project for Statistical Computing*, 2015.

<sup>30</sup>*The R Project for Statistical Computing*, 2015.

<sup>31</sup>*SPSS Statistics: Put the power of advanced statistical analysis in your hands* [online]. IBM, [cit. 2016-01-20]. Dostupný z: (<http://www-01.ibm.com/software/analytics/spss/products/statistics/buy-now.html>).

Pro účely této práce byly jako softwarové nástroje datové analýzy zvoleny jazyk R a jeho IDE prostředí R Studio. Oba nástroje umožňují vytvářet skripty, které mohou být znovu využívány v případě opakování analýz na novějších datech.

### Datový sklad, datové tržiště a OLAP

Pro usnadnění datových analýz se v komerční sféře často ukládají připravená schémata dat do speciální relační databáze, které se říká datový sklad. Data jsou do ní vkládána z různých zdrojů,<sup>32</sup> klade se přitom důraz na objekt (komplexnost i za cenu redundantnosti). Proto jsou tyto databáze mnohem větší než provozní. Slouží však k jiným účelům, a data jsou v ní aktualizována třeba jen jednou ročně.<sup>33</sup> Vytvoření datového skladu je časově náročné, nicméně, je v něm jednodušší analýzy opakovat.

Podmnožinami datového skladu jsou datová tržiště, která zpřístupňují data konkrétního charakteru. Umožňují provádění analýz určitému pracovnímu úseku (např. marketingu, výrobě).<sup>34,35</sup>

Čistě pro vizualizaci dat slouží nástroj OLAP (On-Line Analytical Processing), kde jsou data uložena v několika-dimenzionálních datových kostkách (tzv. data cube), umožňujících rychlou analýzu dat. Jejich zaměření je jiné než u DM. Tento nástroj poskytuje rychlý přehled. Bývá využíván při zpracovávání zpráv jako nástroj business intelligence, neumí ale objevovat nové souvislosti v datech.<sup>36</sup>

Všechny zmíněné metody mají své výhody vzhledem k řízení instituce a jejímu informačnímu managementu obecně. Bohužel, ani jedna z nich nebývá v České republice běžně implementována do provozu knihoven.

Konstrukce datových skladů vyžaduje plánování, jasnou organizační strukturu dat a softwarové zabezpečení, aby se staly pevnou součástí oběhu informací v instituci. Podobně tomu je u technologie OLAP a datových tržišť.

Ačkoliv mohou být tyto technologie knihovnam přínosem, nebudeme se jim v rámci této práce věnovat.

---

<sup>32</sup>Při vkládání dat musí dojít k načtení, čištění a transformaci dat a zároveň k zestejnění označení, či ukazatelů v databázi. (BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9, s. 41)

<sup>33</sup>BERKA, Petr, 2003, s. 41–43.

<sup>34</sup>BERKA, Petr, 2003, s. 41–43.

<sup>35</sup>Datová tržiště. In: *IBM Knowledge Center* [online]. IBM, listopad 2012, [cit. 2016-01-20]. Dostupný z: ([http://www-01.ibm.com/support/knowledgecenter/SSGU8G\\_12.1.0/com.ibm.acc.doc/ids\\_acc\\_012.htm?lang=cs](http://www-01.ibm.com/support/knowledgecenter/SSGU8G_12.1.0/com.ibm.acc.doc/ids_acc_012.htm?lang=cs)).

<sup>36</sup>BERKA, Petr, 2003, s. 35–41.

## 1.3 Metody data miningu

Volba metody souvisí s úlohami, které chceme provést a zároveň souvisí s daty, na něž je hodláme aplikovat. Nejdříve se podívejme na typ úloh, které můžeme data miningu zadat:

- klasifikace,
- predikce,
- deskripce,
- hledání „zajímavých“ zjištění.<sup>37</sup>

Klasifikace znamená nalézt v datech takové vzory, které by pomohly rozlišit jednotlivé kategorie od sebe, aby bylo možné klasifikovat nové případy. Příklad z knihovnického prostředí může být takovýto: je-li nový uživatel studentem prvního ročníku určitého oboru a má-li zapsány jisté předměty, bude potřebovat tuto literaturu.<sup>38</sup>

Predikce se snaží z dosavadních dat zjišťovat budoucí trendy. Když například knihovna pořadí konkrétní akci a má k dispozici data o vytížení knihovny z několika předchozích podobných akcí, může predikovat jejich průběh. Podobně u zavádění nových služeb, nákupu fondu, příchodu nových uživatelů atd. Možnosti predikce závisí na účelu a dostupných datech.

Deskriptivní úlohy pomáhají pochopit zákonitosti, které jsou v datech zachyceny. Jsou účinným nástrojem managementu a marketingu. V knihovnickém světě se například mluví o „data-driven acquisition“ (také DDA).<sup>39</sup> DDA zahrnuje takové postupy, které podle záznamů o půjčovnosti fondu (preferenci formátu, tématické oblasti, nebo určitého autora), vyhledávanosti termínů (dle titulu, tématu, autora), nebo poskytnutím meziknihovní výpůjční služby, zjišťuje publikace, které by uživatelé rádi v knihovně měli.

Hledání „zajímavých“ zjištění je v podstatě obecné opsání podstaty problému. Může jít o detekci podvodů (například v pojišťovnictví), anomálií, které je nutné opravit (např. chyba v nahrávání určitých souborů), odchylek od normálního chování atd. V prostředí knihoven může jít například o detekci náročných uživatelů, nebo abnormálně půjčovaných publikací.

---

<sup>37</sup>BERKA, Petr, 2003, s. 18.

<sup>38</sup>Toto konkrétní užití DM je mezi univerzitními knihovnami poměrně vzácné, protože vyžaduje propojení více systémů, nicméně není vyloučené. O propojení školního systému s knihovním systémem pojednává například Renaud. (RENAUD, John et al. Mining Library and University Data to Understand Library Use Patterns. *The Electronic Library*. 2015, roč. 33, č. 3. Dostupný také z: (<http://dx.doi.org/10.1108/EL-07-2013-0136>))

<sup>39</sup>MITCHELL, Anna. Data Driven Acquisition in the Library. *YouTube* [online]. Publikováno 9. 11. 2015, [cit. 2016-01-02]. Dostupný z: (<https://youtu.be/CKbtmRCRemc>).

Máme-li představu o úloze, kterou bychom nechali zpracovat, musíme se podívat na data, která jsou k dispozici. Příklad z klasifikační úlohy vyžaduje, aby měla knihovna k zpřístupněna data o vyučovaných a zapsaných předmětech jednotlivých studentů ze systému školy. V případě, že tato data není možné získat, nebude možné analýzu provést. Existence a přístup k datům jsou limitujícím faktorem stejně, jako struktura dat. Strukturou dat myslíme v tomto případě to, co je zaznamenáváno. Jestliže máme záznam o studentovi, jeho studentském čísle, jméně, adrese, ročníku i předmětech, ale nemáme záznam o tom, zda je stále studentem, bude mít naše analýza jen omezenou platnost. S tímto aspektem na mysli přikročíme k rozdělení metod.

Jasně rozlišení, které z metod informačních technologií se používají pro DM, může být zrádné. Některé knihovnické články, ačkoliv mají v názvu data mining, při bližším zkoumání pojednávají ve skutečnosti spíše o datové analýze.<sup>40</sup> Knihovnický svět tedy nemá úplně jasno, co do DM metod zařadit. Proto budeme zatím diskutovat definice a úzy z oboru informačních technologií spíše, než ty, prostudované v knihovnických člancích.

Berka<sup>41</sup> například řadí mezi metody KDD: rozhodovací stromy, asociační pravidla, rozhodovací pravidla, neuronové sítě, evoluční algoritmy, bayesovskou klasifikaci, strojové učení založené na analogiích a induktivní logické programování. V části knihy „Tři zdroje“ ukazuje, že hlavní oporou DM jsou samozřejmě databáze, statistika a strojové učení. Pod DM zařazuje ze statistiky hlavně regresní, diskriminanční a shlukovou analýzu.<sup>42</sup> Jak je uvedeno v pramenech, rozdíl mezi statistikou a DM v použití statistických nebo jiných metod je v účelu použití.<sup>43</sup> Použijeme-li statistické metody pro DM, jde o DM. Použijeme-li techniku užívanou k DM k jinému účelu než je objevování závislostí v datech, o DM se nejedná.

### 1.4 Data mining a právní otázky

Existuje několik právních a etických otázek, které by bylo dobré ujasnit, než se budeme věnovat vlastnímu data miningu. Jsme si vědomi rozsáhlosti tématu; proto se nebudeme pokoušet o celkový popis, ani o právní rozbor. Půjde nám hlavně o objasnění postojů, které v této souvislosti zastáváme, a o jejich podpoření právním rámcem či precedentem.

První otázka souvisí s právem organizace ukládat údaje o uživatelích či zákaznících. Komerční firma si o své činnosti buduje databázi sama. Uživatelé jsou s uložením údajů seznámeni a podnik od nich žádá formální souhlas. Ten jej

<sup>40</sup>Použijeme-li pro datovou analýzu a DM vymezení, která jsme zmínili.

<sup>41</sup>BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9.

<sup>42</sup>BERKA, Petr, 2003, s. 46–59.

<sup>43</sup>GREINER, Lynn. What is Data Analysis and Data Mining? In: *Database Trends and Applications* [online]. c2015, [cit. 2015-06-13]. Dostupný z: (<http://www.dbta.com/Editorial/Trends-and-Applications/What-is-Data-Analysis-and-Data-Mining-73503.aspx>).

opravňuje k případnému dalšímu využívání dat pro firemní potřeby. Podobný souhlas udělují studenti knihovně, když se stávají jejími čtenáři, jak je zaneseno například v knihovním řádu Univerzitní knihovny Slezské univerzity v Opavě.

Právní řád české republiky o problematice ukládání osobních dat v knihovnách hovoří takto: „Ve smyslu zákona č. 101/2000 Sb., o ochraně osobních údajů a změn některých zákonů, ve znění pozdějších předpisů je knihovna správcem osobních údajů. Osobním údajem je podle §4 zákona č. 101/2000 Sb., jakýkoliv údaj, který se týká konkrétní osoby, jejíž identitu lze z osobních údajů přímo či nepřímo zjistit. V podmínkách knihovny jsou to identifikační údaje uživatelů, údaje o jejich výpůjčkách a jiných údajích.“<sup>44</sup> A dále: „Uživatel, který nedá souhlas se zpracováním základních identifikačních údajů, nemůže využívat služeb knihovny. Knihovna dále vede o uživateli služební údaje (registrace výpůjček, odeslané upomínky, poznámky vztahující se k porušení Knihovního řádu, údaje o provedených finančních transakcích dle §11, odst. 1 zákona č. 563/1991 Sb., o účetnictví, ve znění pozdějších předpisů).“<sup>45</sup> Knihovna tedy má právo ukládat údaje uživatelů za účelem evidence a plnění svých funkcí. Uživatel s touto evidencí dává souhlas, když se stává čtenářem.

Druhá otázka souvisí s vlastnictvím bibliografických údajů. Konkrétně knihovny se začaly touto problematikou zabývat až v době, kdy se objevil koncept Open Data,<sup>46</sup> a knihovny se do něj jako veřejné instituce měly zapojit.<sup>47</sup> Koncept Open Data se u knihoven týká bibliografických údajů v databázích; bylo důležité zjistit, zda mají tento obsah právo zpřístupňovat. První z oficiálních právních vyjádření k této problematice si nechala zhotovit Národní technická knihovna.<sup>48</sup> Pro účely tohoto právního vyjádření bylo nutné definovat, čím dílem jsou bibliografické záznamy a databáze samotná a zda má knihovna právo s nimi nakládat.

---

<sup>44</sup>Univerzitní knihovna Slezské univerzity v Opavě. *Knihovní řád Univerzitní knihovny Slezské univerzity v Opavě* [online]. [cit. 2016-01-28]. Dostupný z: (<http://www.slu.cz/slu/cz/univerzitni-knihovna/pracoviste-opava/o-nas/KnihovnaUniverzity-knihovnySlezskuniverzity.pdf>), s. 4–5.

<sup>45</sup>Univerzitní knihovna Slezské univerzity v Opavě. *Knihovní řád Univerzitní knihovny Slezské univerzity v Opavě*, 2016, s. 5.

<sup>46</sup>3.0 Česká republika. *Opendata.cz* [online]. [cit. 2016-01-28]. Dostupný z: (<http://opendata.cz/cs/node/29>).

<sup>47</sup>HRABINOVÁ, Světlana. Úvod k Open Bibliographic Data aneb i knihovny mají svá skrytá místa! *Knihovna plus*. 2013, roč. 1. Dostupný také z: (<http://oldknihovna.nkp.cz/knihovnaplus131/hrabin.htm>).

<sup>48</sup>*Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition: opinio: stanovisko k právní otázce*. Objednatel Národní technická knihovna. Zhotovitel Masarykova univerzita. Právnícká fakulta. Zpracovatel Radim POLČÁK, Matěj MYŠKA. Jaromír ŠAVELKA. Brno, 2012. Dostupný z: (<http://www.techlib.cz/files/download/id/3157/open-bibliographic-data-ntk-studie-2012.pdf>).

Z tohoto vyjádření vyplývá, že bibliografické záznamy jako takové nejsou chráněny Autorským zákonem, protože nejsou definovány jako dílo.<sup>49</sup> Zároveň se na ně nevztahují předpisy o obchodním tajemství.<sup>50</sup> Dále databáze knihovny uložená v systému je chráněna autorským právem jako původní databáze.<sup>51</sup> Vykonavatelem práv k této databázi je knihovna jako majitel práv.

Následující otázku symbolizují etické zásady výzkumníka. Pro zajištění správného průběhu je třeba pracovat s anonymizovanými daty, aby nebylo možné identifikovat osobu, a pokud je to možné, zabránit možnosti jejího zpětného dohledání (například nezveřejňovat údaje typu uživatel, který si půjčil určitou knihu minulý týden).

Naposled je nutné vymezit, v jakém právním rámci budeme v této práci operovat. Přístup k datům knihovny (akademické) byl autorce této práce umožněn na základě souhlasu ředitelky Univerzitní knihovny Slezské univerzity v Opavě, Mgr. Zuzany Tiché, jakožto představitelky Univerzitní knihovny v Opavě. Tato knihovna je držitelem autorských práv k databázi uložené v knihovním softwaru Univerzitní knihovny v Opavě, jenž bude předmětem našeho zkoumání. Dále byl přístup k datům umožněn předložením podepsaného dodatku o mlčenlivosti. Vzhledem k tomu, že tato práce bude po obhajobě zveřejněna v Repositáři závěrečných prací Univerzity Karlovy, bude její obsah podléhat schválení ze strany Univerzitní knihovny Slezské univerzity v Opavě. V příloze práce se ze zmíněných důvodů nemůže nacházet originální databáze poskytnutá knihovnou, ze které byly analýzy zhotoveny. Budou se tam však nacházet skripty, které sloužily k úpravě dat.

---

<sup>49</sup> *Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition: opinio: stanovisko k právní otázce*, 2012, s. 15.

<sup>50</sup> *Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition: opinio: stanovisko k právní otázce*, 2012, s. 9.

<sup>51</sup> *Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition: opinio: stanovisko k právní otázce*, 2012, s. 15.

## 2 Data mining v knihovnách

Nyní s představou o tom, co data mining je, přikročíme k tomu, jak jej lze využít (konkrétně k jakým účelům) a to hlavně v oblasti knihoven. Existuje řada článků, které byly na toto téma napsány, některé z nich použijeme jako referenci nebo jako modelový příklad.

V roce 2009 provedl Hsu-Hao Tsai bibliometrickou studii, ve které porovnával publikace z rejstříku SSCI<sup>1</sup> o DM. Během analýz věnovaných období od roku 1989 po rok 2009 zjistil, že obor Informační věda a knihovnictví v publikacích na toto téma vede. Obor Informatika a informační systémy byl až na druhém místě.<sup>2</sup> Z toho jasně vyplývá, že knihovníci a informační pracovníci nejen DM znají, ale také jej aplikují.<sup>3</sup>

V knihovnách je použití data miningu limitováno několika faktory; zaprvé, existencí dat, čili, zda se data o dané činnosti vůbec uchovávají. Zadruhé je limitován propojením systémů, protože mnoho menších knihoven používá pro evidenci různých činností různé způsoby. Zatřetí pracovníkem, musí být zaměstnána způsobilá síla, která data vyextrahuje, vyčistí, analyzuje a interpretuje. První dvě omezení odpadají v případě digitálních knihoven, u kterých je většinou zavedeno jedno prostředí, přes které se uživatelé registrují, vytvářejí si své čtenářské profily, vyhledávají, stahují nebo čtou dokumenty, atd. Data o těchto činnostech zůstávají v databázi k případné analýze, a protože se jedná o jeden systém, není třeba data stahovat, reformátovat, přepisovat či shánět z jiných zdrojů. Společným omezením pro oba druhy knihoven, digitální i fyzickou, je zaměstnání informatika/analytika, který je schopen s daty aktivně pracovat, a také z nich vyvozovat relevantní závěry. Nicméně, některé z knihoven těmito pracovníky disponují, protože svá zjištění publikovaly v odborných časopisech.

Aktivní správy dat a jejich vytěžování jsou schopny převážně knihovny akademické, tj. knihovny jejichž zřizovatelem je univerzita, vysoká škola či výzkumný

---

<sup>1</sup>Bibliografický rejstřík Social Science Citation Index od Institute for Scientific Information od Thomson Reuters.

<sup>2</sup>TSAI, Hsu-Hao. Global data mining: An empirical study of current trends, future forecasts and technology diffusions. *Expert Systems with Applications*. 2012, roč. 39, č. 9, s. 8172–8181. Dostupný také z: (<http://dx.doi.org/10.1016/j.eswa.2012.01.150>).

<sup>3</sup>Jsmo si samozřejmě vědomi případného omezení této studie z hlediska terminologie, protože informatika může pro popis článků o DM používat konkrétní metody spíše než zastřešující termín DM. O počet článků nám však nejde, důležitější je pro nás informace, že obor „Library and Information Science“ tento pojem zná a používá.

ústav.<sup>4</sup> V některých případech pochází data použitá pro DM z jiných zdrojů, například od odborných organizací (jako je OCLC nebo Association of Research Libraries). Někdy jsou dokonce data z veřejných knihoven kombinována s daty z knihoven akademických. V každém případě, našim záměrem není porovnávat, jaké knihovny zpracovávají která data, ale účel, za jakým jsou data použita. Proto nebudeme v následujících částech striktně rozlišovat akademické a veřejné knihovny až na místa, kde to bude nezbytně nutné.

Použití data miningu v knihovnách rozdělíme na použití v digitálních knihovnách a v knihovnách fyzických (tím se myslí veřejné<sup>5</sup> i univerzitní knihovny). Při využití DM ve fyzických knihovnách se zpracovávají data z procesů, které knihovna denně produkuje za účelem analyzovat chování uživatelů, optimalizovat stávající služby a predikovat výsledky zavedení nových služeb. Digitální knihovny<sup>6</sup> oproti tomu umožňují analyzovat data o veškeré elektronické interakci, proto umožňuje DM širší uplatnění.

Vzhledem k výše zmíněným omezením nás nepřekvapí, že procento zastoupení digitálních knihoven v problematice DM analýz je větší.<sup>7</sup>

### 2.1 Použití data miningu ve fyzických knihovnách

Data knihoven mohou pocházet z různých zdrojů, od dat z turniketů, přes data ze sociálních sítí, webových logů, selfchecků, data z výpůjčního protokolu atd. Pro DM analýzy se mohou tato data spojovat<sup>8</sup> a ukládat do datových skladů. Proto by nebylo směrodatné při rozdělování vycházet čistě z typu dat. Různí se také zpracování dat pomocí užitých metod a metody bývají v mnohých případech

---

<sup>4</sup>Jsmo si vědomi toho, že Knihovní zákon neboli Předpis č. 257/2001 Sb. Zákon o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb nezná termíny „akademická knihovna“, „veřejná knihovna“, „městská knihovna“, „obecní knihovna“ nebo „odborná knihovna“, (ČESKÁ REPUBLIKA. Předpis č. 257/2001 Sb. Zákon o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb: knihovní zákon. In: *Sbírka zákonů ČR*. ročník 2001. 98/2001) které budeme používat v této práci. Nicméně knihovnická terminologie tyto pojmy rozeznává a pro účely této práce vycházet z definic České terminologické databáze knihovnictví a informační vědy (dále jen TDKIV). (*KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003–, [cit. 2016-02-15]. Dostupný z: [http://aleph.nkp.cz/F/GJMY8FQJM38NRR8S2KTFEJTBDU47HR8YJX62AUNC9VRSU516-53-47429?func=file&file\\_name=find-b](http://aleph.nkp.cz/F/GJMY8FQJM38NRR8S2KTFEJTBDU47HR8YJX62AUNC9VRSU516-53-47429?func=file&file_name=find-b)))

<sup>5</sup>Myšleno v anglickém termínu „public library“.

<sup>6</sup>Jejich provoz často souvisí s existencí fyzické knihovny, ale ne nevyhnutelně.

<sup>7</sup>Některé články dokonce doporučují aplikaci DM pouze do digitálních knihoven, protože nevidí efektivní využití v knihovnách fyzických. (BANERJEE, Kyle. Is Data Mining Right for Your Library? *Computers in Libraries*. November/December 1998, s. 28–31)

<sup>8</sup>GUENTHER, Kim. Applying Data Mining Principles to Library Data Collection. *Computers in Libraries*. 2000, roč. 20, č. 4.



kombinovány, takže ani toto rozdělení by nebylo rozumné. Rozdělit data dle využití, s ohledem na povahu dat, je nevhodnější.

### 2.1.1 Akvizice s využitím DM

„Data-driven acquisition“<sup>9</sup> je, jak již bylo zmíněno výše, nově diskutovaným tématem.<sup>10</sup> Se snižováním rozpočtu knihovnam se zvýšil tlak na akviziční činnost nejen v zahraničí, ale také v Čechách.<sup>11</sup> Knihovny hledají co nejefektivnější způsob, jak obohacovat svůj fond. S tímto termínem souvisí také dříve známý termín uživatelem řízená akvizice, (v angličtině patron-driven acquisition, user-driven acquisition či demand-driven acquisition),<sup>12</sup> kdy uživatelé sami navrhnou knihovně koupit dokumentu, česky se pro tyto dokumenty vžil termín dezideráta.<sup>13</sup> „Data-driven acquisition“ pak zahrnuje jak návrhy uživatelů, tak vyhodnocení dat, která má knihovna k dispozici (tj. např. data o půjčovanosti knih).

O vytěžování z dat aplikovaném na akvizici mluví ve svém článku například C. H. Wu, který navrhuje model pro opakované získávání a vytěžování z dat, vhodných pro podporu akvizice. Tento model spoléhá na data z výpůjčního protokolu a akvizičního modulu. Navrhuje rozdělit zakoupené dokumenty dle jednotlivých témat pod katedry či ústavy univerzity (samozřejmě počítá i s tématickými přesahy) a vyhodnotit půjčovnost materiálů. Stanovuje také další kritéria, která mohou ovlivnit výši přiděleného rozpočtu na nákup knih, a zavádí výpočet, který má pomoci objektivizovat proces rozhodování pro vedení knihovny. Zároveň uvádí, že jím navrhovaný modul slouží pouze jako podpora rozhodování, umožňuje lépe získávat potřebné informace, a říká, že finální rozdělení je vždy závislé také od jiných kritérií, která si stanovuje management sám. Ve svém článku pro tento účel navrhuje konstrukci datového skladu.<sup>14</sup> O rok později v jiném článku doplnil

---

<sup>9</sup>Český termín zatím neexistuje.

<sup>10</sup>MITCHELL, Anna. Data Driven Acquisition in the Library. *YouTube* [online]. Publikováno 9. 11. 2015, [cit. 2016-01-02]. Dostupný z: <https://youtu.be/CKbtmRCRemc>.

<sup>11</sup>Ministerstvo kultury ČR. *Koncepce rozvoje knihoven ČR na léta 2011–2015 včetně internetizace knihoven* [online]. [cit. 2016-02-01]. Dostupný z: [http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce\\_rozvoje\\_knihoven\\_2011-2015.pdf](http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce_rozvoje_knihoven_2011-2015.pdf), s. 1.

<sup>12</sup>LEVINE-CLARK, Michael. Developing a Multiformat Demand-Driven Acquisition Model. *Collection Management*. 2010, roč. 35, č. 3–4, s. 201–207. Dostupný také z: <http://dx.doi.org/10.1080/01462679.2010.486965>.

<sup>13</sup>CELBOVÁ, Iva. Dezideratum. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: NK ČR, c2014 [cit. 2016-02-01]. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001182&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001182&local_base=KTD).

<sup>14</sup>WU, Chien-Hsing. Data mining applied to material acquisition budget allocation for libraries: design and development. In: *Expert Systems with Applications*. 2003, roč. 25, č. 3, s. 401–411. ISSN 0957-4174.

svůj koncept o použití asociačních pravidel, které by usnadňovaly objevení nových závislostí.<sup>15</sup>

Analýza půjčovnosti s doporučením pro akvizici reprezentuje přínosný přístup na jedné straně, ovšem také jednostranné pojetí na straně druhé. Tento systém si například neporadí s tématickými oblastmi, které ve fondu zastoupeny nejsou. Uživatelé si logicky nemohou půjčit něco, co se ve fondu nevyskytuje i kdyby si to přáli. Snad proto Wu říká, že jeho modul slouží pouze pro podporu rozhodování při akvizici, ne jako univerzální měřítko kvality akvizice.

Podobný systém jako Wu navrhuje také Kao, který extrahuje z databáze data o výpůjčkách, rozděluje je do kategorií a pak s pomocí dotazníků nechává akademické zaměstnance hodnotit tzv. sémantickou sílu jednotlivých kategorií pro jejich katedru či ústav. Sémantická síla reprezentuje míru, jak moc se daná kategorie dokumentů shoduje se zaměřením katedry či ústavu. Všechny hodnoty jsou poté porovnány, jak množství výpůjček dle kategorií, tak počet dokumentů v jednotlivých sekcích, nebo množství materiálů spadajících pod katedry či ústavy. Z hodnot je pak možné vyvodit rozpočet akvizice pro další akademický rok.<sup>16</sup>

Dále takovýto přístup aplikoval Mitsuro Kitajima při evaluaci akvizice Kyushu University Library. Podle NDC (Nippon Decimal Classification) rozdělil publikace do kategorií a rozdělil data o jejich půjčovnosti dle sledovaných období od roku 2000 do roku 2013 a od roku 2012 do roku 2013. Dle půjčovnosti pak rozhodoval, jak se mění půjčovnost a akviziční činnost v závislosti na oborech a jestli jsou nějaké rozdíly mezi časovými obdobími. V průzkumu nezkoumal podobnost mezi tématy. Zjistil, že složení fondu ve velké míře odpovídá frekvenci půjčovnosti publikací. Zvyky ve vypůjčování knih se mezi jednotlivými obory (např. sociální a přírodní vědy) dramaticky liší, a že by bylo vhodné doplnit fond o publikace v konkrétních kategoriích, které nebyly v knihovně tak bohatě zastoupeny.<sup>17</sup>

Nevýhody předchozích pojetí postrádá Yangův na webu založený systém pro doporučování nákupu knih. Ten využívá slov z vyhledávání, která nevrátila žádné hity.<sup>18</sup> Tato slova pak řadí dle frekvence a podle společného výskytu sestavuje

---

<sup>15</sup>WU, Chien-Hsing et al. Knowledge discovery applied to material acquisitions for libraries. In: *Information Processing and Management*. 2004, roč. 40, č. 6, s. 709–725. ISSN 0306-4573.

<sup>16</sup>KAO, S.-C. et al. Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management*. 2003, roč. 39, č. 1, s. 133–147. ISSN 0306-4573.

<sup>17</sup>KITAJIMA, Mitsuro et al. An Evaluation of Book Selection in a University Library by Loan Record Analysis. *International Journal of Information and Education Technology*. 2015, roč. 5, č. 10, s. 728–731. ISSN 2010-3689.

<sup>18</sup>V definici TDKIV se pod pojmem hit myslí: „Měrná jednotka výsledku rešeršního dotazu, která udává počet vyhledaných, věcně relevantních záznamů.“ (CELBOVÁ, Ludmila. hit. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003– [cit. 2016-02-15]. Dostupný z: ([http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000532&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000532&local_base=KTD)))

tezaurus slov. Slova pak spojuje s kategoriemi knih a popisy od prodejců a automaticky sestavuje seznamy doporučených knih ke koupi.<sup>19,20</sup>

Podobně jako Wuiv modul umožňoval evaluaci výpůjček a zanedbával chybějící témata/knihy, Yangův systém sice chybějící témata či knihy pokrývá, ale zhodnocení nejpůjčovanějších knih či témat nenabízí. Ideální by z hlediska akvizičního procesu bylo samozřejmě provést obojí.<sup>21</sup>

Bylo by dobré říci, že všechny navrhované systémy slouží jak pro podporu akvizičního procesu, tak pro podporu rozhodování při akvizici. Bylo by je tedy možné zařadit do podkapitoly „Systémy pro podporu rozhodování“, avšak kvůli jejich specifickému zaměření jim byla vyčleněna tato kategorie.

### 2.1.2 Analýzy záznamů v katalogu

Průzkum Williama Dentona se zaměřuje na data z katalogu. Snaží se vizualizovat informace, schované v záznamech milionů knih, kterými knihovna disponuje. Použil data těchto knihoven: San Francisco Public Library, Toronto Public Library, University of Toronto, University of Prince Edward Island a York University Libraries. Ze záznamů extrahoval kódy DDC (Dewey Decimal Classification) nebo LCC (Library of Congress Classification), tyto promítl do grafů a typově pak výsledky porovnal mezi sebou. Článek zároveň obsahuje části zdrojových kódů do programů Ruby a R, aby čtenáři mohli postup opakovat pro data z jiné knihovny.<sup>22</sup>

Do tohoto druhu datových analýz lze zařadit i snahy výzkumníků při OCLC a jejich články (jedná se o tyto autory: Timothy Dickey, Brian F. Lavoie, Lynn Silipigni Connaway a Edward T. O'Neill). Vzhledem k tomu, že OCLC disponuje největší databází knih na světě (WorldCat), mají jejich datoví analytici v rukou neocenitelný zdroj informací. Je samozřejmě, že výsledky analýz se OCLC snaží proměnit v něco, co přinese další finance.

Timothy Dickey například využil kontinuity a širokého záběru WorldCatu a analyzoval jednotlivé sbírky ve vybraných zemích v čase. Hledal zajímavé výkyvy v datech (v počtu publikací, počtu svazků, počtu překladů a počtu knihoven, které jsou členem OCLC). Vybral si šest zemí, Bolívii, Chile, Německo, Polsko, Jihoafrickou republiku a Thajsko. Na první pohled jde o země, které jsou velmi rozdílné, ovšem to bylo od počátku záměrem. Autor se zaměřil na

---

<sup>19</sup>YANG, Shih-Ting. An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*. 2012, roč. 10, č. 2, s. 163–173. ISSN 1741-5063.

<sup>20</sup>Článek lze v tématice DM zařadit pod text mining a klasifikaci.

<sup>21</sup>Je nutné poznamenat, že oba výzkumníci svůj systém či model aplikovali ve svých knihovnách, takže nejde čistě o teoretické koncepty. Zároveň je tedy jisté, že praktické nasazení podobných technologií je proveditelné.

<sup>22</sup>DENTON, William. On dentographs, a new method of visualizing library collections. *Code4Lib Journal*. 2012, roč. 16, s. 171–179. Dostupný také z: (<http://journal.code4lib.org/articles/6300>). ISSN 1940-5758.

diverzitu a jedinečnost jednotlivých národních sbírek nikoli na statistické zobecnění. V závěrech zdůraznil, že vývoj sbírek kopíroval historický vývoj v jednotlivých zemích.<sup>23</sup> V datech byly objeveny stopy po druhé světové válce, ekonomické krizi, politických či náboženských převratech i uvolnění režimů. Například v Evropě lze vyzorovat v publikační aktivitě euro-centrismus, naproti tomu mimoevropské země jsou ovlivněny rozmachem Angličtiny.<sup>24</sup>

Článek Lynn Silipigni Connaway také využívá jako zdroj dat WorldCat. Snaží se identifikovat ve fondech tzv. poslední kopie, které jsou ve srovnání s fondy jiných knihoven unikátní, a které lze označit za ohrožené ztrátou nebo degradací. Podle výsledků navrhuje upravit digitalizační či akviziční politiku knihovny. Obecně lze říci, že se digitalizační snahy týkají svazků, které jsou unikátní pro výzkum, ekonomicky nákladné, častým užíváním se poškozují, nebo je nutné je zachovat pro další zkoumání.<sup>25</sup> Stejným způsobem jako ve jmenované studii by bylo možné promítnout tato hlediska do kritérií pro klasifikaci<sup>26</sup> záznamů knih v katalogu knihovny.

Analýzami a srovnáváním údajů o fondech členů ARL (Association of Research Libraries) a příspěvatelů do databází OCLC (jiných než z USA) se zabývala také zpráva „Changing Global Book Collection Patterns in ARL Libraries“ kolektivu autorů v čele s Mary Jackson. Během průzkumu prováděli statistická porovnání za pomoci databázových nástrojů.<sup>27</sup> Zkušenosti pak zúročili při konstrukci nástroje OCLC WorldMap, jehož úkolem bylo provádět DM nad zvolenými daty pro účely efektivní akvizice, rozvoj fondu a hlavně pro podporu rozhodování.<sup>28</sup> Kromě nástroje WorldMap OCLC přišlo také s nástrojem pro knihovny, který na základě toho, kdo danou knihu vlastní, určuje vhodnost publikace pro daný okruh knihoven.<sup>29</sup> Obě aplikace tedy vznikly za pomoci DM, využitím dat z WorldCatu a za účelem analyzovat data o fondu.

---

<sup>23</sup>Aby výsledky správně interpretoval, musel být velmi dobře obeznámen s dějinami dané země.

<sup>24</sup>DICKEY, Timothy J. Books as Expressions of Global Cultural Diversity. *Library Resources and Technical Services*. 2011, roč. 55, č. 3, s. 148–162.

<sup>25</sup>SILIPIGNI CONNAWAY, Lynn, O'NEILL, Edward T. et al. Last Copies: What's at Risk? *College and research libraries*. 2006, roč. 67, č. 4, s. 370–379. Dostupný také z: (<http://www.oclc.org/research/publications/archive/2006/connaway-cr107.pdf>). ISSN 0010-0870.

<sup>26</sup>Myšleno provedení klasifikace dat.

<sup>27</sup>JACKSON, Mary Ellen. *Changing Global Book Collection Patterns in ARL Libraries*. 2006.

<sup>28</sup>SILIPIGNI CONNAWAY, Lynn, OLSZEWSKI, Larry. *A Geographical Representation of WorldCat Resources: A Decision-Making Tool for Acquisitions and Collection Management* [online]. Charleston Conference, November 10, 2006, [cit. 2015-06-01]. Dostupný z: (<http://www.oclc.org/content/dam/research/presentations/connaway/20061110-charleston-worldmap.pdf>); WorldMap. In: *OCLC Research* [online]. OCLC, c2015, [cit. 2015-06-01]. Dostupný z: (<http://www.oclc.org/research/activities/worldmap.html>).

<sup>29</sup>Audience Level. In: *OCLC Research* [online]. OCLC, c2015, [cit. 2015-06-01]. Dostupný z: (<http://www.oclc.org/research/activities/audience.html>).

### 2.1.3 Systémy pro podporu rozhodování

Existuje několik odborných článků, ve kterých se autoři snaží vymezit, s jakými úkony může DM vedoucím knihovny pomoci.<sup>30</sup> Autorem jednoho z nich je Kevin Cullen,<sup>31</sup> který navrhuje, aby byl data miningový modul zabudovaný přímo do knihovního softwaru. Vedoucí knihoven by pak měli lepší přehled o tom, jaký typ uživatelů využívá služeb knihovny, popřípadě kterých služeb. Věděli by, která média jsou mezi uživateli nejvíce populární, do kterých dokumentů se vyplatí investovat, jak vysoké má být pojištění, popřípadě jak efektivní je práce knihovny a jednotlivých oddělení. Cullen povzbuzuje k objevování nových závislostí v datech, vedoucích k novým, nepředpokládaným zjištěním.<sup>32</sup> Navrhuje knihovníkům, aby zvyšovali požadavky na dodavatelské firmy, aby takovéto funkce do systému zahrnuli. Vybízí také ke standardizaci, aby bylo možné výsledky porovnávat mezi knihovnami.<sup>33</sup>

O standardizaci v oblasti datových skladů knihoven mluví také Nicholson ve svém článku se Su Cleyle „Approaching librarianship from the data: using bibliomining for evidence-based librarianship“. Jejich hlavním záměrem je ukázat jiný účel, ke kterému lze DM v knihovně využít. Mluví o na „důkazech založeném knihovnictví“.<sup>34</sup> Označují tak snahu nalézt v oboru knihovnictví či informační vědy takový ekvivalent výzkumu či studie, který by pokrýval problém, jenž knihovna řeší, a z jehož výsledků by mohla čerpat bez toho, aby průzkum na své náklady prováděla.

Cleyle a Nicholson argumentují, že nalezení studie s velmi podobnými podmínkami je těžké, proto navrhují, aby knihovny ukládaly svá data (transakční i výzkumná) do standardizovaných formátů datových skladů, tyto záznamy měsíčně aktualizovaly, agregovaly společně na jednom místě a umožnily tak vzniku jakési testovací platformy, kde si prostě knihovník či výzkumník navolí, jaká kritéria by měla data splňovat, a pak jen analyzuje soubor. Umožní tak vyhodnotit například nové služby, které ještě nebyly zavedeny. Jako výhodu vidí rychlost, nízkou cenu a možnost rozšíření výzkumu o mnohá hlediska. Jako nevýhodu pak vidí problém zachování soukromí uživatelů.<sup>35</sup>

---

<sup>30</sup>SCHULMAN, Sandy. Data mining: Life After Report Generators. *Information Today*. 2003, roč. 15, č. 3, s. 52. ISSN 0363-0277.; NICHOLSON, Scott, STANTON, Jeffrey M. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. 2003, s. 247–262.

<sup>31</sup>CULLEN, Kevin. Delving into data. *Library Journal*. 2005, roč. 130, č. 13, s. 30–32. ISSN 0363-0277.

<sup>32</sup>Uvádí příklad knihovny, která v datech objevila zvyšující se počet výpůjček, ale snižující se počet skříněk, které uživatelé používají. Zjistili tak, že se jim klíče od skříněk ztrácí.

<sup>33</sup>Cullen, Kevin, 2005.

<sup>34</sup>Neboli „evidence-based librarianship“ přesný český termín zatím chybí.

<sup>35</sup>CLEYLE, Su et al. Approaching librarianship from the data: using bibliomining for evidence-based librarianship. *Library hi tech*. 2006, roč. 24, č. 3, s. 369–375. ISSN 0737-8831.

Vyvstává tedy otázka, je k dispozici nějaký software, který by knihovnám pomohl orientovat se v jejich datech? Odpověď zní: je jich celá řada. Základní nástroj pro statistický přehled o tom, kolik čtenářů navštívilo knihovnu obsahuje asi každý systém, ovšem hlubší analýzu dovoluje například Director's Station (od SirsiDynix), Decision Center (od společnosti Sierra), Alma Analytics (od Ex Libris), nebo Intota Assessment (od ProQuestu), atd.<sup>36</sup> Každý z nich samozřejmě odpovídá určitému zaměření; avšak seznam dostupného softwaru nejen demonstruje zájem ze strany knihoven o podrobnější analýzu a daty podloženější rozhodování, ale také smysl pro příležitost předních světových databázových center či společností dodávajících knihovní software. Abychom však nemluvili jen o softwaru za peníze, Harvardská univerzitní knihovna vytvořila software pro analýzu a vizualizaci knihovních dat, který je dostupný jako Open source software.<sup>37</sup>

V předchozí sekci jsme zmínili snahy OCLC o provádění DM nad bibliografickými daty a zde jsme také zmínili dva autory, kteří se zasazovali o zavedení standardů, zvýšení dostupnosti dat a o možnost porovnávat knihovny mezi sebou. S takovými nástroji přišlo OCLC. Díky široké základně zákazníků z řad knihoven umožnili klientům sdílet data s různými organizacemi kvůli porovnání, ověření, či vylepšení některého aspektu knihovnické práce. K tomuto účelu slouží celá platforma WorldShare, obsahující moduly od akvizice po meziknihovní výpůjční službu.<sup>38</sup> Z nich zmiňme jen tři moduly, umožňující přímo analyzovat data, určená přímo vedoucím pracovníkům: WorldShare License Manager, WorldShare Collection Evaluation a WorldShare Report Designer.

WorldShare License Manager kromě evidence zprostředkovaných zdrojů zvládá také statistickou analýzu návštěvnosti.<sup>39</sup> WorldShare Collection Evaluation zabezpečuje krom jiného sdílení dat s ostatními knihovnami, porovnávání fondů a sledování skupin uživatelů.<sup>40</sup> A WorldShare Report Designer umožňuje stahovat data z ostatních modulů, analyticky s nimi pracovat a vytvářet zprávy o činnosti knihovny.<sup>41</sup>

---

<sup>36</sup>BREEDING, Marshall. Mining Data for Library Decision Support. *Computers in Libraries*. 2013, roč. 33, č. 5, s. 23–25. ISSN 1041-7915.

<sup>37</sup>Harvard Library. Library Analytics Toolkit. *Harvard Library Lab* [online]. Date Funded: February 1, 2011, [cit. 2016-02-01]. Dostupný z: (<https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>).

<sup>38</sup>OCLC. WorldShare *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<http://www.oclc.org/worldshare.en.html>).

<sup>39</sup>OCLC. WorldShare License Manager. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<https://www.oclc.org/license-manager.en.html>).

<sup>40</sup>OCLC. WorldShare Collection Evaluation. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<https://www.oclc.org/collection-evaluation.en.html>).

<sup>41</sup>OCLC. WorldShare Report Designer. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<http://www.oclc.org/worldshare-report-designer.en.html>).

OCLC u těchto modulů mluví o pojmu „daty řízené rozhodování“.<sup>42,43</sup> Podobně například také Nicholson hovoří o DM v knihovnách jako o efektivní podpoře rozhodování<sup>44</sup> a to zvláště při strategickém plánování.<sup>45</sup> Jeho vize jde za možnosti modulů od OCLC.

Nicholson navrhuje vytěžování z dat použít k zjišťování chování uživatelů a jejich preferencí, k doporučení literatury pomocí analýzy nákupního košíku,<sup>46</sup> k prozkoumávání chyb v systému pro vyladění služeb, k nalézání vzorů chování při ztrátě drahých či vzácných dokumentů pro úpravu výpůjčního řádu, předpovědi trendů v chování uživatelů (jako je preference elektronických dokumentů), zjišťování nedostatků ve fondu (za primární zdroj informací v tomto směru označuje referenční služby), k vylepšování webových stránek, nebo k tvorbě expertního systému pro zodpovídání dotazů přes virtuální referenční službu.<sup>47</sup>

Z organizačního hlediska považuje Nicholson DM za mocný nástroj k evaluaci práce v knihovně, k optimalizaci toku zaměstnanců, k nalézání vzorů mezi zakoupenými produkty, které se například ukázaly jako špatná investice. Podobě vyhodnocení málo půjčovaných dokumentů může ukazovat na to, že byly třeba špatně zařazeny či popsány při katalogizaci. A povzbuzuje k vytváření profilů prodejců pro lepší akvizici. Vedení může použít výsledky DM k ospravedlnění svých výdajů a jako základ pro argumenty vzhledem ke zřizovateli. Knihovna může také získat lepší přehled o skupinách svých uživatelů a případně určit jejich oblasti zájmu. Podporou jejich specifických potřeb může rozvíjet znalostní management organizace. Vytěžováním dat o vyhledávání může získat přehled, jak jsou na tom uživatelé například s informační gramotností, a vedení knihovny snáze obhájí speciální kurzy či materiály. Na základě DM tak mohou organizace upravit výpůjční řád, přístup k nákupu knih, skupinám uživatelů a sbírkám dokumentů.<sup>48</sup> Podrobný metodický popis k procesu DM v knihovnách, který by vedl k lepšímu rozhodovacímu procesu, popisuje Nicholson v dalším ze svých článků.<sup>49</sup>

Tématem evaluace služeb knihovny z hlediska podpory na důkazech založeného rozhodování se Nicholson zabývá v článku „A conceptual framework for the holistic measurement and cumulative evaluation of library services“. K tomuto

---

<sup>42</sup>OCLC. WorldShare Report Designer. *OCLC.org*, 2016.

<sup>43</sup>Tj. data-driven decisions, ustálený český termín zatím neexistuje

<sup>44</sup>Nicholson, Scott, 2003b.

<sup>45</sup>NICHOLSON, Scott, STANTON, Jeffrey M. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. 2003, s. 247–262.

<sup>46</sup>Tím se myslí zjišťování, které zboží zákazníci kupují nejčastěji spolu. V našem případě to může znamenat, které dokumenty si uživatelé půjčují dohromady.

<sup>47</sup>Nicholson, Scott, Jeffrey M. Stanton, 2003.

<sup>48</sup>Nicholson, Scott, Jeffrey M. Stanton, 2003.

<sup>49</sup>NICHOLSON, Scott. The bibliomining process: Data warehousing and data mining for library decision making. *Information technology and libraries*. 2003, roč. 22, č. 4, s. 146. ISSN 2163-5226.

účelu zhotovil proces, rozdělený do dvou částí. První část se věnuje knihovnímu systému, druhá jeho využití. První se dělí na dvě fáze: Dotazování knihovníků a průzkum fondu má prověřit standardy a procesy v knihovně. Dále dotazník, rozhovor, focus group<sup>50</sup> a think-aloud<sup>51</sup> mají zjistit, jak podle uživatelů systém funguje. Druhá část se věnuje využití dokumentů, zde DM, analýza transakčních a webových logů a pozorování mají specifikovat zaznamenané interakce uživatele se systémem a případně s dokumentem. Poslední fází je pak zjišťování, jestli uživatelé získaných informací využili a to sledováním jejich citační aktivity pomocí průzkumů, rozhovorů, focus group a sledování citací.<sup>52</sup>

Nicholsonovo závěrečné schéma vychází z toho, že způsobů hodnocení služeb knihovny může být velké množství. Například, analýza nákladů a přínosů, analýza efektivnosti nákladů, analýza nákladů a užítku;<sup>53</sup> dále úplně jiné hodnocení přidělují knihovně a jejím službám uživatelé. Nicholson se nesnaží počet metod omezit, spíše vypracovat systém, který by hodnotil knihovnu z celistvého pohledu a využíval co nejvíce postupů.<sup>54</sup> Tento systém je samozřejmě výhodný, protože ukazuje celistvý obraz organizace a jejích služeb, může být však málo přehledný, zvláště, pokud porovnává více subjektů.

Jiný přístup k evaluaci pak nabízí Papavlasopoulos a Poulos, kteří se snaží o zjednodušení vyjádření hodnocení knihovny (především kvůli srovnání); zavedením jedné hodnoty, která by v sobě obsáhla několik hledisek. Při určování těchto hledisek vychází ze systému komplexního řízení kvality, používaném v řeckých akademických knihovnách. Stanovují tak devět hlavních kategorií (např. využívání elektronických služeb knihovny) a v každé několik proměnných (celkový počet je 43). Každá z nich je pak porovnána dvakrát, z hlediska účinnosti a z hlediska výkonnosti. Váha je u každého zhodnocení nastavena dle názoru experta či expertů a celý model je zachycen v neuronové síti. Neuronová síť je učena učitelem tak, aby zachycovala důležitost jednotlivých kategorií či hledisek, dle kterých je

---

<sup>50</sup>Focus group je metoda skupinového rozhovoru, při kterém respondenti reagují na moderátora, jenž jim pokládá otázky, a vzájemně na své odpovědi. (Guidelines for Conducting a Focus Group, 2016)

<sup>51</sup>Think-aloud je metoda používaná často UX designéry, při testování uživatelských rozhraní. Posadí respondenta k PC s daným prostředím, dají mu úkol a pak jej sledují a nutí, aby popisoval své myšlenky a pocity při jeho plnění. (NIELSEN, Jakob. Thinking Aloud: The #1 Usability Tool. *Nielsen Norman Group* [online]. January 16, 2012, [cit. 2016-02-09]. Dostupný z: ([https://assessment.trinity.duke.edu/documents/How\\_to\\_Conduct\\_a\\_Focus\\_Group.pdf](https://assessment.trinity.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf)))

<sup>52</sup>NICHOLSON, Scott. A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*. 2004, roč. 60, č. 2, s. 164–182. ISSN 0022-0418.

<sup>53</sup>Definice a popis jednotlivých metod, viz literatura. (SOUKOPOVÁ, Jana. Nákladově-výstupové metody hodnocení (CMA, CEA, CUA). In: *Veřejné zakázky a veřejné projekty a jejich hodnocení* [PDF]. Masarykova univerzita, 2013 [cit. 9. 2. 2016]. Dostupný také z: ([https://is.muni.cz/e1/1456/jaro2013/MPV\\_VZVP/um/33148301/Studijni\\_text\\_nakladove\\_vystupove\\_metody\\_CMA\\_CEA\\_CUA.pdf](https://is.muni.cz/e1/1456/jaro2013/MPV_VZVP/um/33148301/Studijni_text_nakladove_vystupove_metody_CMA_CEA_CUA.pdf)))

<sup>54</sup>Nicholson, Scott, 2004.



knihovna posuzována. Výsledkem je tedy model, do kterého je možné dosadit data konkrétních knihoven; výsledky je možné mezi sebou porovnat bez ohledu na to, jak je knihovna velká, proslulá, financovaná atd. Systém vrátí vždy objektivní vyjádření její úspěšnosti v jedné hodnotě.<sup>55</sup>

Je jisté, že některé články z této sekce bychom mohli přeradit například do sekce „Služby a data mining“, nicméně byly zařazeny sem především pro jejich globální povahu. Navrhované systémy, modely a aplikované postupy v této části krom jiného hodnotí služby knihoven, avšak hodnotí je z celkového pohledu na organizaci.

### 2.1.4 Rozdělování financí

Data mining a rozdělování financí v knihovnách bychom mohli začlenit jako podtéma k předchozím systémům pro podporu rozhodování, nicméně, zatímco předchozí část byla zaměřena na systémy pro podporu rozhodování a způsoby vyhodnocování práce v knihovnách, následující se bude zabývat jiným problémem, stanovováním kritérií pro rozdělování financí mezi obory (myšleno obory studia na univerzitě). Waltersova studie k tomu používá data o jednotlivých součástech, uživatelích a fondu k vytvoření vhodného modelu pro sestavení rozpočtu. Protože se data z roku na rok mění, využívá regresi k předpovědi vhodného rozdělení. Jako kritéria ve studii slouží:<sup>56</sup>

- Celkový počet studentů daného oboru.
- Elektronické informační zdroje (dále také EIZ) k oboru (počítá se samozřejmě poměr i s jinými obory).
- Počet studentů bakalářských oborů.
- Průměrná cena za jednu oborovou publikaci.
- Přítomnost magisterského oboru.<sup>57</sup>

Walters dále umožňuje každé z hodnot přiřadit váhu dle institucionálního zaměření. Pro směrodatné rozdělení fondů je samozřejmě nutné mít data za posledních několik let a brát ohled na situaci akademické knihovny. Zajímavost jeho přístupu je v tom, že ke stanovení rozdělení využívá jedné jednoduché rovnice.<sup>58</sup>

---

<sup>55</sup>PAPAVLASOPOULOS, Sozon et al. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, roč. 33, č. 4/5, s. 261–271. ISSN 0143-5124.

<sup>56</sup>WALTERS, William H. A regression-based approach to library fund allocation. *Library Resources and Technical Services*. 2006, roč. 51, č. 4, s. 263–278. ISSN 0024-2527.

<sup>57</sup>Walters, William H. 2006.

<sup>58</sup>Walters, William H. 2006.

### 2.1.5 Služby a data mining

Využití DM v oblasti služeb je poměrně široké téma; část článků, které se zabývají fungováním služeb, byla již zmíněna v podkapitole „Systémy pro podporu rozhodování“, část bude zmíněna zde.

Možnosti použití DM v této oblasti jsou rozmanité. Spojení vytěžování z dat s bibliometrií, respektive s bibliometrickou analýzou, využila Nicole Will z knihovny Technické univerzity v Delftu, aby zjistila, zda je pro univerzitu výhodné založit nový časopis. Knihovna měla zmapovat okruh potenciálních autorů a publika, a tak zjistit, jestli se vydávání nového časopisu vyplatí. Do průzkumu musela zahrnout konkurenční žurnály, publikační praktiky členů odborných organizací a potenciální autory. Během průzkumu vyplynuly zajímavé souvislosti, které pak bylo možné prezentovat univerzitě jako zadavateli, a umožnit tak rozhodnutí založené na znalostech, které nebyly předem známé.<sup>59</sup>

Druhou oblastí, kde je DM využíván, je zlepšování služeb. V tomto smyslu použila Nadaleen Tempelman-Kluit spojení vytěžování z dat s tvorbou person, která je nástrojem user-centered design.<sup>60</sup> Analyzovala záznamy referenčních dotazů New York University Library; komunikaci s každým uživatelem vyjádřila na dvou osách, jako povahu uživatelské informační potřeby, a úroveň jeho motivace. Nad daty bylo provedeno shlukování, čímž byly objeveny čtyři skupiny uživatelů. Těm byly přiřazeny čtyři osoby, tzv. data-driven personas, reprezentující uživatele dané knihovny. Potenciál výsledků netkví v lepší orientaci knihovníků ve skupinách uživatel, ale zvláště ve zlepšování výukových kurzů, webových stránek a nástrojů pro federativní vyhledávání dle konceptů user-centered design.<sup>61</sup>

DM byl také použit pro vyhodnocování rozsáhlého sociologického průzkumu o službách knihoven Reinholda Deckera and Antonie Hermelbracht. Cílem bylo zjistit preference uživatelů ohledně služeb knihoven co nejkomplexnějším způsobem. Nejdříve byly návrhy a nápady ohledně nových služeb spolu se stávajícími službami zapracovány do průzkumu, konstruovaného pomocí „adaptive conjoint analysis“ a „choice-based conjoint analysis“.<sup>62</sup> Průzkum pokrývá názor asi 3 500 respondentů z akademických obcí tří univerzit (dvou v Německu a jedné v USA).

---

<sup>59</sup>WILL, Nicole. Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*. 2006, roč. 73, č. 8, s. 1045–1050. ISSN 0040-1625.

<sup>60</sup>„Na uživatele zaměřený design“ označuje takové postupy ve webovém designu, které zajišťují jeho použitelnost.

<sup>61</sup>TEMPELMAN-KLUIT, Nadaleen et al. Invoking the User from Data to Design. *College and Research Libraries*. 2014, roč. 75, č. 5, s. 616–640. Dostupný také z: (<http://dx.doi.org/10.5860/crl.75.5.616>). ISSN 0010-0870.

<sup>62</sup>„Conjoint analysis“ je metoda používaná v marketingu, která umožňuje zjistit, která kombinace hodnot je pro respondenta nejvhodnější, jaké jsou jeho preference v souboru produktů či služeb. Názor na daný soubor alternativ či produktů bývá často vyjádřen tím, jak je dotazovaný seřadí. „Adaptive conjoint analysis“ pak znamená, že konkrétní otázky jsou pro respondenta generovány dynamicky. U „choice-based conjoint analysis“ pak respondenti vybírají jednu konkrétní službu z daného souboru alternativ. (Decker, Reinhold, 2006, s. 559–560)

Protože výsledky zahrnovaly poměrně velké množství dat, a protože dotazník nebyl konstruován standardní cestou, bylo použití DM k vyhodnocení nasnadě. Výsledků je samozřejmě velké množství; jejich vyhodnocování stále probíhá. Avšak, použitím netradiční metody výzkumu se podařilo dosáhnout dvou věcí. Zaprvé, účast respondentů byla vysoká, zadruhé, počet vyplňujících, jenž nedošli až na konec dotazníku, byl nízký.<sup>63</sup>

Zajímavou aplikaci DM metod provedli Fontane a Finnell,<sup>64</sup> kteří analyzovali záznamy o referenčních dotazech (a později také připojené záznamy sylabů vyučovaných předmětů) dvou univerzitních knihoven, aby zjistili, jak fond pokrývá témata, o která se čtenáři zajímají (nebo témata, jenž se na univerzitě vyučují). Snažili se také zjistit, jaké jsou vzorce využívání fondu, zda uživatelé, kteří se ptají, také doporučené zdroje využívají.<sup>65</sup>

Během průzkumu autoři objevili mnohé závislosti mezi vyučovanými kurzy a referenčními dotazy, které přicházely od studentů. Knihovníci tak mohli na základě prokázaných závislostí obhájit informační lekce, které měly studentům usnadnit práci během semestru. Po dobu průzkumu se také povedlo zavést novou službu „Book-A-Librarian“, vycházející z požadavků uživatelů, mající jim usnadnit získávání potřebných informací. Pro účely informačních lekcí také vznikla řada referenčních materiálů. Díky zjištěním byla upravena akviziční politika, aby více vycházela z potřeb vyučujících i studentů.<sup>66</sup>

Jody Condit Fagan pojednává o prověřování závislostí mezi využíváním referenčních služeb, účastí na školeních, vyhledáváním v databázích a nákupem dokumentů s požadavky na stažení plnotextových článků. Obecně existuje předpoklad, že čím více uživatelé využívají referenčních služeb a školení v oblasti databází, tím větší by mělo být využití zdrojů knihovny. K vyhodnocení byla použita data National Center for Education Statistics a Association of Colledge and Research Libraries. Výsledky potvrdily vazbu proměnné „požadavky na stažení plnotextových článků“ a proměnné „vyhledávání v odborné databázi“. Dále nebyla nalezena přímá vazba mezi využíváním referenčních služeb a požadavky na stažení plnotextových článků, nicméně nalezené závislosti napovídaly, že je ve hře ještě jiná proměnná, která nebyla do průzkumu zahrnuta.<sup>67</sup>

I podobné výsledky mají samozřejmě ve vědě svou hodnotu. V tomto případě by bylo vhodné shromáždit více dat, a provést nejdříve explorační analýzu, která

---

<sup>63</sup>DECKER, Reinhold et al. Planning and evaluation of new academic library services by means of web-based conjoint analysis. *The Journal of academic librarianship*. 2006, roč. 32, č. 6, s. 558–572. ISSN 0099-1333.

<sup>64</sup>FINNELL, Joshua et al. Reference question data mining: A systematic approach to library outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.

<sup>65</sup>Finnell, Joshua, 2010.

<sup>66</sup>Finnell, Joshua, 2010.

<sup>67</sup>CONDIT FAGAN, Jody. The effects of reference, instruction, database searches, and ongoing expenditures on full-text article requests: an exploratory analysis. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 3–4, s. 264–274. ISSN 0099-1333.

by umožnila blíže popsat situaci a nalézt správné proměnné. Teprve poté by bylo vhodné ověřovat závislosti pomocí regrese. Příkladem takové explorační studie je článek Eve-Marie Lacroix a Marie Elizabeth Collins.<sup>68</sup>

Tato studie se snažila popsat vývoj požadavků na meziknihovní výpůjční službu v oblasti lékařských knihoven v USA. Porovnávala rok 2005 a 1992. Jako zdroj dat sloužily záznamy z databáze Národní lékařské knihovny USA DOCLINE.<sup>69,70</sup>

Přestože členů používajících DOCLINE přibýlo, množství vyřízených požadavků na knihovnu ubylo. Zároveň lze vyzorovat změnu v požadavcích; v roce 1992 zaměřovaly ve velké většině na nové dokumenty, v roce 2005 už ale byla polovina vyžádaných dokumentů starších pěti let. Autorky si to vysvětlují rozmachem přístupů k odborným databázím přes konsorcia. V nich jsou dostupné hlavně nové články, takže není třeba je objednávat přes MVS. Naopak, ke starším dokumentům je problematictější se dostat, proto počet požadavků na MVS klesá (v celkových číslech), avšak zároveň stoupá stáří požadovaných dokumentů.<sup>71</sup>

DM nad daty o službách a dokumentech může vést k různým permutacím proměnných při snaze nalézt v nich závislosti. Otázkou spíše zůstává, jakou metodu použít a kdy.

### 2.1.6 Evaluace OPACu a webu knihovny

Transakční analýzy OPACu (podobně webové analýzy) mohou sloužit analýze uživatelského chování v elektronickém prostředí katalogu, k evaluaci jeho funkcí a zlepšování designu. Tento přístup je nasnadě hlavně proto, že pohyb uživatele v katalogu je vždy automaticky zaznamenáván knihovním softwarem. Nevyžaduje poskytnutí dat třetích stran jako v případě přístupů do elektronických informačních zdrojů.

Studie skupiny autorů pod vedením Debory Blecic analyzovala transakční logy OPACu knihovny University of Illinois. Sledovala hlavně přístup uživatelů k vyhledávání a zjistila, že mnoho uživatelů nerozumí terminologii; zmateně si proklikává cestu katalogem, a běžně se setkává s chybovými hláškami. Jako reakce byl zjednodušen design, jazyk vyhledávacího rozhraní a úvodních obrazovek. Dále byly zpřeházeny vyhledávací možnosti. Pro ověření změn byla provedena druhá analýza na datech s provedenými změnami. Výsledek byl okamžitý. Z transakčních logů vyplývalo, že průchod online katalogem je daleko plynulejší.<sup>72</sup>

<sup>68</sup>LACROIX, Eve-Marie et al. Interlibrary loan in US and Canadian health sciences libraries 2005: update on journal article use. *Journal of the Medical Library Association*. 2007, roč. 95, č. 2, s. 189–194. ISSN 1536-5050.

<sup>69</sup>Lacroix, Eve-Marie, 2007.

<sup>70</sup>Tento systém používají lékařské a nemocniční knihovny v USA a v Kanadě. Jako typ dokumentu jsou nejvíce požadovány odborné články. (Lacroix, Eve-Marie, 2007)

<sup>71</sup>Lacroix, Eve-Marie, 2007.

<sup>72</sup>BLECIC, Deborah D. et al. Using transaction log analysis to improve OPAC retrieval results. *College and Research Libraries*. 1998, roč. 59, č. 1, s. 39–50. ISSN 0010-0870.

Podobnou analýzu, ale na datech o přístupu uživatelů k elektronickým zdrojům, provedl Paul Bracke z Lékařské knihovny Arizonské univerzity. Jeho hlavním cílem bylo zjistit, zda má typ dokumentu (ekniha, ežurnál) nějaký vliv na způsob přístupu (vyhledávač, webová stránka), a jaký má vliv na návštěvnost webových stránek. Pomocí logistické regrese ověřil, že uživatelé nejčastěji ke zdrojům přistupují z kampusu a ze stránek knihovny, nikoli z katalogu. Dále, že velké množství přístupů na stránky se odehrává při vyhledávání dokumentu. Protože však existovalo jisté procento uživatelů mimo univerzitu, kteří se snažili přistoupit ke zdrojům zvenku (ačkoli toto spojení není podporováno), navrhl novou službu zpřístupňování dokumentů za poplatek mimo univerzitu. Dále navrhoval vycházet z výsledků analýzy při redesignu stránek knihovny. Navrhoval přestavění stránek tak, aby se uživatelé rychleji dostali k obsahu, který hledají. Dále změnu umístění některých málo užívaných služeb a hlavně vyhledávacího boxu.<sup>73</sup>

Jiann-Cherng Shieh využil analýzu webových logů k rekonstrukci webových stránek knihovny, tak aby bylo možné zlepšit jejich vnitřní strukturu. Ve svém článku k tomuto účelu použil analýzu rozptylu.<sup>74</sup>

Jednu z mála českých aplikací DM v knihovním prostředí reprezentuje bakalářská práce Michaly Soškové, která zpracovávala pomocí nástrojů SQL databáze a programu Excel údaje ze souborného online katalogu Masarykovy univerzity.<sup>75</sup> Měla k dispozici data ze čtyř měsíců fungování online katalogu v roce 2005 (červen až říjen), ve kterých se snažila zmapovat chování uživatelů. Výsledkem byly popsány vzorce chování, které se objevovaly nejčastěji, a také souhrn doporučení pro vývojáře a správce online katalogu, akviziční a katalogizační oddělení a vedení knihovny.<sup>76</sup>

Nevíme, proč se autorka rozhodla analyzovat data právě z těchto čtyř měsíců, které jsou pravděpodobně (kromě října) nejklidnějším časovým obdobím v katalogu knihovny, a tedy obdobím nejméně zajímavým. Studentka přesto dokázala najít a přesvědčivě specifikovat vzorce chování uživatelů v katalogu. Provedení analýzy umožnil také systém, který knihovna používá (Aleph). Pokud by systém data o průchodu čtenáře neukládal, nebylo by z čeho usuzovat.

---

<sup>73</sup>BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *J. Med. Libr. Assoc.* 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>).

<sup>74</sup>SHIEH, Jiann-Cherng. From website log to findability. *The Electronic Library*. 2012, roč. 30, č. 5, s. 707–720. ISSN 0264-0473.

<sup>75</sup>Šlo o náročný, ručně provedený proces bez použití specializovaného softwaru pro DM.

<sup>76</sup>SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného on-line katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)).

### 2.1.7 Průzkum chování uživatelů pomocí DM

Tato oblast reprezentuje poměrně různorodý blok. Nazvali jsme ji „Průzkum chování uživatelů pomocí DM“, protože data, ze kterých analýzy vychází, referují o uživatelském chování. Z průzkumů jako takových pro knihovnu většinou plynou další konsekvence, které mají svůj vlastní dopad: evaluace funkcí knihovny, nebo plánování akcí, změny v politice vedení knihovny atd. Pokud bychom toto rozdělení brali do důsledků, spadaly by sem jistě i některé případy zmíněné výše. Všechny studie na uvedená témata však měly výraznější vliv v jiné oblasti, proto nebyly zařazeny sem.

Zde zmíněné články se dělí přibližně na dva směry, na průzkum toho, jak knihovna ovlivňuje schopnost univerzity (či školy) udržet si studenty, a na analýzu využívání fondu knihovny. Publikace ve většině případů pocházejí od akademických knihovníků. Výjimku mezi nimi tvoří článek Douga Johnsona, který navrhuje využití DM ve školní knihovně na základní škole. Školní databáze zachycuje výkony studentů a databáze knihovny zase jejich aktivitu vzhledem k dokumentům a účasti na akcích. Navrhuje tato data proložit, aby výsledky zajistily lepší spolupráci školy s knihovnou a rodičů se školou.<sup>77</sup>

#### Průzkum vlivu knihovny na retenci studentů

Myšlenka průzkumu chování uživatelů, která by vedla k ověření teorie, že návštěvnost knihovny (akademické) ovlivňuje studijní výsledky k lepšímu, je pro knihovníky velmi lákavá. Pomohla by akademickým knihovnám lépe obhájit svou existenci. Jde o ideu natolik populární, že na toto téma bylo provedeno hned několik výzkumů.

Některé z nápadů, zmíněných Dougem Johnsonem, aplikuje studie kolektivu autorů v čele s Johnem Renaudem z University v Miami.<sup>78</sup> Informuje o průzkumu, jakým způsobem koreluje návštěvnost knihovny se studijními výsledky studentů. K měření byly využity informace z evidenčního systému školy, katalogu knihovny a data z turniketů. Ačkoliv se jasná korelace neprokázala (předpokládá se, že je to dáno dalšími vlivy, které nebyly brány do úvahy), byly v datech objeveny nečekané způsoby využívání knihovny. Autoři tak narazili na nový trend ke zmapování. Kromě nových výsledků a nepotvrzených korelací tým výzkumníků samozřejmě objevil v datech závislosti, které očekával, a podařilo se mu poměrně podrobně popsat, jaký typ uživatelů do knihovny chodí, a jaké materiály si půjčuje.<sup>79</sup>

---

<sup>77</sup>JOHNSON, Doug. A data mining primer and implications for school library media specialists. *Knowledge Quest*. 2004, roč. 32, č. 5, s. 32–35.

<sup>78</sup>RENAUD, John et al. Mining Library and University Data to Understand Library Use Patterns. *The Electronic Library*. 2015, roč. 33, č. 3. Dostupný také z: (<http://dx.doi.org/10.1108/EL-07-2013-0136>).

<sup>79</sup>Renaud, John et al, 2015.

Například Ethelene Whitmire ve svém článku analyzuje spojení mezi používáním knihovních zdrojů a služeb, a zlepšováním kritického myšlení studentů. Pro výzkum byla použita data shromážděná ze staršího průzkumu z několika univerzit. Zahrnutí byli studenti všech stupňů studia. Pomocí multidimenzionální regresní analýzy byly zpracovány jejich demografické informace, informace o používání knihovny a také respondenty posouzené změny v jejich kritickém myšlení. Ačkoliv nebyla potvrzena základní myšlenka, že studenti, kteří tráví v knihovně více času, nahlásili větší posun v kritickém myšlení, byla nalezena jiná závislost. Studenti s lepším kritickým myšlením sice tráví v knihovně méně času, ale více využívají zdrojů knihovny. Nabízí se tedy dvě vysvětlení, buď byli při lekcích v knihovně tak dobře instruováni, že potřebují k vyhledání zdrojů kratší dobu, nebo tato knihovna nenaplňuje svým fondem dostatečně jejich informační požadavky (v dané knihovně bylo zaznamenáno nejvíce žádostí o MVS), a proto v ní studenti tráví méně času.<sup>80</sup>

Kromě těchto skutečností bylo objeveno ještě několik dalších závislostí, které nebyly očekávány a vyžadovaly by hlubší prozkoumání.<sup>81</sup> Celý tento proces demonstruje dobře potenciál vytěžování z dat; dokud máme data, jsme v nich schopni nacházet závislosti, o nichž jsme neměli ponětí, a zpřesňovat naši představu o situaci (např. o tom, kdo jsou naši uživatelé, jak se chovají v knihovně, či jaké je jejich chování v oblasti informačních zdrojů).

Studie kolektivu autorů v čele s Kristou Soria z univerzitní knihovny Minnesotské univerzity se snažila prokázat vztah mezi využíváním univerzitní knihovny studenty prvních ročníků a jejich akademickými úspěchy, případně postoupením do dalšího ročníku. Povedlo se prokázat, že studenti, kteří využívají služeb knihovny a navštěvují ji, mají lepší známky; také pravděpodobnost postupu do druhého ročníku je vyšší. S úspěšnými studenty nebylo asociováno půjčování knih, ale přístup k elektronickým žurnálům a odborným databázím, využívání referenčních služeb a používání studijního místa s PC.<sup>82</sup> Je jasné, že po podobném závěru touží každá akademická knihovna. Důležité však je také správně interpretovat výsledky. Korelace těchto jevů je prokázána; zůstává otázkou, zda je návštěva knihovny či využívání jejích služeb důsledkem lepších studijních výsledků studentů, nebo obráceně, zda studenti, kteří mají potenciál na univerzitě zůstat, využívají knihovnu automaticky. Podobně by bylo třeba prozkoumat další faktory, které mohou podmiňovat lepší výsledky studentů.

Rozdílný přístup ke studijním výsledkům zaujal Mark Emmons a Frances Wilkinson na Univerzitě v Novém Mexiku. Zpracovali data od devadesáti devíti knihoven

---

<sup>80</sup>WHITMIRE, Ethelene. Academic library performance measures and undergraduates' library use and educational outcomes. *Library and information science research*. 2002, roč. 24, č. 2, s. 107–128. ISSN 0740-8188.

<sup>81</sup>Whitmire, Ethelene, 2002.

<sup>82</sup>SORIA, Krista M. et al. Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 1, s. 84–91. ISSN 0099-1333.

Association of Research Libraries a několika dalších, aby zjistili, zda existuje pozitivní korelace mezi počty absolventů a studentů, kteří postoupili do druhého ročníku, s počtem profesionálních knihovníků v knihovnách. Kromě hlavního cíle bylo nalezeno několik dalších závislostí (například v otázkách etnické příslušnosti a socioekonomických faktorů) a vyvstalo několik nových otázek, které by bylo třeba dále prozkoumat. V každém případě, pozitivní vztah mezi počty studentů, kteří se dostali do druhého ročníku, počty absolventů, a počtem odborných knihovníků se podařilo potvrdit.<sup>83</sup>

Je zajímavé, že reakce univerzit na úbytek finančních prostředků bývá většinou opačná, snížit počty knihovníků. Čím méně je studentů (ať už samovolně odpadajících, nebo málo nově se hlásících), tím více je knihovna nucena snižovat náklady na svůj provoz. V tomto případě článek Marka Emmonse přímo nahrává knihovnám v tom, aby se snažily snižování rozpočtu zabránit. Je pravda, že Emmons sám neumí korelace v datech spolehlivě vysvětlit, na to by byl třeba specializovaný výzkum; avšak, výsledky nelžou; chce-li univerzita získat nebo si zachovat větší počet studentů, měla by zvážit jeho závěry.

Autorem poslední studie v této podkapitole je Sharon Weiner. Zabývala se otázkou, zda je knihovna spolutvůrcem reputace univerzity, zaměřovala se na školy v USA, které mají doktorské studium. K ověření použila data sesbíraná ze čtyř různých zdrojů, celkově pokrývajících 247 univerzit. Údaje o knihovnách porovnávala s „Peer Assessment Score“ z „U.S. News and World Report“. Knihovníky asi nepřekvapí její závěr, kdy se jí pomocí regresní analýzy podařilo ověřit, že knihovna je spolutvůrcem reputace univerzity. Jak však dále sama uvádí, oproti jiným součástem bohužel silně podfinancovaným.<sup>84</sup>

### Analýza využívání fondu

Proč se tato část nejmenuje Analýza půjčovnosti fondu? Odpověď je jednoduchá, knihovny nedisponují jen tištěnými dokumenty, ale umožňují studentům přístup k odborným databázím, nebo půjčují elektronické knihy. Část níže zmíněných článků provedla analýzy právě na těchto datech, protože elektronický přístup samozřejmě generuje větší množství dat. Oblasti přístupů pouze do elektronických databází se budeme částečně věnovat v oddíle o digitálních knihovnách, zde zmíníme přehledovou studii Manoramy Tripathi, zabývající se možnostmi analýzy transakčních logů.<sup>85</sup>

---

<sup>83</sup>EMMONS, Mark et al. The academic library impact on student persistence. *College and Research Libraries*. 2011, roč. 72, č. 2, s. 128–148. ISSN 0010-0870.

<sup>84</sup>WEINER, Sharon. The Contribution of the Library to the Reputation of a University. *The Journal of Academic Librarianship*. 2009, roč. 35, č. 1, s. 3–13. ISSN 0099-1333.

<sup>85</sup>TRIPATHI, Manorama et al. A selective review of research on e-resource usage in academic libraries. *Library Review*. 2013, roč. 62, č. 3, s. 134–156. ISSN 0024-2535.



První z článků, jehož autory jsou Stephen Zweibel a Zachary Lane, analyzuje data z výpůjček fyzických knih a o elektronických interakcích. Data byla analyzována za účelem nalezení vzorců chování a ověření výpůjční politiky knihovny. Podařilo se nalézt tři hlavní trendy ve vypůjčování dokumentů. Zaprvé, na nejdější dobu si půjčují dokumenty vyučující. Zadruhé, celková aktivita výpůjček se snížila, což naznačuje přechod k elektronickému způsobu práce. A zatřetí, prodloužení výpůjční doby u studentů ze 30 dní na semestr mělo za následek pokles nejen prodlužování výpůjček (který se dal očekávat), ale také pokles výpůjční aktivity celkově. Studenti zřejmě došli k závěru, že čekat na vrácení určité knihy jeden semestr nemá smysl a opatřili si dokument jinak. Také se zvýšil počet elektronických interakcí.<sup>86,87</sup>

Druhá ze studií, provedených na toto téma, se týká fondu Marstonské vědecké knihovny Floridské univerzity, kterou má na svědomí kolektiv autorů v čele s Michelle Leonard. Ze začátku analyzovali data nejen o výpůjčkách, ale také o využívání e-knih a MVS. Rozdělili je dle typů uživatelů, skupin témat a konkrétních oddělení. Pomocí statistiky se jim pak podařilo odvodit výpůjční profily pro dané oddělení a danou kategorii čtenáře. Stanovili tak přibližné struktury, ve kterých se projevuje informační chování jedinců. Důsledky zjištění jsou validní, jak pro akviziční, tak pro katalogizační oddělení.<sup>88</sup>

O vytvoření výpůjčních profilů se pokusila také jiná studie Sumita Goswamiho, avšak ta svoje zkoumání postavila na komunitním přístupu k datům o půjčování dokumentů. Skupiny, které si půjčovaly stejné dokumenty, byly zařazeny do jedné komunity (komunit vzniklo několik). Na základě těchto údajů byly vytvořeny síťové grafy, které mohou dále knihovně sloužit například při doporučování knih čtenářům v online katalogu. Když se v sítích vytvořily shluky, které byly propojeny jedním bodem, bylo jasné, že se jedná o rozhraní jednotlivých skupin. Na základě sítěmi reprezentovaných výpůjčních profilů by mohly být uživatelům doporučovány publikace, díky definovaným rozhraním by jedné skupině nebyly doporučovány publikace od skupiny druhé.<sup>89</sup>

Nyní se budeme věnovat tématu sledování přístupu k elektronickým zdrojům a jaké druhy analýz to může podnítit. V tomto směru využijeme článek Manoramy Tripathi, které se analýzou literatury podařilo vymezit problematiku, a stanovit

---

<sup>86</sup>ZWEIBEL, Stephen et al. Probing the Effects of Policy Changes by Evaluating Circulation Activity Data at Columbia University Libraries: Edited by Rick J. Block. *The Serials Librarian*. 2012, roč. 65, č. 1, s. 17–27. ISSN 0361-526X.

<sup>87</sup>Článek bohužel nedostatečně vystihuje, co do těchto interakcí zahrnuje.

<sup>88</sup>LEONARD, Michelle F. et al. Metrics and science monograph collections at the Marston Science Library, University of Florida. *Issues in Science and Technology Librarianship*. 2010, č. 62, s. 1. Dostupný také z: (<http://dx.doi.org/10.5062/F4PC308T>). ISSN 1092-1206.

<sup>89</sup>GOSWAMI, Sumit et al. Visualisation of Relationships Among Library Users Based on Library Circulation Data. *DESIDOC Journal of Library & Information Technology*. 2010, roč. 30, č. 2, s. 26–39. ISSN 0974-0643.

celou řadu oblastí, kterými se knihovníci v průzkumech přístupu k elektronickým zdrojům zabývají:<sup>90</sup>

- Plánování podnikových zdrojů, systém pro informovanou akvizici na základě dat o využívání zdrojů.
- Srovnání mezi lokálními statistikami a statistikami dodavatele databáze.
- Analýza transakčních logů.
- Analýza využívání elektronických žurnálů.
- Srovnávání využitelnosti databází a elektronických žurnálů.
- Využívání e-zdrojů v různých institucích, oborech a různými kategoriemi čtenářů.
- Změny a trendy ve využívání e-zdrojů.
- Počty stažení a hodnocení článků.
- Relevance statistik dodavatele databáze.
- Role statistik v při výběru kolekcí žurnálů, optimalizaci rozpočtu, rozhodování, vzniku národních konsorcií a stanovování cen.
- Využití knihovnických databází.
- Informační chování a vyhledávání.
- Citace, využití zdrojů a stanovování měřítek.
- Odvozování nových technik k odhadování využití e-zdrojů.
- Kvalita a kvantita využití e-zdrojů.
- Vliv Open Access na publikování.
- Na využívání založená cenová politika e-žurnálů.
- Vliv na katalog.
- Federativní vyhledávání.
- Vliv na akademickou produktivitu.
- Metriky používání e-zdrojů.

---

<sup>90</sup>TRIPATHI, Manorama et al. A selective review of research on e-resource usage in academic libraries. *Library Review*. 2013, roč. 62, č. 3, s. 134–156. ISSN 0024-2535.

- Pokusy o standardizaci (SUSHI, ISO).
- Limity hodnocení přístupů.
- Role výukových a informačních lekcí na využívání e-zdrojů.<sup>91</sup>

Data mining byl zatím fyzicky využit jen v části zmíněných témat, ovšem, je logické, že dokud jsou přístupná všechna potřebná data, je DM možný v jakékoliv oblasti. Nyní se podíváme, v jakých případech bylo vytěžování z dat přístupů k elektronickým zdrojům zatím použito.

Podobným způsobem, jako se Whitmire<sup>92</sup> a Soria<sup>93</sup> snažily spojit využívání knihoven a studijní výsledky, se Sue Samson ve své práci pokusila spojit známky studentů s využitím elektronických informačních zdrojů. V tomto případě byla potvrzena hypotéza, že lepší známky (respektive lepší průměrné hodnocení studenta) jsou asociovány s používáním e-zdrojů.<sup>94</sup> Krom jiného se podařilo získat zajímavou zpětnou vazbu, totiž, že studenti i vyučující využívají e-zdrojů ve velké míře. Největší počet přístupů byl z domova. Demografická specifika jednotlivých skupin (např. ženy/muži a počty přístupů) nebyla statisticky významná, avšak z výsledků za jednotlivé součásti bylo možné získat přesnější informaci o tom, které fakulty či ústavy používají elektronické informační zdroje více, a které méně. V průběhu se vynořily další otázky, které by bylo dobré dalším průzkumem prověřit. Například vyučující dle výsledků využívají e-zdrojů nejvíce ze všech – zajímavé by bylo zjistit, jaký vliv to má na jejich vědecké publikování.<sup>95</sup>

Pervaiz Ahmad se snažil využít analýzy transakčních logů databáze elektronické knihovny (speciální rozhraní pro přístup k eknihám Edith Cowan University Library) k tomu, aby dokázal identifikovat v systému „zkušeného uživatele“. To je uživatel, který: „se odpoutal od tradiční formy učení, demografických charakteristik nebo technických bariér a konstantně užívá, sdílí, vytváří, publikuje nebo mění informace kreativním, inovativním nebo nezáměrným způsobem, tak, že se stává jejich násobitelem ve svém prostředí.“<sup>96</sup> Ačkoli je obecná definice známá, není jasné,

---

<sup>91</sup>Tripathi, Manorama, 2013.

<sup>92</sup>WHITMIRE, Ethelene. Academic library performance measures and undergraduates' library use and educational outcomes. *Library and information science research*. 2002, roč. 24, č. 2, s. 107–128. ISSN 0740-8188.

<sup>93</sup>SORIA, Krista M. et al. Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 1, s. 84–91. ISSN 0099-1333.

<sup>94</sup>Znovu vyvstává otázka, zda úspěšnější studenti využívají elektronických informačních zdrojů a nebo jestli studenti, kteří využívají e-zdrojů jsou ve studiu úspěšnější. V každém případě knihovna může tímto výsledkem zaštitit své výdaje v oblasti přístupů k odborným databázím.

<sup>95</sup>SAMSON, Sue. Usage of e-resources: Virtual value of demographics. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 6, s. 620–625. ISSN 0099-1333.

<sup>96</sup>AHMAD, Pervaiz et al. The e-book power user in academic and research libraries: Deep log analysis and user customisation. *Australian Academic & Research Libraries*. 2014, roč. 45, č. 1, s. 35–47. ISSN 0004-8623., s. 36.

jakým způsobem je chování těchto uživatelů zaznamenáno v systému knihovny, proto se Ahmad pokusil najít stopy jejich informačního chování a oddělit je od zbytku uživatelů.<sup>97</sup>

Zjistil, že se „zkušený uživatel“ vyznačuje častějším připojováním i vyhledáváním. Tráví více minut čtením a prochází více druhů titulů než jiní čtenáři. Tyto hodnoty spolu korelují; rozdíly mezi běžnými a „zkušenými uživateli“ jsou významné. Zároveň se ve studii pokusil zhotovit model, který by s určitou pravděpodobností byl schopen na jiných datech předpovědět, zda se jedná o tento druh čtenáře, či nikoliv. Díky této klasifikaci by byla knihovna schopna nabídnout „zkušeným uživatelům“ zvýšenou funkcionalitu nebo přizpůsobení jejich profilu.<sup>98</sup>

### DM analýza k organizaci fondu

Častým problémem ve veřejných knihovnách bývá „navigace“ čtenářů ve fondu. Pro rychlou orientaci v odborných dokumentech se používá např. Mezinárodní desetinné třídění (jako systematický selekční jazyk), a pro orientaci v beletrii základní kategorie vyjadřující např. literární formu (detektivka, životopis aj). Nicméně, jakkoliv se může zdát kategorizace beletrie knihovníkům vyhovující, je otázkou, nakolik vhodnou ji shledávají uživatelé. A pokud vhodná není, jak se to dá zjistit? I když se zeptáme dostatečného množství uživatelů, aby odpovědi reprezentovaly odpovídající statistický vzorek, jak prokážeme, že jiný systém byl vhodnější, když jediné, s čím mají uživatelé zkušenost, je systém současný?

Co reprezentuje poměrně složitý problém pro sociologický průzkum, nepředstavuje stejně obtížný problém pro DM. O co jednodušší je analyzovat data knihovny o využívání fondu všemi uživateli a zkusit kategorizaci fondu odvodit z dat, která knihovna už má? O něco podobného se pokusil americko-nizozemský tým výzkumníků vedený Jaapem Boterem. Shromáždili transakčních logy z pěti různých knihoven. Zkombinovali je s demografickými údaji a vyhledali vzorce chování uživatelů v půjčování dokumentů. Zjistili například, že uživatelé ve výpůjčkách dělají rozdíl mezi romány přeloženými (nejčastěji z Angličtiny) a původními nizozemskými.<sup>99</sup>

Oba druhy dokumentů vycházejí z různého pozadí a hodnotových žebříčků. Pro uživatele to znamená takový rozdíl, že je možné skupiny dle ve vypůjčování dokumentů jasně oddělit. Podobných zjištění bylo více. Výzkumníci proto navrhli podrobnější rozčlenění konkrétních děl do jednotlivých kategorií (např. upřesnili thrillery na akční thrillery, thrillery z lékařského prostředí, thrillery pro ženy a thrillery pro muže) a podkategorií, podle toho, jak podrobně bylo nutné knihy dělit.

---

<sup>97</sup> Ahmad, Pervaiz, 2014.

<sup>98</sup> Ahmad, Pervaiz, 2014.

<sup>99</sup> BOTER, Jaap et al. User categorization of public library collections. *Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. Dostupný také z: (<http://dx.doi.org/10.1016/j.lisr.2005.01.004>). ISSN 0740-8188.

Autoři provedli kategorizaci pětkrát, pro každou knihovnu zvlášť, a zdůrazňují, že ve všech by měl být aplikován jim příslušný model, který bude odpovídat konkrétnímu složení uživatelů. Nedoporučovali příliš výsledky generalizovat.<sup>100</sup>

### 2.1.8 Analýzy dat ze sociálních sítí

Zajímavý koncept byl publikován v článku „Who is following us“, který se zabývá analýzou dat uživatelů sociální sítě Twitter, sledujících profil jedné z poboček univerzitní knihovny Texaské Univerzity A&M.<sup>101</sup> Robin Sewell v něm zpracoval hlavně statistickou analýzu. Ovšem dnešní nástroje umožňují data ze sociálních sítí stahovat, provádět segmentaci a vizualizovat,<sup>102</sup> prostor pro datovou analýzu je zde velký. „Social Network Analysis“ je rozvíjejícím se odvětvím, těšícím se zájmu sociologů.<sup>103</sup> Vytěžování z tohoto druhu dat je teoreticky velmi zajímavá oblast. Ačkoliv se v tomto článku nejedná o aplikaci DM, myslíme si, že reprezentuje zajímavé téma, na které bychom rádi v praktické části navázali několika poznámkami. Proto jsme se jej rozhodli uvést.

## 2.2 Použití data miningu v digitálních knihovnách

Zajímavé je, že starší články viděly aplikaci DM do digitálních knihoven jako omezenou. Novodobější články vidí aplikaci DM do digitálních knihoven jinak. Těžiště této práce však neleží v pojednávání o digitálních knihovnách, ale v pojednání o knihovnách fyzických, proto zde uvedeme pouze modelové studie. Jelikož jsou články k tématu poměrně různorodé, rozdělíme je pouze do tří kategorií, obsahujících práce teoretické, praktické a ty s použitím text miningu. Teoretické práce se zabývají návrhy algoritmů, aplikací, či konceptů. Studii z Taiwanu uvádíme jako zástupce praktických užití DM do digitálních knihoven.

### 2.2.1 Teoretické práce

Jedním z předních autorů v této oblasti je Scott Nicholson. Ve svém článku „A framework for Internet Archeology...“ převádí metodu z archeologie do tzv. „internetové archeologie“. Vychází z toho, že u digitálních knihoven je spíše možné najít stopy interakce uživatele se systémem. Stanovuje tři měřitelné (na datech založené) informace: „informace a služby, jež systém nabídl (interní systémová

---

<sup>100</sup>Boter, Jaap, 2005.

<sup>101</sup>SEWELL, Robin R. Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*. 2013, roč. 31, č. 1, s. 160–170. Dostupný také z: (<http://dx.doi.org/10.1108/07378831311303994>). ISSN 0737-8831.

<sup>102</sup>Například software Mathematica. (New in Mathematica 9: Social Network Analysis. In: *Wolfram Mathematica* [online]. Wolfram, c2015, [cit. 2015-06-18]. Dostupný z: (<http://www.wolfram.com/mathematica/new-in-9/social-network-analysis/>))

<sup>103</sup>Vydělilo se z teorie grafů.

informace)“, „informace o tom, které záznamy byly relevantní“ a „informace o tom, které byly pertinentní (obě externí informace)“. K získání zpětné vazby navrhuje zaslat uživateli dotazník, který pomůže odvodit efektivnost, nebo zjistit typického uživatele a sledovat jeho jednání v systému.<sup>104</sup>

Evaluaci služeb digitálních knihoven pak navrhuje rozdělit do pěti částí:<sup>105</sup> sběr digitálních artefaktů o knihovních službách, indukci (neboli odvození modelu pomocí DM), dedukci, (logickou analýzu s generalizací), predikci (vytvoření hypotéz z generalizovaných skutečností) a testování hypotéz výzkumnou metodou (pomocí dat a uživatelů).<sup>106</sup>

Tento model „nejdříve sbírejme, hledejme závislosti, generalizujeme, tvořme hypotézy, pak ověřujeme“, je určitě použitelný mnohými způsoby a nejen v digitálních knihovnách, ovšem jiná otázka přichází s tím, zda se některý z knihovníků do něčeho podobného pustí. Jedním z problémů je samozřejmě náročnost takového postupu, a druhým je validita vzhledem k různým scénářům, které může knihovník zvolit.

Čistě technické řešení zpracování dat z digitální knihovny ukazují Chen a Chen ve svém článku „Using data mining technology to provide a recommendation service in the digital library“, kde navrhuje rozdělit uživatele do skupin shlukováním mravenčí kolonií<sup>107</sup> a následně doporučovat literaturu podle zařazení pomocí asociačních pravidel.<sup>108</sup> Jde o poměrně dobře popsany a promyšlený postup, který ambiciózně využívá (mezi knihovníky) ne příliš známých technik. Navržen je pro digitální akademickou knihovnu, ovšem jeho využití může být při dostatečném zásobení daty daleko širší.

### 2.2.2 Praktické práce

Studie Chan-Chine Changa a Ruey-Shun Chena z National Chiao Tung University na Taiwanu obsahuje analýzu záznamů o vypůjčkách v akademické digitální knihovně. Výzkumníci pracovali s údaji od roku 2000 do roku 2003, které obsahovaly kolem milionu řádků. Nad těmito záznamy postavili datový sklad, který byl poté využit ke shlukové analýze. Jednotlivé shluky pak popsaly typy uživatelů a jejich chování vzhledem k vypůjčovaným dokumentům. Podařilo se tak potvrdit

---

<sup>104</sup>NICHOLSON, Scott. A framework for Internet archeology: Discovering use patterns in digital library and Web-based information resources. *First Monday*. 2005, roč. 10, č. 2.

<sup>105</sup>Dle hypoteticko-deduktivně-induktivního cyklu, který používá moderní archeologie.

<sup>106</sup>Nicholson, Scott, 2005.

<sup>107</sup>Je to technika strojového učení bez učitele k určování a popisu skupin pomocí shlukování. (JAFAR, Mohamed et al. Ant-based clustering algorithms: A brief survey. *International Journal of Computer Theory and Engineering*. 2010, roč. 2, č. 5, s. 787–796. Dostupný také z: (<http://www.ijcte.org/papers/242-G730.pdf>). ISSN 1793-8201. )

<sup>108</sup>CHEN, Chia-Chen et al. Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*. 2007, roč. 25, č. 6, s. 711–724. Dostupný také z: (<http://dx.doi.org/10.1108/02640470710837137>). ISSN 0264-0473.

zvyšující se zájem uživatelů o digitální materiály spolu s upřesněním skupiny zájemců o multimediální dokumenty.<sup>109</sup> Velkou pozornost autoři věnují konstrukci datového skladu a popisu práce při analýzách. Vytváří tak možnost jejich postup opakovat.

### 2.2.3 Text mining

Z hlediska využitelnosti DM v knihovnictví je samozřejmě jedním z nejlákavějších konceptů automatické zpracování redukovaných textů. V určitých konkrétních případech k automatickému zpracování dokumentu došlo, ovšem použití má stále své limity. Tyto limity potvrzuje o něco starší článek Zdeňka Jonáka o nástroji pro automatickou obsahovou analýzu TextQuest. Tento je schopen vytvořit slovníky slovních výrazů z textu pro jednoduchý popis dokumentů, tj. jako podklad pro obsahovou analýzu vykonávanou člověkem.<sup>110</sup>

Automatickým zpracováním článků stažených z internetu do katalogizačního formátu před jejich zařazením do digitální knihovny, se zabývá článek „Bibliographic Data Mining“ kolektivu autorů v čele s Ying Li. Použili programovací jazyk PERL k automatickému získávání popisných údajů z webových stránek poskytovatelů elektronických žurnálů. Systém je schopen deskriptivní katalogizace, ovšem neobejde se bez pomoci lidské ruky, která musí systém správně nastavit, tj. označit, odkud má brát jaké údaje. Ačkoliv je nutné systému mnohé ukázat, jeho výhoda tkví v tom, že je schopen podobné záznamy ze stejného žurnálu vytvářet sám, čímž šetří čas knihovníků.<sup>111</sup>

## 2.3 Bibliomining

V souvislosti s DM používaným v knihovnách se objevil také specializovaný termín bibliomining. Jeho tvůrcem je jedna z nejvýraznějších osobností, která se kdy DM v knihovnách zabývala. Scott Nicholson, profesor na School of Information Studies při Syracuse University, je autorem mnoha metodických článků,<sup>112</sup> ve kterých tento pojem zavedl.<sup>113</sup>

---

<sup>109</sup>CHANG, Chan-Chine et al. Using data mining technology to solve classification problems: A case study of campus digital library. *The Electronic Library*. 2006, roč. 24, č. 3, s. 307–321. Dostupný také z: (<http://dx.doi.org/10.1108/02640470610671178>). ISSN 0264-0473.

<sup>110</sup>JONÁK, Zdeněk. TEXTQUEST: software pro obsahovou analýzu. *Ikaros*. 2000, roč. 4, č. 5. Dostupný také z: (<http://ikaros.cz/textquest-software-pro-obsahovou-analyzu>). ISSN 1212-5075.

<sup>111</sup>LI, Ying et al. Bibliographic data mining: automatically building component part records for e-journal articles on the Internet. *Journal of internet cataloging*. 2002, roč. 5, č. 1, s. 29–41. ISSN 1091-1367.

<sup>112</sup>NICHOLSON, Scott. *Bibliomining.com* [online]. [cit. 2015-05-20]. Dostupný z: (<http://www.bibliomining.com/>).

<sup>113</sup>NICHOLSON, Scott, STANTON, Jeffrey M. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries.

Definuje jej doslova jako: „Bibliomining, neboli data mining v knihovnách, je aplikace data miningu a bibliometrických nástrojů na data generovaná službami knihovny.“<sup>114</sup> Dle něj zahrnuje pojem vytváření datového skladu (s daty, které knihovna produkuje u své činnosti) a provádění DM analýz nad tímto skladem.<sup>115</sup> Termín byl odvozen od slov bibliometrie a data mining pro odlišení článků s tématy „data mining prováděný v knihovnách“ a „softwarové knihovny používané pro data mining“.

Autor se ve svých článcích zabývá vytvářením datového skladu v knihovně,<sup>116</sup> použitím metod logistické regrese, nelineární diskriminanční analýzy, rozhodovacích stromů a neuronových sítí pro bibliomining.<sup>117</sup> Také se zabývá využitím bibliominingu, např. pro akvizici, meziknihovní výpůjční službu, podporu rozhodování, knihovní služby atd.<sup>118</sup>

Ačkoli je Scott Nicholson považován za významného odborníka v této oblasti, jeho termín bibliomining se příliš neujal. Někteří z autorů byli termínem nadšeni, ale je nutné podotknout, že nejkonzistentněji jej používá sám Nicholson. Velká většina vědců, softwarových knihovníků a jiných odborníků používá ve svých článcích nadále spojení knihovna a data mining. Proto ani v této práci nebudeme termín bibliomining používat.

### 2.4 Data mining a knihovny v Česku a na Slovensku

Bohužel, data mining je v českých a slovenských knihovnách poměrně opomíjeným tématem. Existují pouze dvě práce, které přináší výsledky aplikace DM na knihovní data. Jednou je případová studie Pavlíny Omastové, Radky Římanové a Ctibora Škuty, kteří v termínech z vyhledávání Národní technické knihovny hledali nové vhodné termíny pro Polytématický strukturovaný heslář.<sup>119</sup> Je to poměrně originální způsob uplatnění DM v knihovnách, a také nesmírně záslužný, podíváme-li se na stav některých jiných tezaurů či řízených slovníků, kterým úplně chybí perspektiva uživatele.

---

*Organizational data mining: Leveraging enterprise data resources for optimal performance.* 2003, s. 247–262.

<sup>114</sup>NICHOLSON, Scott. The bibliomining process: Data warehousing and data mining for library decision making. *Information technology and libraries.* 2003, roč. 22, č. 4, s. 146. ISSN 2163-5226., s. 1.

<sup>115</sup>NICHOLSON, Scott. Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works. *Journal of the American Society for information science and technology.* 2003, roč. 54, č. 12. ISSN 1081-1090., s. 786.

<sup>116</sup>Nicholson, Scott, 2003b.

<sup>117</sup>Nicholson, Scott, 2003a.

<sup>118</sup>Nicholson, Scott, Jeffrey M. Stanton, 2003.

<sup>119</sup>OMASTOVÁ, Pavlína et al. Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety. In: *Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety.* 2011, s. 101–109. ISBN 978-80-86249-62-9.



Druhou aplikaci DM na data knihovny představuje již zmiňovaná práce Michaly Soškové, která analyzovala data ze Souborného online katalogu knihovny Masarykovy univerzity.<sup>120</sup>

Kromě praktických prací existují i dvě práce teoretické, snažící se vymezit pojmy, postupy a způsoby, ke kterým lze DM v knihovnách použít. Jednou z nich prací je vymezení Beáty Sedláčkové, která jako jedna z prvních referovala o této metodě, a jejím použití v českém jazyce.<sup>121</sup> Její článek poukazuje na možnosti využití DM při komplexním řízení kvality, které jsme jako praktické využití zmínili také v části Systémy pro podporu rozhodování.<sup>122</sup>

O něco kratší práci s vysvětlením pojmu bibliomining, a několika příklady publikovala na Slovensku také Katarína Lалуhoová.<sup>123</sup> Mimo těchto prací byl pojem bibliomining či data mining v knihovnách zmíněn také ve dvou prezentacích.<sup>124</sup>

Vzhledem k požadavkům na výzkumníka a k technické obtížnosti je DM náročným procesem. Čím méně finančních prostředků knihovna na práci s daty má, tím více se proces komplikuje; a tím více znalostí je také vyžadováno od toho, kdo data analyzuje. To je zřejmě také důvod, proč se DM v českých knihovnách masivněji nerozšířil.

---

<sup>120</sup>SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného on-line katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)).

<sup>121</sup>SEDLÁČKOVÁ, Beáta. Data mining a jeho uplatnění při poznávání knihovnických jevů a zákonitostí. *Knižnica*. 2007, roč. 8, č. 11–12, s. 21–22. Dostupný také z: ([http://www.snk.sk/images/snk/casopis\\_kniznica/2007/november-december/21.pdf](http://www.snk.sk/images/snk/casopis_kniznica/2007/november-december/21.pdf)). ISSN 1335-7026.

<sup>122</sup>PAPAVLASOPOULOS, Sozon et al. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, roč. 33, č. 4/5, s. 261–271. ISSN 0143-5124.

<sup>123</sup>LALUHOVÁ, Katarína. Bibliomining – data mining pre knižnice. *ITlib: Informačné technológie a knižnice*. 2007, roč. 4. Dostupný také z: (<http://www.cvtisr.sk/itlib/itlib074/laluhova.htm>).

<sup>124</sup>ŘÍMANOVÁ, Radka. Bibliomining – dobývání znalostí z knihovnických databází a systémů [online]. 3. 11. 2008, [cit. 2016-02-15]. Dostupný z: (<http://uisk.ff.cuni.cz/detail.do?articleId=6569>); PILECKÁ, Věra. Zpráva z konference Využívání informací v informačnej spoločnosti. *Ikaros*. 2006, roč. 10, č. 11. Dostupný také z: (<http://ikaros.cz/node/12255>). ISSN 1212-5075.

## 3 Analýza dat z knihovního softwaru Clavius

V předchozím textu jsme potvrdili, že vytěžování dat v knihovnách je možné využít. Také jsme ukázali, jakým způsobem je možné knihovnám pomoci. Pro referenci jsme použili akademické a veřejné knihovny, které jej aplikovaly.

Víme, že možnosti DM přímo závisí na datech, která má knihovna k dispozici. Z povahy dat lze usoudit, pro jaké analýzy jsou vhodná. Mějme však na paměti, že knihovní systém, ukládající data ve způsobilém formátu, je ještě nemusí zachycovat tak, aby provedení analýz bylo jednoduché, nebo aby výsledky byly průkazné. Data, produkovaná systémem (a jejich vhodnost k provádění datových analýz) mohou být pro knihovnu jedním z rozhodujících kritérií při nákupu nového systému.

Uvedli jsme dva příklady využití DM v českém prostředí; oba se týkaly akademických knihoven,<sup>1</sup> oba vycházely z dat systému Aleph.

Aleph je zavedeným knihovním softwarem od firmy Ex Libris, sídlící v Izraeli. Tento software je používán v sedmdesáti zemích světa. V ČR mezi uživatele patří převážně krajské, vědecké, akademické a jiné odborné knihovny.<sup>2</sup> Jedná se o systém všestranně podporující provoz knihovny. V záložce „Služby“ umožňuje vyhodnocování jednotlivých činností. Sám podrobnější analýzy nenabízí, ale z provedených studií víme, že jeho data jsou jeho data pro analýzu vhodná.

Například víme, že záznamy o průchodu online katalogem systému jsou tak podrobné, že analýzou jeho transakčních logů lze zjistit chování uživatelů při

---

<sup>1</sup>Národní technická knihovna je nejen odbornou technickou knihovnou, ale také knihovnou akademickou, protože slouží třem vysokým školám ve svém okolí. (O NTK. *Národní technická knihovna* [online]. NTK, c2006–2016, [cit. 2016-02-15]. Dostupný z: (<https://www.techlib.cz/cs/82794-o-ntk>))

<sup>2</sup>Knihovny. *SU Aleph* [online]. SU Aleph: Neformální sdružení uživatelů knihovního systému Aleph, [cit. 2016-02-15]. Dostupný z: (<http://www.sualeph.cz/clients/>).

procházení.<sup>3</sup> Některé jmenované zahraniční studie využily data z Alephu při datových analýzách.<sup>4,5</sup>

Jak jsou na tom však jiné knihovní softwary? Zajímá-li nás čistě české prostředí, prvním systémem, který nám přijde na mysl, je Clavius, protože se jedná o systém nejrozšířenější.

Je známo, že Clavius je určen převážně veřejným knihovnám městského typu. Skladba jeho modulů podporuje práci s dětmi (modul Dětský ONLINE katalog) a zpřístupňování internetu veřejnosti (modul Dispečink internetu), ale neumožňuje například spravovat elektronické informační zdroje,<sup>6</sup> které jsou jedním ze základních zdrojů informací studentů a učitelů na VŠ při jejich práci. Jeho určení také potvrzuje skladba zákazníků. Z přibližně tří tisíc knihoven tvoří většinu knihovny městské a obecní. Pouze vybrané odborné a krajské knihovny jej používají a zcela výjimečně pak Clavius využívají univerzitní a vysokoškolské knihovny – Slezská univerzita v Opavě, Metropolitní univerzita Praha, Veterinární a farmaceutická univerzita Brno, Univerzita Palackého Olomouc, Česká zemědělská univerzita, New York University in Prague, Západomoravská vysoká škola, Moravská vysoká škola, Vysoká škola uměleckoprůmyslová a Vysoká škola ekonomie a managementu.<sup>7</sup>

Výběr knihovního softwaru je samozřejmě otázkou priorit a dostupných fondů. Knihovna musí zvážit mnoho hledisek, než se rozhodne pro konkrétní systém. V každém případě, jedno z těchto hledisek by mělo znít: zda software nabízí nákup modulu pro vyhodnocování činností v knihovně, nebo zda jsou data alespoň vhodná pro stažení a vyhodnocení v jiném programu. Cílem praktické části práce tak bude ověření, jaké jsou možnosti použití DM na datech z Clavia. Využití bude ověřováno na příkladech z teoretické části.

Skladba modulů vyhovuje více městským knihovnám; souvisí s daty, která jsou o činnostech v systému ukládána. Clavius umožňuje pomocí statistik vyhodnocovat návštěvnost knihovny (ve dnech a hodinách), výpůjčky (ve dnech a hodinách) rozložení čtenářů (procentuálně mezi oddělení, kategorie, dle pohlaví a věku) a

---

<sup>3</sup>SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného on-line katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)).

<sup>4</sup>LEONARD, Michelle F. et al. Metrics and science monograph collections at the Marston Science Library, University of Florida. *Issues in Science and Technology Librarianship*. 2010, č. 62, s. 1. Dostupný také z: (<http://dx.doi.org/10.5062/F4PC308T>). ISSN 1092-1206.; SORIA, Krista M. et al. Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 1, s. 84–91. ISSN 0099-1333.

<sup>5</sup>Data mining byl na těchto datech proveden jak s použitím, tak bez použití specializovaného softwaru.

<sup>6</sup>Clavius – základní informace. *Lanius* [online]. Lanius, [cit. 2016-02-15]. Dostupný z: (<http://www.lanius.cz/>).

<sup>7</sup>Reference knihovního systému Clavius v ČR. *Lanius* [online]. Lanius, [cit. 2016-02-15]. Dostupný z: (<http://www.lanius.cz/>).

zastoupení různých druhů dokumentů ve půjčených dokumentech.<sup>8</sup> Ve statistikách ale chybí možnost propojení a hlubší analýzy. Systém například nedokáže kromě vyhodnocení aktivity čtenářů vyhodnotit také aktivitu knihovníků. Neumí oddělit vyhodnocení jednotlivých služeb od sebe a vzájemně je porovnat, o vizualizaci obsahu fondu nemluvě.

Clavius je mezi knihovnamí populární hlavně pro nízkou cenu a modularitu. Otázkou zůstává, zda právě nízká cena vyváží funkce, které Claviu chybí – jako možnost doplnění systému o manažerský modul, který by podrobně vyhodnocoval data o činnosti knihovny, podobně jako software Alma a produkty od OCLC. Podpora rozhodování je jednou z klíčových činností, kterou DM umožňuje. Proto je dalším z cílů praktické části této práce ověřit, zda i přes absenci manažerského modulu lze data ze softwaru stáhnout a vyhodnotit tak, aby bylo výsledky možné použít k manažerským rozhodnutím v knihovně.

Pro tento účel máme k dispozici data Univerzitní knihovny Slezské univerzity v Opavě, která Clavius používá od roku 2011. Clavius není systém, který by byl primárně určen akademickým knihovnám, přesto jej akademické knihovny používají. Pohledem do databázových tabulek systému můžeme zjistit, jak ukládání dat v systému plní potřeby konkrétní akademické knihovny z hlediska datové analýzy.

## 3.1 Cíle praktické části

Pro praktickou část jsme si tedy stanovili tyto cíle:

1. Cílem je zjistit, jaké jsou možnosti využití DM na datech ze softwaru Clavius. Nejdříve chceme ověřit, zda jsou data, ukládaná systémem, dostatečná pro provedení analýz. A pokud ano, jakého druhu? Jako možnosti použití dat budeme uvažovat oblasti, které jsme ověřili průzkumem literatury. Vhodnost dat pro DM samozřejmě můžeme posuzovat také z hlediska jednoduchosti práce s daty, náročnosti na vědomosti knihovníka (budeme specifikovat konkrétní oblasti) a průkaznosti výsledků analýz.<sup>9</sup> Tyto budou dále rozvedeny v konkrétních částech. V průběhu plnění tohoto cíle bude prověřeno,<sup>10</sup> zda lze data ze systému Clavius využít k DM v těchto oblastech:

- akvizice,

---

<sup>8</sup>Emailová korespondence s Evou Lackovou, 7. 3. 2016.

<sup>9</sup>V průběhu zpracovávání tohoto cíle nebudeme uvažovat, jaký program lze pro analýzy využít. Volbu necháme na knihovnicích, pokud budou schopni provést analýzy pomocí SQL (jako Michaela Sošková), bude to jejich volba. Způsob analýzy bude vyžadovat jejich invenci a bude záviset na řešeném problému. Pokud se budeme zmiňovat o jazyce SQL, bude to kvůli úpravě dat, nikoli kvůli analýze.

<sup>10</sup>Prověření bude sestávat hlavně ze zjištění, zda systém ukládá data vhodná pro datové analýzy.

- analýza záznamů v katalogu,
  - systémy pro podporu rozhodování,
  - rozdělování financí,
  - služby,
  - průzkum chování uživatelů,
  - analýzy dat ze sociálních sítí.
2. Cílem je provést DM na datech ze softwaru Clavius, které by bylo možné využít k podpoře manažerských rozhodnutí v knihovně.

Manažerské využívání datových analýz je asi nejširším okruhem, který se DM v knihovnách dotýká. Naším úkolem v této fázi bude nejdříve zjistit, k čemu může knihovna data z Clavia skutečně využít, a jaké okruhy řízení provozu knihovny může takto podpořit. Součástí tohoto cíle bude samozřejmě provedení datových analýz; již předem však víme, že se tato část bude dále specializovat (na analýzu chování uživatelů).

- Jaké okruhy problémů při řízení knihovny lze daty podpořit.
- Na základě předchozích zjištění provést data mining nad daty z Clavia.

## 3.2 Postup

Při plnění prvního cíle budeme vycházet z dat knihovního softwaru Clavius, poskytnutých Univerzitní knihovnou Slezské univerzity v Opavě, a z pramenů. Z rozboru literatury víme, že při použití DM ve fyzických knihovnách můžeme teoreticky data použít k různým účelům. Provedeme mapování systémem zaznamenaných skutečností, a následně zhodnocení. Z výsledků pak budeme vycházet.

Dále se soustředíme na podporu rozhodování. Zjistíme, jaká data systém ukládá, a potom k podpoře kterých manažerských rozhodnutí by je bylo možno použít. Následně se budeme ve finální datové analýze specializovat, a provedeme DM.

DM lze provést dvojím způsobem; jako otevřený DM s nestanovenými předpoklady, nebo s použitím inferenčních statistických metod (chceme-li zjistit více o konkrétním jevu). V této fázi by nám však použití těchto metod (například regresní analýza) nepomohlo. Pokud bychom předem stanovili proměnné, a hledali mezi nimi vztah bez toho; abychom věděli, jaké proměnné jsou ve hře, podobalo by se to hledání jehly v kupce sena. Výsledek by mohl vyjít falešně pozitivní, falešně negativní, nebo prostě neprůkazný. Z tohoto důvodu se na datech nejdříve provádí otevřený data mining.<sup>11</sup> DM bude v tomto bodě využívat vizualizačních technik, statistiky a histogramů.

<sup>11</sup>Příkladem, kde nebyly správně stanoveny proměnné, mohou být některé studie. (FINNELL, Joshua et al. Reference question data mining: A systematic approach to library outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286; CONDIT FAGAN, Jody. The effects

V rámci této práce jsme využili přímé exporty dat z databáze SQL. Načetli jsme je R Studiem, kde jsme data vyčistili, transformovali, analyzovali a ukládali do různých formátů.

### 3.3 Současná situace Univerzitní knihovny Slezské univerzity v Opavě

Popisovat budeme stav Univerzitní knihovny Slezské univerzity v Opavě (dále jen UKSU) v rámci období, sledovaného v datech, tedy od 1. 9. 2011 do 3. 8. 2015. Jinými slovy, nebudeme popisovat aktuální stav. Knihovna má dvě hlavní části, první se nachází v Opavě, druhá v Karvině. Obě mají několik pracovišť s profesionálními knihovnicemi. Na těchto pracovištích lze provádět výpůjčky (zaznamenané do systému) a provozovat odborné služby:

- Opava,
  - Studovna a půjčovna,
  - Výpůjční sklad Hauerova,
  - Knihovna matematického ústavu,
  - Rakouská knihovna,
  - Německá knihovna.
- Karviná,
  - Budova Na Vyhlídce,
  - Čítárna.<sup>12</sup>

Kromě výše zmíněných lokací je fond dále rozložen mezi ústavní knihovny, administrované představiteli ústavů (vybranými vyučujícími, doktorandy, sekretářkami atd.) a mezi pracovny vyučujících. V těchto knihovnách se zaznamenává výpůjčka pouze ručně. Vzhledem k absenci těchto výpůjček v systému je nutné počítat s jistým omezením (v datech).

---

of reference, instruction, database searches, and ongoing expenditures on full-text article requests: an exploratory analysis. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 3–4, s. 264–274. ISSN 0099-1333.; BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *J. Med. Libr. Assoc.* 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>; RENAUD, John et al. Mining Library and University Data to Understand Library Use Patterns. *The Electronic Library*. 2015, roč. 33, č. 3. Dostupný také z: <http://dx.doi.org/10.1108/EL-07-2013-0136>; EMMONS, Mark et al. The academic library impact on student persistence. *College and Research Libraries*. 2011, roč. 72, č. 2, s. 128–148. ISSN 0010-0870. )

<sup>12</sup>Univerzitní knihovna *Slezská univerzita v Opavě* [online]. Slezská univerzita v Opavě, c2011, [cit. 2016-02-15]. Dostupný z: <http://www.slu.cz/slu/cz/univerzitari-knihovna>.

Knihovně se po dlouhá léta nepodařilo umístění fondu Opavské části sjednotit. Částečně je to způsobeno umístěním v nevyhovujících prostorách. Změna by měla proběhnout v příštím akademickém roce, protože bude dokončena vestavba ve dvoře univerzitní budovy na Bezručově náměstí 14, kam by měla Univerzitní knihovna přesídlit. Jiná situace je v Karviné, kde se veškerý fond nachází v budově Na Vyhlídce. Čítárna v Karviné slouží jako prodejna skript a studijní prostor (výjimku tvoří regál s referenčními materiály k zapůjčení).

Standardně knihovna poskytuje absenční i prezenční výpůjční službu, meziknihovní výpůjční službu, přístup k elektronickým informačním zdrojům (i vzdáleně).<sup>13</sup> Na začátku semestru pořádá informační lekce, a v prostorách knihovny poskytuje rešeršní a reprografické služby.<sup>14</sup>

Z celkového počtu 16 knihovníků jich 11 vykonává svou profesi v Opavě a 5 v Karviné (včetně vedoucích a systémového knihovníka).<sup>15</sup> Uvážíme-li počet míst, kde jsou poskytovány profesionální služby a fakt, že ne všichni zaměstnanci v knihovně pracují na celý úvazek, dojdeme k názoru, že knihovna musí mít napjatý rozvrh služeb, aby pokryla otevírací dobu všech svých součástí a zajistila všechny další činnosti.

Bohužel, podobně jako v jiných akademických knihovnách, také UKSU musí bojovat o finanční podporu ze strany Slezské univerzity. Přístup jednotlivých představitelů hlavních součástí univerzity ke knihovně se od sebe velmi liší. Knihovna v podstatě již od vzniku univerzity bojovala o lepší podmínky pro provozování svých služeb. Ačkoliv se vyhlídky na důstojnější umístění pomalu naplňují, boj knihovny o lepší rozpočet a prostředky na pracovní sílu nekončí.

UKSU má oficiálně statut součástí Slezské univerzity, ale jakoukoliv změnu musí projednat a schválit nejen vedení univerzity, ale také vedení každé jednotlivé součásti. Není od věci zmínit, že ne všichni jsou v názorech vždy za jedno; knihovně by se hodilo mít k dispozici materiály vztažené ke každé součásti zvlášť (kterými by mohla podpořit své argumenty). Vyjednávání o financování ztěžuje nedostatek informací v různých oblastech.<sup>16</sup> Proto se jeden z našich cílů zaměřuje právě na podporu manažerských rozhodnutí.

Knihovna operuje od 1. 9. 2011 v systému Clavius. Před nástupem Clavia používala T Series. Uživatelům je ze školy i z domu přístupný online katalog Carmen. Do fondu knihovny jsou kromě fyzických knih nově (od léta 2014) zařazeny elektro-

---

<sup>13</sup>Pracoviště Opava: Elektronické informační zdroje *Slezská univerzita v Opavě* [online]. Slezská univerzita v Opavě, c2011, [cit. 2016-02-15]. Dostupný z: ([http://www.slu.cz/slu/cz/univerzitni-knihovna/pracoviste-opava/fondy/elektronicke\\_informacni\\_zdroje](http://www.slu.cz/slu/cz/univerzitni-knihovna/pracoviste-opava/fondy/elektronicke_informacni_zdroje)).

<sup>14</sup>Nyní nově poskytuje přístup ke digitalizovaným dokumentům ze svého fondu, k závěrečným pracím, provozuje Videotéku a e-shop publikací Slezské univerzity.

<sup>15</sup>Univerzitní knihovna *Slezská univerzita v Opavě*, 2016.

<sup>16</sup>Například v podpoře elektronických informačních zdrojů se jednotlivé součásti neshodnou. Přístupy jsou drahé a každá součást jim přiděluje jinou prioritu.

Tabulka 3.1: Počty studentů Slezské univerzity v Opavě v akademických letech 2011/2012 až 2014/2015

Rok	Bc.	Mgr.	Nav. Mgr.	Dokt.	Celkem
2011/12	4 052/2 329	134	947/727	89/180	8 430
2012/13	3 815/2 046	83	1 020/697	73/180	7 884
2013/14	3 518/1 987	44	1 079/697	70/149	7 530
2014/15	3 020/1 727	10	975/614	66/117	6 508

nické knihy z online digitální knihovny Ebrary, jejichž záznamy uživatelé naleznou v katalogu.<sup>17</sup>

Pro přiblížení skupiny uživatelů, které má UKSU sloužit, budeme vycházet z dat zveřejněných ve výročních zprávách univerzity a zprávách Ministerstva školství, mládeže a tělovýchovy. Tato skupina je složená ze studentů a akademických pracovníků univerzity. Údaje o nich nám také pomohou správně interpretovat situaci na univerzitě, ovlivňující knihovnu.

Data o studentech ve sledovaných letech naleznete v Tabulce č. 3.1.<sup>18</sup> Do hodnot nejsou započítáváni studenti, kteří byli přijati, ale své studium předčasně ukončili.<sup>19</sup> Data jsou převzata ze záznamů MŠMT, které je pravidelně aktualizuje (záznamy jsou aktuální k 20. 1. 2016).<sup>20</sup>

Údaje o pracovnících můžete vidět v Tabulce č. 3.2.<sup>21</sup> Také k jejímu zpracování bylo využito statistik MŠMT. V tabulce jsou uvedeni jak pracovníci akademičtí, tak součet všech zaměstnanců.

<sup>17</sup>I když záznamy v katalogu nejsou dlouho, bylo by například možné zjistit množství rezervací nebo výpůjček těchto knih a požádat danou součást o finance na nákup knih na základě oblíbenosti uživatelů z jejich fakulty.

<sup>18</sup>Výkonové ukazatele: VŠ studenti podle formy a typu studijního programu: podle vysoké školy/fakulty. *MŠMT Odbor analyticko-statistický*, 2015.

<sup>19</sup>Každoročně je těchto studentů asi 2 000 a jejich množství se dramaticky nemění. (*Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2014* [online]. Slezská univerzita v Opavě, 2015, [cit. 2016-02-20]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocnizpravy-a-dlouhodobyzamer/2014/vz-cinnost-2014.pdf>))

<sup>20</sup>Prezenční studenti jsou vždy uvedeni jako první, pak jsou uvedeni kombinovaní. Magisterské pětileté studium má pouze prezenční formu.

<sup>21</sup>Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2011. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/22951/download/>); Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2012. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/29843/download/>); Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2013. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/34251/download/>); Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2014. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/36358/download/>).



Tabulka 3.2: Počty pracovníků na SU v Opavě v akademických letech 2011/2012 až 2014/2015.

Akademický rok	Počet všech pracovníků	Počet akad. pracovníků
2011/2012	621	336
2012/2013	647	323
2013/2014	650	336
2014/2015	604	298

Tabulka 3.3: Počty potenciálních uživatelů UKSU v akademických letech 2011/2012 až 2014/2015.

Akademický rok	Počet potenciálních uživatelů
2011/2012	9 051
2012/2013	8 531
2013/2014	8 180
2014/2015	7 112

UKSU je knihovnou sloužící především akademickým pracovníkům a studentům Slezské univerzity v Opavě; ovšem v průběhu zpracovávání dat jsme zjistili, že knihovnu v určité frekvenci využívají také ostatní pracovníci univerzity. Proto jsme se rozhodli je započítat do tzv. potenciálních uživatelů, kteří tvoří součet výše zmíněných (viz Tabulka č. 3.3). Tyto údaje nám pomohou normalizovat výsledky dle případných trendů v počtech studentů či pracovníků na univerzitě; viz výroční zprávy.<sup>22,23</sup> Z nich vyvozujeme pokles potenciálních uživatelů v posledních letech.

<sup>22</sup> *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2011* [online]. Slezská univerzita v Opavě, 2012, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-zamer/2011/vz-cinnost-2011.pdf>); *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2012* [online]. Slezská univerzita v Opavě, 2013, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-zamer/2012/vz-cinnost-2012.pdf>); *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2013* [online]. Slezská univerzita v Opavě, 2014, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-zamer/2013/vz-cinnost-2013.pdf>); Výkonové ukazatele: VŠ studenti podle formy a typu studijního programu: podle vysoké školy/fakulty. *MŠMT Odbor analyticko-statistický* [online]. MŠMT, 2016, [cit. 2015-06-15]. Dostupný z: ([http://dsia.uiv.cz/vystupy/vu\\_vs\\_f2.html](http://dsia.uiv.cz/vystupy/vu_vs_f2.html)).

<sup>23</sup> *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2014*, 2016; Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2011. *MŠMT*, 2016; Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2012. *MŠMT*, 2016; Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2013. *MŠMT*, 2016; Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2014. *MŠMT*, 2016.

Tabulka 3.4: Statistiky Univerzitní knihovny Slezské univerzity v Opavě v akademických letech 2011/2012 až 2014/2015.

Akademický rok	2011/2012	2012/2013	2013/2014	2014/2015
Výpůjčky	32 336	36 238	31 798	33 858

Celkově knihovna disponuje fondem složeným převážně z odborné literatury o asi 175 000 svazcích. Z toho cca 135 000 svazků je umístěno v Opavě a cca 40 000 v Karviné. Ve volném výběru se nachází přibližně 60 000 svazků. Studijní místa v Karviné se nachází v obou lokacích, celkem je jich 86. Oproti tomu Opavské pracoviště disponuje pouze 35 studijními místy, z nichž většina je ve Studovně a půjčovně na Masarykově třídě. Omezené prostory tohoto pracoviště časem přinutily knihovnu, aby začala půjčovat také ve skladu na Hauerově ulici (kde se nachází také několik studijních míst).<sup>24</sup>

Počet výpůjček<sup>25</sup> za jednotlivé akademické roky naleznete v Tabulce č. 3.4. Data jsou převzata přímo z dostupných záznamů. Rádi bychom uvedli více popisných údajů o knihovně a jejích čtenářích, ale podrobné analýze se budeme přímo věnovat ve výsledcích, proto zde dalšího rozboru zanecháme.

## Popis dat

Systém Clavius ukládá data do databáze MySQL. Z této databáze byl zhotoven dump<sup>26</sup> zprovozněný v systému WampServer.<sup>27</sup> Z něj byly vytvořeny exporty do formátu CSV, který byl dále zpracováván v softwaru R<sup>28</sup> a Microsoft Excel. Výhody využití softwaru R jsme vyzvedli dříve. MS Excel byl použit převážně jako nástroj pro vizualizace.

Zdrojem dat ke zpracování jsou tedy data z databázových tabulek UKSU. Začátek sledovaného období je 1. 9. 2011, se začátkem nového akademického roku 2011/2012. Konec sledovaného období je 3. 8. 2015, kdy byl zhotoven dump. V záznamech jsou promítnuty čtyři akademické roky, rok 2011/2012, rok 2012/2013, rok 2013/2014 a rok 2014/2015. Rozdělení na akademické roky je v případě akademické knihovny logičtější, protože s novým rokem přichází nová vlna studentů, starší studenti naopak koncem semestru univerzitu opouští. Zároveň činnosti v rámci

<sup>24</sup>Data jsou vztažena ke sledovanému období.

<sup>25</sup>Bez prodlužování.

<sup>26</sup>Databázový dump je výstupem z databáze, který umožňuje zálohovat nebo duplikovat databázi. (Database dump. *Technopedia* [online]. [cit. 2016-02-15]. Dostupný z: <https://www.techopedia.com/definition/23340/database-dump>)

<sup>27</sup>*WampServer* [online]. [cit. 2015-06-20]. Dostupný z: <http://www.wampserver.com/en/>.

<sup>28</sup>Je software pod GNU licenci pro zpracování a vizualizaci dat. (*The R Project for Statistical Computing* [online]. R Foundation, [cit. 2015-06-20]. Dostupný z: <http://www.r-project.org/>)

předmětů jsou směřovány na ročníky či semestry, data jsme proto zpracovali s vazbou na akademické roky ne na roky kalendářní.

Data byla anonymizována, ovšem v databázi je stále uveden jednoznačný identifikátor uživatele, díky kterému jej můžeme přiřadit k jednotlivým činnostem. Čtenáře tak symbolizuje číslo, ze kterého sice nelze odvodit jakékoli osobní údaje, ale lze díky němu člověka identifikovat v systému.

UKSU má k dispozici moduly Katalogizace, OPAC (Carmen), Revize fondu a Výpůjční protokol.<sup>29</sup> Pokud by měla jiná knihovna k dispozici data také z jiných modulů, bylo by možné, že by se její možnosti analýz mohly lišit.

## 3.4 Využití dat z knihovnického softwaru Clavius

Zde rozdělíme vyhodnocení dat pod jednotlivé cíle a body, aby bylo možné dojít k závěrům. Především, že jako zdroj informací u prvního cíle budou sloužit články, které jsme jmenovali v teoretické části. Dále to budou data UKSU.

Po prostudování literatury jsme pro využití DM nad daty z Clavia vymezili tyto konkrétní oblasti:

- akvizice,
- analýza záznamů v katalogu,
- systémy pro podporu rozhodování,
- rozdělování financí,
- služby,
- průzkum chování uživatelů,
- analýzy dat ze sociálních sítí.

Mnoho ze zmíněných studií vycházelo z dat knihovny, propojených s jinými daty. Například s daty školy, nebo s daty dodanými od prodejců knih. My samozřejmě nemůžeme vycházet z údajů, které nemáme, ale můžeme posoudit, zda jsou uložené údaje v knihovním softwaru dostatečné.

### 3.4.1 Akvizice a data z Clavia

Prvním zvažovaným použitím dat je použití pro akvizici. Začneme tématem, které představil C. H. Wu, počítající s rozdělením dat o výpůjčkách pod jednotlivá oddělení (ústavy či katedry) a s jejich přepočtením na hodnotu, kterou pro konkrétní oddělení znamenají. Data mají být uložena do datového skladu, který

<sup>29</sup>Emailová korespondence s Evou Lackovou, 7. 3. 2016.

má pomáhat akvizitérům k rozvíjení fondu.<sup>30</sup> Podobný postup navrhuje i Kao; také on počítá s rozdělením výpůjček pod jednotlivá oddělení.<sup>31</sup>

Ačkoliv by bylo možné data o výpůjčkách uložit do datového skladu, rozdělit a analyzovat, chybí v datech z Clavia UKSU pro tuto analýzu podstatné údaje.<sup>32</sup> Konkrétně se jedná o zařazení pod jednotlivá oddělení (ústavy či katedry). Nejsou zde uložena ani data o příslušnosti studentů k oborům. Aplikace těchto postupů čistě na základě dat UKSU tedy není možná. Nevylučujeme, že by nebylo možné propojit databázi knihovny s databází SU a z těchto výsledků něco odvozovat; bohužel se však obáváme, že by ani tento postup nebyl jednoduchý. V databázi knihovny se totiž nevyskytuje studentské číslo, pod kterým jsou studenti SU identifikováni ve školním systému STAG.

Mitsuro Kitajima prováděl analýzy na základě záznamů knih v katalogu. Využíval znaků systematického selekčního jazyka, klíčová slova a data z výpůjčního protokolu. Pomocí nich pak hodnotil vývoj akvizičního procesu.<sup>33</sup> Tuto analýzu lze na datech z Clavia vykonat spojením tabulek `linear`,<sup>34</sup> `tituly`, `svazky` a `kpujcky`. První obsahuje bibliografické záznamy knih s názvy, klíčovými slovy a MDT, druhá obsahuje záznamy vlastněných titulů a umožní propojit bibliografické záznamy v `linear` se `svazky`. Tabulka `svazky` nese údaje o čase nákupu a ceně; umožní propojení se záznamy o půjčených dokumentech v tabulce `kpujcky`.

Na základě dat z Clavia je možné i provedení evaluace akvizice. Na druhou stranu, data v každé z tabulek jsou uložena jinak, a provedení vhodných SQL dotazů k získání dat a jejich následné propojení není triviální záležitost. Rozhodně vyžaduje dobrou znalost databáze Clavia (jazyka SQL), znalost práce s daty, statistikami, vizualizacemi a interpretací dat. Na druhou stranu, o průkaznosti výsledků nebude pochyb. Výsledky podobné analýzy by poskytly výborný přehled o tom, jakým způsobem jsou finance na akvizici využívány.<sup>35</sup>

Yang dále navrhuje zřízení systému nabízení knih k nákupu. Tento systém krom jiného zohledňuje nenalezené dotazy při vyhledávání<sup>36</sup> a z dat z Clavia je možné jej

---

<sup>30</sup>WU, Chien-Hsing. Data mining applied to material acquisition budget allocation for libraries: design and development. In: *Expert Systems with Applications*. 2003, roč. 25, č. 3, s. 401–411. ISSN 0957-4174.

<sup>31</sup>KAO, S.-C. et al. Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management*. 2003, roč. 39, č. 1, s. 133–147. ISSN 0306-4573.

<sup>32</sup>UKSU nemá zakoupen modul akvizice.

<sup>33</sup>KITAJIMA, Mitsuro et al. An Evaluation of Book Selection in a University Library by Loan Record Analysis. *International Journal of Information and Education Technology*. 2015, roč. 5, č. 10, s. 728–731. ISSN 2010–3689.

<sup>34</sup>Názvy tabulek Clavia budeme zvýrazňovat, aby bylo jasné, že jde o název.

<sup>35</sup>Toto je obecné zjištění. Pokud se jedná o konkrétní situaci UKSU, je to trochu složitější. Knihovna má vlastní rozpočet oddělený od jednotlivých součástí jen několik posledních let.

<sup>36</sup>YANG, Shih-Ting. An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*. 2012, roč. 10, č. 2, s. 163–173. ISSN 1741-5063.

zprovoznit. Bylo by ale třeba, aby některý z knižních dodavatelů doplnil (a pravidelně dodával) záznamy o nabízených knihách. Clavius ukládá nalezené výrazy do tabulky `car_cloud` a nenalezené výrazy do tabulky `car_expression_not_found`; ty by sloužily jako ideální zdroj výrazů pro sestavení tezauru. Dále by bylo třeba provést mapování fondu dle klíčových slov pomocí údajů v tabulce `linear`. Také by bylo nutné zjistit půjčovnost fondu pomocí tabulek `tituly`, `svazky` a `kpujcky` jako v předešlém případě. Poslední část úkolu by spočívala v nalezení shody nabízených knih a požadovaných témat.

Pro provedení tohoto úkolu platí podobné požadavky jako v předchozím případě, a ještě jeden navíc. Pro sestavení tezauru je potřeba poměrně vysoké odbornosti v oblasti zpracování přirozeného jazyka a znalosti v oboru fuzzy systémů. Ačkoli je tedy Yangův systém nejkompaktnější a jako pomoc akvizitérům také nejvhodnější, je také nejnáročnější.

#### 3.4.2 Analýza bibliografických záznamů v na datech z Clavia

Další z možností využití dat z Clavia je průzkum, vizualizace či popis sbírky dokumentů na základě jejich bibliografických údajů v katalogu. Výše zmíněné studie (William Dentona,<sup>37</sup> Timothy Dickeyho<sup>38</sup>), jsou samozřejmě příklady zpracování takových přehledů o různém rozsahu a o různé hloubce. V mnohých případech jsou zpracovávána data několika subjektů. My se budeme zabývat jen těmi příklady, které byly realizovány na omezeném vzorku. Analýzy bibliografických záznamů z hlediska jedné knihovny možná nemají tak silný sociokulturní či vědecký náboj, viz jiné studie (Dickeyho článek<sup>39</sup> a článek Mary Jackson<sup>40</sup>), pořád však mohou sloužit jako validní nástroj managementu.

První je průzkum William Dentona. Ten se snažil vizualizovat záznamy ve fondu pomocí Deweyho desetinného třídění a Třídění Kongresové knihovny.<sup>41</sup> Vzhledem k tomu, že má UKSU tolik pracovišť a její fond je navíc rozdělen mezi Opavu a Karvinou, byla by podobná analýza jistě vítána. Dobrou zprávou je, že by bylo možné na datech z Clavia analýzu provést a nebylo by to ani tak složité. Stačí využít data jedné tabulky; konkrétně `linear`. Tabulka obsahuje také záznamy o lokaci knihy; záznamy by tedy bylo možné třídit také dle umístění.

---

<sup>37</sup>DENTON, William. On dentographs, a new method of visualizing library collections. *Code4Lib Journal*. 2012, roč. 16, s. 171–179. Dostupný také z: <http://journal.code4lib.org/articles/6300>. ISSN 1940-5758.

<sup>38</sup>DICKEY, Timothy J. Books as Expressions of Global Cultural Diversity. *Library Resources and Technical Services*. 2011, roč. 55, č. 3, s. 148–162.

<sup>39</sup>Dickey, Timothy J. 2011.

<sup>40</sup>JACKSON, Mary Ellen. *Changing Global Book Collection Patterns in ARL Libraries*. 2006.

<sup>41</sup>Denton, William, 2012.

### 3 Analýza dat z knihovního softwaru Clavius

Pro provedení této analýzy není nutná ani rozsáhlá znalost jazyka SQL, ani podrobné seznámení s databází. Zato je užitečné mít přehled v polích MARCu,<sup>42</sup> pod jejichž označeními se zmíněné informace nacházejí, dále mít zkušenosti s vizualizačními nástroji, prací s daty a znalosti ve statistice. Co se týká průkaznosti výsledků, zpracované vizualizace budou vypovídat přesně o obsahu fondu. Interpretace samozřejmě závisí na knihovně.

Další možností, jak využít data z Clavia je identifikace knih k digitalizaci. V literatuře jsou tyto publikace vybírány na základě porovnávání obsahu fondu jednotlivých knihoven a jako kandidáti jsou označovány knihy, které jsou ohroženy degradací nebo ztrátou.<sup>43</sup> Pokud za jeden z faktorů ohrožení dosadíme také nadměrnou půjčovanost knihy, která může ohrozit její fyzický stav, můžeme analýzu provést i bez nutnosti mít k dispozici data jiných knihoven.<sup>44</sup> Stačí nám k tomu propojit tabulky `tituly`, `svazky` a `kpujcky`, ze kterých zjistíme nejpůjčovanější dokumenty; ty pak po prověření fyzického stavu můžeme navrhnout k digitalizaci.<sup>45</sup>

K provedení analýz je nutná znalost databáze Clavia, jazyka SQL, práce s daty, statistiky, vizualizace a interpretace dat. Vzhledem k tomu, že z dat nelze jasně určit, jaký je přesný fyzický stav knihy, je výsledek méně průkazný než výše uvedené analýzy.

---

<sup>42</sup>„Výměnný formát pro zápis a přenos bibliografických dat; specifikuje pole údajů, návěští polí (tagy), indikátory a kódy podpolí, které mají být přiděleny jednotlivým údajům při popisu dokumentů ve strojem čitelné podobě. Usnadňuje výměnu dat mezi bibliografickými agenturami a jinými institucemi.“ (VODIČKOVÁ, Hana. KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV). In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. [cit. 2016-02-15]. Praha: Národní knihovna ČR, 2003–. Dostupný z: ([http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001236&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001236&local_base=KTD)))

<sup>43</sup>SILIPIGNI CONNAWAY, Lynn, O'NEILL, Edward T. et al. Last Copies: What's at Risk? *College and research libraries*. 2006, roč. 67, č. 4, s. 370–379. Dostupný také z: (<http://www.oclc.org/research/publications/archive/2006/connaway-cr107.pdf>). ISSN 0010-0870.

<sup>44</sup>Při digitalizaci publikací je nutné vyřešit autorská práva. (BOUCHET, Emilie. Autorské právo a digitalizace v archivech, muzeích a knihovnách. *Knihovna*. 2008, roč. 19, č. 1, s. 48–55. Dostupný také z: (<http://oldknihovna.nkp.cz/knihovna81/81048.htm>). ISSN 1801-3252. )

<sup>45</sup>Je nutné poznamenat, že UKSU již má v Digitální knihovně Krameria 4 přístupné některé publikace. Jsou to převážně skripta, u kterých je jisté, že je studenti ve zvýšené míře potřebují a ke kterým škola drží autorská práva. (*Digitální knihovna K4* [online]. [cit. 2016-02-15]. Dostupný z: (<http://kramerius.slu.cz/search/#intro6>))

### 3.4.3 Systémy pro podporu rozhodování a Clavius

Pomineme-li teoretická témata, která probírali Kevin Cullen<sup>46</sup> a Nicholson<sup>47</sup>, jenž jsme zmínili v části „Systémy pro podporu rozhodování“ a články o konkrétních systémech (např. od OCLC),<sup>48</sup> dostaneme se k tématu zpracovaném ve studii Papavlasopoulou a Poulou. Prověřují použití neuronové sítě k přehlednému hodnocení knihovny. Snaží se převést 43 indikátorů, které se používají pro hodnocení a porovnávání výkonů knihoven v Řecku, do jedné hodnoty.<sup>49</sup>

Také v českých knihovnách se používají výsledky zaznamenané ve statistikách systému k tomu, aby pomáhaly porovnávat knihovny mezi sebou.<sup>50</sup> Kromě jiného tyto hodnoty slouží k vykazování činnosti popř. úspěšnosti zřizovateli.<sup>51</sup>

Clavius umí zpracovávat základní statistiky, které jsou pak součástí výkazů o činnosti. Pro podobné zpracování analýz jako u Papavlasopoulou a Poulou, by bylo třeba mnohem většího množství dat z různých knihoven. Clavius data pro porovnání s ostatními knihovnami<sup>52</sup> neobsahuje. Konkrétním využitím dat z Clavia v rámci podpory rozhodování se budeme zabývat při vyhodnocování druhého cíle práce. Zmíníme se také o statistikách systému Clavius.

### 3.4.4 Rozdělování financí a data z Clavia

Clavius, jakožto systém primárně pro městské knihovny, neobsahuje data o počtu studentů (magisterských i bakalářských studií) ani data o EIZ, která Walters navrhuje pro tento účel zpracovávat.<sup>53</sup> Waltersův systém by umožnil výpočet

---

<sup>46</sup>CULLEN, Kevin. Delving into data. *Library Journal*. 2005, roč. 130, č. 13, s. 30–32. ISSN 0363-0277.

<sup>47</sup>CLEYLE, Su et al. Approaching librarianship from the data: using bibliomining for evidence-based librarianship. *Library hi tech*. 2006, roč. 24, č. 3, s. 369–375. ISSN 0737-8831.; NICHOLSON, Scott, STANTON, Jeffrey M. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. 2003, s. 247–262; NICHOLSON, Scott. A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*. 2004, roč. 60, č. 2, s. 164–182. ISSN 0022-0418.

<sup>48</sup>Při vyhodnocení tohoto cíle nás zajímají pouze praktické aplikace, které mluví o konkrétních datech, ne teoretické koncepty. A nezajímá nás ani konkrétní software.

<sup>49</sup>PAPAVLASOPOULOS, Sozon et al. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, roč. 33, č. 4/5, s. 261–271. ISSN 0143-5124.

<sup>50</sup>Adresáře institucí: statistika: knihovna. *Odborný útvar NIPOS* [online]. Ministerstvo kultury ČR, [cit. 2016-02-15]. Dostupný z: <http://www.nipos-mk.cz/?cat=88>.

<sup>51</sup>KRČÁL, Martin. Sběr a využití statistických dat v českých knihovnách. *ProInflow*. 2015, roč. 7, č. 2, s. 41–52. Dostupný také z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/1163>.

<sup>52</sup>Jako například software od OCLC.

<sup>53</sup>WALTERS, William H. A regression-based approach to library fund allocation. *Library Resources and Technical Services*. 2006, roč. 51, č. 4, s. 263–278. ISSN 0024-2527.

průměrné ceny za odbornou publikaci a to z tabulky **svazky**; ovšem tuto cenu by nedokázal vztáhnout k danému oboru, protože data o oboru (studovaném či vyučovaném) neukládá. Vztahení odborné publikace k tématu by bylo možné udělat přes klíčová slova nebo přes symbol MDT (Mezinárodního desetinného třídění), ale znovu bychom narazili na problém vztahení dat k oborům vyučovaným na univerzitě. Tato data by bylo potřeba dodat z externího zdroje. Již dříve jsme vyjádřili pochybnost nad možnostmi propojení s daty z univerzitního systému STAG. Úspěšné zopakování Waltersovy studie čistě na datech z Clavia proto není možné.

#### 3.4.5 Služby a DM na datech z Clavia

Vynecháme-li téma o zpracování dat z bibliometrické analýzy (služba pro univerzitu)<sup>54</sup> a pomineme-li také téma studie Reinholda Deckera (jehož výzkum byl založen na datech z dotazníkového šetření),<sup>55</sup> dostaneme se k tématu analýza dat z referenčních dotazů. Toto téma reprezentuje například článek Nadaleen Tempelman-Kluit, která jich využívala k tvorbě person (persona sloužila jako nástroj user-centered design).<sup>56</sup>

V UKSU jsou referenční služby poskytovány pouze na základě fyzické přítomnosti uživatele a nejsou nijak zaznamenávány. V Claviu ani není prostor, kam nahrávky nebo zápisy pro podobnou analýzu umístit. K provedení něčeho podobného nám tedy chybí data.<sup>57</sup>

Ze stejného důvodu musíme vynechat také studii Fontaneho a Finella, kteří analyzovali záznamy o referenčních dotazech v souvislosti s fondem,<sup>58,59</sup> nebo Jody Condit Fagan, která hledala závislosti v datech mezi referenčními službami, účastí na lekcích, vyhledáváním, požadavky na nákup a stažením dokumentů.<sup>60,61</sup>

---

<sup>54</sup>WILL, Nicole. Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*. 2006, roč. 73, č. 8, s. 1045–1050. ISSN 0040-1625.

<sup>55</sup>DECKER, Reinhold et al. Planning and evaluation of new academic library services by means of web-based conjoint analysis. *The Journal of academic librarianship*. 2006, roč. 32, č. 6, s. 558–572. ISSN 0099-1333.

<sup>56</sup>TEMPELMAN-KLUIT, Nadaleen et al. Invoking the User from Data to Design. *College and Research Libraries*. 2014, roč. 75, č. 5, s. 616–640. Dostupný také z: (<http://dx.doi.org/10.5860/cr1.75.5.616>). ISSN 0010-0870.

<sup>57</sup>Pro analýzu, kterou představila Tempelman-Kluit, by bylo možné využít tabulku **ctenari**, kterou by bylo nutné doplnit o data o referenčních službách.

<sup>58</sup>FINNELL, Joshua et al. Reference question data mining: A systematic approach to library outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.

<sup>59</sup>Pro studii Fontaneho a Finella platí stejná omezení, jako u Tempelman-Kluit.

<sup>60</sup>CONDIT FAGAN, Jody. The effects of reference, instruction, database searches, and ongoing expenditures on full-text article requests: an exploratory analysis. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 3–4, s. 264–274. ISSN 0099-1333.

<sup>61</sup>Pokud bychom se rozhodli vynechat proměnnou referenčních služeb, bohužel bychom si nepomohli. V systému nejsou záznamy ani o účasti na lekcích, ani o požadavcích na nákup a už vůbec ne o stažených člancích.



Jedinou analýzou, pro kterou Clavius obsahuje data, je hledání trendů v požadavcích MVS. Oproti koncepci Eve-Marie Lacroix<sup>62</sup> by šlo sice o čistě lokální záležitost,<sup>63</sup> ale také z této analýzy by se dalo mnohé pro UKSU odvodit. Bylo by to možné provést s využitím tabulek z Clavia tituly, svazky a kpujcky. Systém v nich zaznamenává požadavky MVS a jejich výpůjčky s časovým údajem, tituly, jejich stáří, autora, nakladatelství a další. Všechny tyto proměnné by se dalo využít k hledání trendů.

Pro provedení těchto analýz by bylo třeba dobré znalosti databáze Clavia, také jazyka SQL, zkušenosti s prací s daty,<sup>64</sup> statistikou, vizualizacemi a interpretací dat. Z hlediska průkaznosti výsledků by knihovna získala, kromě jiného, solidní základ pro návrh nákupu literatury, která se v požadavcích studentů a vyučujících opakuje.

#### 3.4.6 Evaluace OPACu či webu knihovny pomocí dat z Clavia

Sledování průchodu katalogem na základě transakčních logů z OPACu, které prováděla Debora Blecic<sup>65</sup> bohužel na datech z Clavia provést nelze. Systém totiž neukládá všechny interakce (například navigaci v hit listu), některé sice uloží, ale bez vazby na uživatele, který je provedl (například nenalezené výrazy). Zachycené údaje by bylo možné spojit s činností jednoho uživatele pouze na základě časového údaje, což by znamenalo, že výsledky by byly velmi přibližné, protože čtenářů může být v systému samozřejmě mnoho; systém je ve většině případů od sebe neodliší. Kromě toho by neposkytovaly úplný odraz průchodu uživatele katalogem.

Paul Bracke ve své práci prováděl analýzu nad daty z webových stránek a OPACu a zjišťoval vzájemné závislosti. Jeho cílem bylo podobně jako u Debory Blecic, zjednodušit cestu uživatele za hledanými informacemi.<sup>66</sup> Bohužel, pro provedení podobné analýzy na datech z Clavia nemáme dostatek dat. Museli bychom spoléhat na data z externího zdroje o průchodu webových stránek. Kromě toho bychom museli počítat s omezeními, zmíněnými v předchozím odstavci, čili museli bychom spárovat data pouze na základě časového údaje, což snižuje kredibilitu výsledků.

---

<sup>62</sup>LACROIX, Eve-Marie et al. Interlibrary loan in US and Canadian health sciences libraries 2005: update on journal article use. *Journal of the Medical Library Association*. 2007, roč. 95, č. 2, s. 189–194. ISSN 1536-5050.

<sup>63</sup>Navíc bez elektronických článků.

<sup>64</sup>Zvláště proto, že je nutné odfiltrovat dokumenty požadované přes MVS z běžných titulů a svazků, které má knihovna k dispozici.

<sup>65</sup>BLECIC, Deborah D. et al. Using transaction log analysis to improve OPAC retrieval results. *College and Research Libraries*. 1998, roč. 59, č. 1, s. 39–50. ISSN 0010-0870.

<sup>66</sup>BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *J. Med. Libr. Assoc.* 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>).

Jak jsme podotkli, analýza webových logů ze stránek knihovny (po příkladu Jiann-Chernga Shieha<sup>67</sup>) by musela být závislá na externím zdroji dat. Proto ji u vyhodnocení vynecháme.

Jako poslední aplikaci průzkumu transakčních logů uvádíme práci Michaly Soškové, vycházející z dat systému Aleph.<sup>68</sup> Ten, na rozdíl od Clavia, zaznamenává pečlivěji průchod čtenáře katalogem a všechny jeho interakce. Proto bylo provedení analýzy transakčních logů na těchto datech možné. Na datech z Clavia tento postup uplatnit nelze.

#### 3.4.7 Průzkum chování uživatelů a data z Clavia – vliv knihovny na retenci studentů

Systém Clavius neobsahuje data o studijních úspěších uživatelů, ani záznamy o začátku či konci studia, stejně jako data o studovaném oboru či příslušnosti k ústavu. Aplikace dat z Clavia je tedy v tomto ohledu značně omezená. Závisela by na dodání dat z externího zdroje a také na úspěšném provázání těchto dat s uživatelskou tabulkou *ctenari*. Tato externí data by musela být poskytnuta systémem školy, jak navrhoval Doug Johnson<sup>69</sup> nebo pomocí dotazníkového šetření, jak to provedla Ethelene Whitmire.<sup>70</sup>

V této části vyhodnocení musíme vynechat témata zmíněná ve studiích, u nichž zdrojem dat nebyl knihovní software.<sup>71,72</sup>

---

<sup>67</sup>SHIEH, Jiann-Cherng. From website log to findability. *The Electronic Library*. 2012, roč. 30, č. 5, s. 707–720. ISSN 0264-0473.

<sup>68</sup>SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného on-line katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)).

<sup>69</sup>JOHNSON, Doug. A data mining primer and implications for school library media specialists. *Knowledge Quest*. 2004, roč. 32, č. 5, s. 32–35.

<sup>70</sup>WHITMIRE, Ethelene. Academic library performance measures and undergraduates' library use and educational outcomes. *Library and information science research*. 2002, roč. 24, č. 2, s. 107–128. ISSN 0740-8188.

<sup>71</sup>SORIA, Krista M. et al. Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 1, s. 84–91. ISSN 0099-1333.; EMMONS, Mark et al. The academic library impact on student persistence. *College and Research Libraries*. 2011, roč. 72, č. 2, s. 128–148. ISSN 0010-0870.; WEINER, Sharon. The Contribution of the Library to the Reputation of a University. *The Journal of Academic Librarianship*. 2009, roč. 35, č. 1, s. 3–13. ISSN 0099-1333.

<sup>72</sup>Whitmire, Ethelene, 2002.

### 3.4.8 Průzkum chování uživatelů a data z Clavia – analýza využívání fondu

Prvním způsobem použití dat z knihovního systému, který navrhuje Stephen Zweibel a Zachary Lane, je analýza výpůjček, vracení dokumentů a online aktivit uživatelů za účelem nalezení vzorců v uživatelském chování.<sup>73</sup> Pro podobnou analýzu bychom v Claviu využili tabulky `ctenari`, `kpujcky`, `car_statistics_login` (která obsahuje přihlašování uživatelů do OPACu) a `car_statistics_access` (která obsahuje všechny přístupy do OPACu). Rozšíření o další tabulky by záviselo na zaměření průzkumu.

Musíme však počítat s nízkou propojitelností tabulek, které nesou údaj o přístupu `car_statistics_access` a o vyhledávání `car_cloud` (úspěšné vyhledávání) a `car_expression_not_found` (neúspěšné vyhledávání). Z hlediska analýz online interakcí by tedy byla průkaznost výsledků omezená.

Propojení tabulek vyžaduje znalosti o databázi Clavius, zkušenosti s prací s daty, vizualizacemi, statistikou a rozhodně dobré interpretační schopnosti. Analýzy nejsou náročné na zpracování, ale aby je bylo možné použít v praxi, je třeba znalost situace UKSU a dobrý vhled do problematiky.

Druhý způsob použití dat demonstruje studie z Marstonské vědecké knihovny. Do svých analýz kromě předchozích proměnných zahrnula ještě využívání e-knih, službu MVS a rozčlenění aktivit dle oddělení, témat a oborů.<sup>74</sup> My víme, že rozdělení dat dle oborů a oddělení na univerzitě není možné u Clavia provést, nicméně pro doplnění analýz o hlediska využívání e-knih a služby MVS by stačilo doplnit analýzu o data z tabulek `tituly` a `svazky`. Požadavky na analytika by byly stejné jako v předchozím případě plus znalost jazyka SQL (kvůli přípravě dat).

Goswami se zabýval vytvářením komunit pomocí dat z výpůjčního protokolu. Každá výpůjčka, provedená uživatelem, reprezentovala spojení s ostatními knihami, které si uživatel vypůjčil.<sup>75</sup> Myšlenka konstrukce uživatelských skupin podle zájmů, reflektovaných ve výpůjčkách, je velmi přitažlivá. V Claviu by pro podobnou analýzu stačila jediná tabulka – `kpujcky`. Nebylo by třeba informací o tom, jaký obor uživatel studuje nebo kde pracuje; ani bychom nemuseli vědět, o čem půjčená kniha je. Nechali bychom data „promluvit“ v pravém slova smyslu.

---

<sup>73</sup>ZWEIBEL, Stephen et al. Probing the Effects of Policy Changes by Evaluating Circulation Activity Data at Columbia University Libraries: Edited by Rick J. Block. *The Serials Librarian*. 2012, roč. 65, č. 1, s. 17–27. ISSN 0361-526X.

<sup>74</sup>LEONARD, Michelle F. et al. Metrics and science monograph collections at the Marston Science Library, University of Florida. *Issues in Science and Technology Librarianship*. 2010, č. 62, s. 1. Dostupný také z: (<http://dx.doi.org/10.5062/F4PC308T>). ISSN 1092-1206.

<sup>75</sup>GOSWAMI, Sumit et al. Visualisation of Relationships Among Library Users Based on Library Circulation Data. *DESIDOC Journal of Library & Information Technology*. 2010, roč. 30, č. 2, s. 26–39. ISSN 0974-0643.

Goswami navrhoval výsledky využívat při doporučování knih. Kromě toho by data mohla nahrazovat údaje o oborech (předpokládáme-li, že si uživatelé stejného oboru půjčují podobnou literaturu), přidáním dalších proměnných by mohla také podávat bližší informace o uživatelských skupinách, uživatelském chování a fondu. Pro vytvoření podobné analýzy by bylo třeba znalostí v oblasti teorie grafů a programování, také by byl nezbytný vhled do datové struktury Clavia. Výsledky by samozřejmě vyžadovaly správnou interpretaci, ale jejich vliv by byl značný.

Téma zpracování přístupů k elektronickým informačním zdrojům můžeme z naší analýzy vynechat. Clavius totiž žádná data o přístupech k EIZ neobsahuje. Tyto údaje by bylo nutné doplnit z externích zdrojů (vyžádat si je od producentů nebo zprostředkovatelů databází).

Identifikací „zkušeného uživatele“ v OPACu, či v databázích knihovny se zabývá Ahmad.<sup>76</sup> Ačkoli by tato analýza měla jistě svůj přínos, možnosti jejího provedení na datech z Clavia jsou omezené. Jak jsme podotkli výše, všechna data v OPACu nelze jednoznačně přiřadit jednomu uživateli. Na druhou stranu, pokud by bylo možné „zkušeného uživatele“ identifikovat v systému pouze na základě údajů, které lze spojit, potom by bylo možné analýzu provést.

Pokud budeme vycházet z dosavadních zkušeností, můžeme usoudit, že je analýzu možné provést pomocí propojení tabulek `car_statistics_login` (přihlášení), `car_cloud` (úspěšné vyhledávání), `car_statistics_score` (označování pomocí like, hvězdičky nebo komentářem), `car_settings_user` (nastavení) a `car_favorite_tag` (seznam literatury). Tyto reprezentují aktivity, které čtenář může v Carmen vykonávat. Musíme však počítat s tím, že propojení dat bude vyžadovat nemalé úsilí, dobré znalosti jazyka SQL, databáze Clavia a vhled do problematiky. Zároveň vyvstanou omezení při interpretaci dat.

#### 3.4.9 Průzkum chování uživatelů a data z Clavia – DM analýza k organizaci fondu

Zpracování analýzy, umožňující lepší organizaci fondu, se věnoval Jaap Boter. Zabýval se zlepšením členění fondu do kategorií ve veřejných knihovnách.<sup>77</sup> Naše data ale pochází z akademické knihovny, pro kterou by neměl tento druh analýzy smysl. Vzhledem k četnosti zastoupení Clavia v městských knihovnách problém prozkoumáme.

Pro provedení analýzy na datech z Clavia by bylo třeba údajů z datových struktur: `kpujcky`, `ctenari`, `tituly`, `svazky` a `linear`. Jejich propojení by nebylo

---

<sup>76</sup>AHMAD, Pervaiz et al. The e-book power user in academic and research libraries: Deep log analysis and user customisation. *Australian Academic & Research Libraries*. 2014, roč. 45, č. 1, s. 35–47. ISSN 0004-8623.

<sup>77</sup>BOTER, Jaap et al. User categorization of public library collections. *Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. Dostupný také z: (<http://dx.doi.org/10.1016/j.lisr.2005.01.004>). ISSN 0740-8188.

triviální, vyžadovalo by dobré znalosti jazyka SQL, databáze Clavia, statistiky, DM metod a zkušenosti s prací s daty.

#### 3.4.10 Analýza dat ze sociálních sítí – vyhodnocení

Analýza dat ze sociálních sítí navrhovaná ve studii „Who is following us“ využívala externích dat.<sup>78</sup> Ani Clavius neobsahuje žádná data související se sociálními sítěmi. Záznamy v Carmen naopak obsahují hodnocení publikací čtenáři v tabulce `car_statistics_score`. Čtenáři mají možnost se k publikaci vyjádřit trojím způsobem; „ohvězdičkováním“, udělením „like“ a sepsáním hodnocení. Všechny činnosti kromě udělení „like“, může čtenář provádět pouze pod přihlášením. Aktivitu čtenářů lze v tomto smyslu sledovat bez vazby na sociální síť. Pokud bereme tzv. lajkování (nebo využívání jiného typu hodnocení) a komentování jako typický projev aktivity uživatele v prostředí Webu 2.0,<sup>79</sup> můžeme také komentáře v Carmen a přidělování hvězdiček či hodnocení „like“ považovat za tento typ chování.

Z toho plyne, pokud by UKSU chtěla zjistit, kdo ji sleduje na sociálních sítích, musela by získat externí data – například ze sociální sítě Facebook, na níž má knihovna profil.<sup>80</sup> Pokud by knihovně šlo o sledování aktivit uživatele v prostředí podobnému Webu 2.0, bylo by možné využít záznamů v tabulce `car_statistics_score` (obsahující hodnocení publikací).

#### 3.4.11 Sumarizace výsledků

Probrané způsoby analýzy dat na knihovních datech a ověření jejich použití na datech ze systému Clavius nyní shrneme v Tabulce č. 3.5. Do červena zabarvené řádky reprezentují témata, do kterých bylo vyhodnocení zařazeno, žlutě jsou v tabulce označena jména vědců, kteří se daným zpracováním zabývali. První sloupec ukazuje, zda je možné data z Clavia přímo vzít a provést na nich danou

<sup>78</sup>SEWELL, Robin R. Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*. 2013, roč. 31, č. 1, s. 160–170. Dostupný také z: (<http://dx.doi.org/10.1108/07378831311303994>). ISSN 0737-8831.

<sup>79</sup>„Technologie, prvky a služby, které kladou důraz na komunitu a její vzájemnou interakci, umožňují uživatelům vytvářet obsah webu (informace, hudbu, obrázky, videa apod.) sdílet jej s jinými uživateli, a komunikovat navzájem. Knihovny mohou využívat prvky Webu 2.0 (instant message, diskusní fóra, chaty, blogy, RSS a další služby) v rámci jedné instituce nebo ve spojení s jinými institucemi po celém světě za účelem rychlejší komunikace. Na principu Webu 2.0 je založena Knihovna 2.0.“ (HAVLOVÁ, Jaroslava et al. Web 2.0. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003–, [cit. 2016-02-15]. Dostupný z: ([http://aleph.nkp.cz/F/?func=direct&doc\\_number=000014569&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000014569&local_base=KTD)))

<sup>80</sup>Knihovna má v současnosti asi 564 fanoušků na sociální síti Facebook. (Timeline Univerzitní knihovny SU v Opavě. *Facebooková stránka Univerzitní knihovny SU v Opavě* [online]. [cit. 2016-02-15]. Dostupný z: ([https://www.facebook.com/UniverzitaSUvOpav%C3%83%C2%AD-knihovna-SU-v-Opav%C3%84%C2%9B-212446575439142/timeline?ref=page\\_internal](https://www.facebook.com/UniverzitaSUvOpav%C3%83%C2%AD-knihovna-SU-v-Opav%C3%84%C2%9B-212446575439142/timeline?ref=page_internal)))

analýzu bez toho, aby k nim bylo cokoli přidáno. V mnohých případech by byla potřeba dodat další data z externího zdroje. My můžeme vycházet pouze z dat, která máme přímo k dispozici, tj. z dat UKSU. Musíme také vzít na vědomí, že knihovna nemá zakoupeny všechny moduly Clavia. Naše výsledky jsou tedy platné pro UKSU a knihovny, které mají k dispozici právě takové moduly jako UKSU.

Sloupec „použitelné tabulky“ reprezentuje tabulky z Clavia, které bude třeba k dané analýze zpracovat. „Průkaznost“ určuje, jak spolehlivý by byl výsledek analýzy. V případě, že to není možné předpovědět (zvláště kvůli nedostatku dat), je údaj vynechán. „Náročnost provedení“ vyjadřuje obtížnost provedení analýzy; vycházíme především z toho, jak jsou data v tabulkách strukturována<sup>81</sup> a jestli by bylo nutné využít dalších znalostí.<sup>82</sup>

## 3.5 Podpora rozhodování v knihovně pomocí DM

Nyní budeme postupovat ve dvou krocích. Nejdříve se pokusíme zjistit, analýzu jakých dat můžeme na základě dat z Clavia provést bez toho, aby nám chyběly údaje. Následně bude například zřejmější, jestli lze oblasti v budoucnu propojit. Druhý krok bude souviset s provedením analýz na datech z Clavia, ovšem jak jsme předeslali v podkapitole „Cíle praktické části“, bude nutné se specializovat pouze na jeden okruh problémů (tento okruh symbolizuje chování uživatelů).

- Jaké okruhy problémů při řízení knihovny lze daty podpořit.
- Na základě předchozích zjištění provést data mining nad daty z Clavia.

Během vyhodnocování vhodnosti dat z Clavia pro provedení konkrétních analýz jsme došli k různým závěrům. V některých případech DM využít lze, a o spolehlivosti výsledků nebude pochyb. Jindy k analýze chyběla data, závěry by nebylo možné udělat, nebo by byly neprůkazné.

Proto v prvním kroku, nemůžeme postupovat odshora dolů, tj. stanovit si okruhy, s nimiž bychom potřebovali pomoci a data zpracovat. Mnohdy totiž potřebná data chybí. Budeme muset postupovat zdola nahoru, tj. zjistit, jaké údaje systém ukládá a co bychom jejich analýzou mohli získat.

Clavius je modulární, to také znamená, že podle modulů data ukládá. UKSU má zakoupeny moduly Katalogizace, OPAC (Carmen), Revize fondu a Výpůjční protokol.<sup>83</sup> Systém ukládá údaje do 139 tabulek. Z těchto tabulek je pro analýzu vhodných pouze několik. Většina obsahuje systémové definice, definice hodnot,

<sup>81</sup>Složitější struktura uložení značí vyšší obtížnost zpracování. Například tabulka `linear`.

<sup>82</sup>Například znalosti ve fuzzy logice nebo v oblasti neuronových sítí.

<sup>83</sup>Emailová korespondence s Evou Lackovou, 7. 3. 2016.

### 3 Analýza dat z knihovního softwaru Clavius

Tabulka 3.5: Souhrn způsobů použití dat z Clavia

	Aplikace na datech UKSU	Použitelné tabulky	Průkaznost v souč. stavu	Náročnost provedení
<b>Akvizice</b>				
Wu + Kao	ne (chybí data z akviz. modulu)	kpujcky, linear, svazky, tituly	X	v souč. stavu není možné zjistit
Kitajima	ano	linear, tituly, svazky, kpujcky	vysoká	vysoká
Yang	ne (potřeba ext. dat)	car_cloud, linear, tituly, svazky, kpujcky, car_expression_not_found	X	vysoká (potřeba spec. znalostí)
<b>Analýza bib. záznamů</b>				
Denton	ano	linear	vysoká	nízká
<b>Sys. pro podp. rozhodování</b>				
Papavlasopoulos	ne (potřeba ext. dat)	kpujcky	X	v souč. stavu není možné zjistit
<b>Rozděl. financí</b>				
Walters	ne (potřeba ext. dat)	svazky	X	v souč. stavu není možné zjistit
<b>Služby</b>				
Tempelman-Kluit + Fontane + Fagan	ne (potřeba ext. dat)	ctenari	X	v souč. stavu není možné zjistit
Lacroix	ano	kpujcky, tituly, svazky	střední	střední
<b>Evaluace OPACu a webu</b>				
Blecic + Soskova	ne (špatné uložení)	data jsou neprovázatelná	X	střední
<b>Průzkum chování</b>				
Zweibel	ano	kpujcky, ctenari, car_statistics_login, car_statistics_access, car_expression_not_found, car_cloud	částečná (kvůli údajům o online aktivitách)	střední
Leonard	ne (potřeba ext. dat)	kpujcky, ctenari, car_statistics_login, car_statistics_access, tituly, svazky	X	vysoká
Goswami	ano	kpujcky	vysoká	vysoká (potřeba spec. znalostí)
Ahmad	ne (potřeba ext. dat)	car_statistics_login, car_cloud, car_statistics_score, car_settings_user, car_favorite_tag	nízká	střední
Boter	ano	kpujcky, ctenari, linear, tituly, svazky	vysoká (nevhodné pro ak. knihovnu)	vysoká
<b>Anal. dat ze soc. sítí</b>				
Sewell	ne (potřeba ext. dat)	X	X	X

nebo nedostatečné množství dat (v případě, že knihovna funkci nepoužívá). Analyzovatelné tabulky<sup>84</sup> lze rozdělit do těchto okruhů: data o uživateli a data o fondu. Všechna ostatní data jsou buď nedůležitá (definice), nebo jsou nedostatečná pro pokrytí problému (například nejsou propojitelná s dalšími tabulkami).

Data o fondu reprezentují bibliografické záznamy knih, autority, statistiky Z39.50,<sup>85</sup> výpůjčky, svazky a tituly, úbytky a ztracené knihy. Data o uživateli reprezentují aktivity v OPACu (zobrazování katalogu, přihlašování, úspěšné vyhledávání, označování, nastavování), aktivity fyzické (příchod, výpůjčka, poplatky) a informace o čtenáři.

#### 3.5.1 Vyhodnocení – okruhy problémů

UKSU donedávna neměla vlastní rozpočet (protože byla původně rozdělena a spadala pod jednotlivé součásti), proto nelze provádět analýzy fondu se zacílením na finanční stránku.<sup>86</sup> Nákupy publikací si obstarávaly jednotlivé ústavy samy, knihovna fond pouze zpracovávala a evidovala. To je také důvod, proč knihovna nevyužívá akvizičního modulu.<sup>87</sup>

Co se týká činností v knihovně a jejich evidence z hlediska knihovníků, systém samozřejmě registruje počet zpracovaných dokumentů v daném dni, počet MVS nebo počet výpůjček, není však stavěn například na sledování zátěže pracoviště. V tomto případě se jedná o malou knihovnu s malým počtem pracovníků. Jejich síly jsou napnuté mezi řadu pracovišť a řadu různých činností. Díky tomu mají o činnostech v knihovně a práci knihovníků daleko lepší představu než systém Clavius jednotliví vedoucí.

Sledování vytíženosti systému na datech z Clavia má smysl jenom vzhledem k činnostem uživatelů. Program sice ukládá zápis o aktivitách v systému do tabulky logů, do souboru na disk a do tabulky s chybami, ale k výpisům neexistuje dokumentace nutná pro další analýzu a logy se pravidelně mažou. Bez přímé pomoci firmy Lanisus (která systém vyvíjí a prodává) nepovažujeme provedení analýzy vytíženosti systému na datech, které generuje systém sám, za možnou. Smysl by měla pouze analýza vytíženosti programu uživateli; protože je jejich činnost zaznamenaná, poměrně dobře čitelná a systém její záznamy nemaže. Dle

<sup>84</sup>Podotýkáme, že mluvíme pouze o datech UKSU. Data knihovny, využívající jiné moduly, budou rozdílná.

<sup>85</sup>„Protokol pro vyhledávání informací v rámci knihovních aplikací, standard NISO (National Information Standards Organisation, USA). Je určen speciálně pro podporu vyhledávání z distribuovaných serverů. V architektuře OSI je na úrovni aplikační vrstvy. Umožňuje provádět z aplikace na jednom počítači dotazy do databáze na jiném počítači. V současnosti protokol oficiálně zajišťuje Library of Congress (USA).“ (CELBOVÁ, Ludmila. Z39.50. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)*. Praha: Národní knihovna ČR, 2003–. Dostupný z: ([http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000575&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000575&local_base=KTD)))

<sup>86</sup>Nemělo by například smysl zjišťovat v jakém období utratila knihovna určité množství peněz, vztaženo třeba na typy dokumentů.

<sup>87</sup>Finance knihovny a všech jejích pracovišť jsou vedeny zcela mimo systém Clavius.



jejich výsledků by bylo možné plánovat například údržbu systému a nastavovat pravidelné akce, které Clavius spouští. Také by podle nich mohla upravovat politiku knihovny.

Dále například analýza umístění fondu a výpůjček by mohla vedení knihovny přinést lepší představu o tom, jak jsou její dokumenty využívány, vůči tomu jak jsou rozmístěny. Pomohla by je rozčlenit dle půjčovnosti do skladu a do volného výběru. S provedením DM bychom však doporučovali počkat, dokud se knihovna nepřestěhuje do nových prostor.<sup>88</sup>

Jednou z možností, jak využít data o fondu a data o činnosti uživatelů je jejich vyhodnocení pro úpravu politiky knihovny. Například Stephen Zweibel a Zachary Lane provedli ve svém článku DM, aby ověřili, jak fungují změny ve výpůjční politice, které knihovna provedla. Výsledky jejich průzkumu by bylo možné porovnat s výsledky UKSU.

Posledním, a nejdůležitějším okruhem, je reporting univerzitě. V tomto smyslu by bylo samozřejmě možné provést analýzy jak nad daty o fondu (pro přehled v tématech, stáří dokumentů, počtu exemplářů a popularitě), tak nad daty o uživateli knihovny (trendy s vazbou na fakulty a činnosti). První analýza by přinesla hlavně přehled, a v budoucnu také základ pro formování akviziční politiky. Druhá pak ověření aktivity čtenářů vzhledem k provozní době, složení uživatelů a jejich chování. Výsledky by poté sloužily jako podklad pro rozhodování a zdůvodňování rozhodnutí součástí univerzity.

#### 3.5.2 Vyhodnocení – Data mining nad daty z Clavia

Při provádění DM nad daty UKSU se budeme specializovat. Ze dvou oblastí, o kterých jsme si řekli, že Clavius shromažďuje dostatek údajů, si budeme muset vybrat pouze jednu. Vzhledem k situaci UKSU, problémům s prosazováním změn a nových rozhodnutí na celouniverzitní úrovni, si pro analýzu<sup>89</sup> vybereme data spojená s chováním uživatelů. Od výsledků si slibujeme lepší popis současné situace a podklady pro prosazení změn v knihovně.<sup>90</sup>

Naším zaměřením je analyzovat chování uživatelů, ovšem analýza chování uživatelů jen na datech ze systému Clavius nebude jednoduchá. Žádná jiná data než data knihovního softwaru Clavius k dispozici nemáme. Clavius neumožňuje analyzovat transakční logy z OPACu a neukládá záznamy dalších interakcí, které uživatelé s knihovnou či s informačními zdroji mají (jako například přístup k EIZ, stahování článků, využívání referenčních služeb, účast na informačních lekcích atd.). Naši pozornost proto budeme směřovat k analýze takových akcí uživatelů, které systém sleduje, a kterých by bylo možné využít k podpoře managementu.

---

<sup>88</sup>Jedno stěhování navíc by momentálně nemělo smysl.

<sup>89</sup>Analýzy budou provedeny podle metodiky CRISP-DM, kterou jsme jmenovali v úvodu práce.

<sup>90</sup>Navázaných zvláště na fakulty.

### 3 Analýza dat z knihovního softwaru Clavius

Proto bychom měli specifikovat, jaké údaje má vedení knihovny vlastně k dispozici. Víme, že Clavius základní statistiky poskytuje. Zpracování statistik se nachází ve Výpůjčním modulu,<sup>91</sup> přičemž je v něm možné generovat rovnou tiskové sestavy.

Databázové dotazy, týkající se transakcí jsou schopny zjistit měsíční statistiky výpůjček (absenčních a prezenčních) obecně, viz Obrázek č. 3.1, dále výpůjčky ve dnech v měsíci, viz Obrázek č. 3.2,<sup>92</sup> statistiku výpůjček podle jazyků, statistiku MVS, viz Obrázek č. 3.3, statistiku návštěv ve dnech v týdnu, viz Obrázek č. 3.4 a podobně vypadající rozložení výpůjček a vracení knih do dnů v týdnu v konkrétním měsíci. V programu lze také zhodnotit zvláště elektronické výpůjčky.

**Statistický výkaz za měsíc 1/2016**

Celkový počet čtenářů registrovaných ve sledovaném období : 0

Počet návštěvníků : 2 , z toho registrační poplatky zaplatilo : 0

Počet výpůjček CELKEM :	25 - z toho prolongací : 22
Počet výpůjček knih :	21 , tj. 84.00 % z celkového počtu
Počet výpůjček periodik :	3 , tj. 12.00 % z celkového počtu
Počet výpůjček ostatních :	1 , tj. 4.00 % z celkového počtu

Počet výpůjček na 1 návštěvníka : 12.50

Vyřízených rezervací : 0 , registrovaných informací : 0

Složení výpůjček podle tematického zaměření ( pouze knihy )			
druh	počet	% z celku	% skupiny
Beletrie půjčená dospělým :	11 , tj. 44.00 %	, tj. 52.38 %	
Naučná literatura dospělým :	10 , tj. 40.00 %	, tj. 47.62 %	
Celkem půjčeno dospělým :	21 , tj. 84.00 %	, tj. 100.00 %	
Beletrie půjčená dětem :	0 , tj. 0.00 %	, tj. 0.00 %	
Naučná literatura dětem :	0 , tj. 0.00 %	, tj. 0.00 %	
Celkem půjčeno dětem :	0 , tj. 0.00 %	, tj. 100.00 %	

Částka vybraná za registrace ve sledovaném období : 0.00 Kč

Částka vybraná za upomínky ve sledovaném období : 0.00 Kč

Obrázek 3.1: Vzorová statistika výpůjček v měsíci z programu Clavius.<sup>93</sup>

Čtenáře lze ve statistikách z konkrétního období rozdělit dle věku (který se počítá z rodného čísla), viz Obrázek č. 3.5; fakult a kategorií (rozdělení vypadá stejně) a dle stavu registrace (dobíhající registrace v daném období). Také lze

<sup>91</sup>Lanius. *Clavius*. c2016. [software]. verze 09.02.2016. Tábor.

<sup>92</sup>Je také možné vygenerovat variantu pro audio-vizuální média, nebo celkový souhrn všech druhů dokumentů.

<sup>93</sup>Lanius, c2016.

<sup>94</sup>Lanius, c2016.

<sup>95</sup>Lanius, c2016.

<sup>96</sup>Lanius, c2016.

### 3 Analýza dat z knihovního softwaru Clavius

**Statistický deník za měsíc 2/2015**

Datum	Regist- rování členů	z toho členů do 15 let	Počet návště- vníků	Výpůjčky celkem	naučná literatura dospělým	krásná literatura dospělým	naučná literatura dětím	krásná literatura dětím	periodika a ostatní	z toho přičten- gací
01.02. ne	0	0	0	0	0	0	0	0	0	0
02.02. po	0	0	0	0	0	0	0	0	0	0
03.02. út	0	0	0	0	0	0	0	0	0	0
04.02. st	0	0	0	0	0	0	0	0	0	0
05.02. čt	0	0	0	0	0	0	0	0	0	0
06.02. pá	0	0	0	0	0	0	0	0	0	0
07.02. so	0	0	0	0	0	0	0	0	0	0
08.02. ne	0	0	0	0	0	0	0	0	0	0
09.02. po	0	0	0	0	0	0	0	0	0	0
10.02. út	0	0	0	0	0	0	0	0	0	0
11.02. st	0	0	0	0	0	0	0	0	0	0
12.02. čt	0	0	0	0	0	0	0	0	0	0
13.02. pá	0	0	0	0	0	0	0	0	0	0
14.02. so	0	0	0	0	0	0	0	0	0	0
15.02. ne	0	0	0	0	0	0	0	0	0	0
16.02. po	0	0	0	0	0	0	0	0	0	0
17.02. út	0	0	0	0	0	0	0	0	0	0
18.02. st	0	0	0	0	0	0	0	0	0	0
19.02. čt	0	0	0	0	0	0	0	0	0	0
20.02. pá	0	0	0	0	0	0	0	0	0	0
21.02. so	0	0	0	0	0	0	0	0	0	0
22.02. ne	0	0	0	0	0	0	0	0	0	0
23.02. po	0	0	0	0	0	0	0	0	0	0
24.02. út	0	0	0	0	0	0	0	0	0	0
25.02. st	0	0	0	0	0	0	0	0	0	0
26.02. čt	0	0	0	0	0	0	0	0	0	0
27.02. pá	0	0	0	0	0	0	0	0	0	0
28.02. so	0	0	0	0	0	0	0	0	0	0
<b>Součet</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Obrázek 3.2: Vzorová statistika výpůjček ve dnech v měsíci z programu Clavius.<sup>94</sup>

#### **Statistika MVS od 01.02.2016 do 29.02.2016**

Požadavky MVS obdržené z jiných knihoven :

Celkový počet obdržených požadavků : 0

Celkový počet kladně vyřízených požadavků : 0

- z toho počet požadavků vyřízených výpůjčkou originálu : 0

- z toho počet požadavků vyřízených zasláním kopie : 0

Požadavky MVS zasláné do jiných knihoven :

Celkový počet odeslaných požadavků : 0

Celkový počet kladně vyřízených požadavků : 0

- z toho počet požadavků vyřízených výpůjčkou originálu : 0

- z toho počet požadavků vyřízených zasláním kopie : 0

Obrázek 3.3: Vzorová statistika výpůjček MVS na měsíc z programu Clavius.<sup>95</sup>

### 3 Analýza dat z knihovního softwaru Clavius

**Časové rozložení návštěv za měsíc 1/2016**

Hodina	Pondělí	Úterý	Středa	Čtvrtek	Pátek	Sobota	Neděle	Celkem
do 08:00	0	0	0	0	0	2	0	2
Celé dny	0	0	0	0	0	2	0	2

Obrázek 3.4: Vzorová statistika návštěv ve dnech v týdnu za měsíc z programu Clavius.<sup>96</sup>

Tabulka 3.6: Výpůjčky UKSU v akademických letech 2011/2012 až 2014/2015.

Akademický rok	2011/2012	2012/2013	2013/2014	2014/2015
Výpůjčky dle zpráv	68 165	65 764	37 019	údaj chybí
Výpůjčky dle dat	32 336	36 238	31 798	33 858

vygenerovat přehled návštěv čtenářů v konkrétním období, viz Obrázek č. 3.6, a obdobně výpis návštěv anonymních čtenářů.

Poslední statistické výpisy se týkají „rekordmanů“, tj. nejžádanějších dokumentů (výpůjčky, rezervace) v daném období, čtenářů s nejvíce výpůjčkami v daném období a seznam čtenářů s momentálně vypůjčenými dokumenty, viz Obrázek č. 3.7. Úroveň podrobnosti a řazení může být různá.

#### 3.5.3 Problémy při počítání statistik

V letech 2011 až 2014 uváděla UKSU ve Výročních zprávách univerzity<sup>100</sup> také údaje za aktivitu knihovny. Když jsme chtěli porovnat tyto statistiky s našimi výsledky, (počet uživatelů a počet absenčních výpůjček) narazili jsme na problémy. Naše čísla se dramaticky lišila od těch, které se nacházejí ve výročních zprávách. Vezměme si kupříkladu počet výpůjček, viz Tabulka č. 3.6.

Z propadu uvedeném ve výroční zprávě v roce 2013/2014 by se mohlo zdát, že knihovna prodělává velký pokles ve výkonnosti nebo v počtu čtenářů. Počet potenciálních uživatelů opravdu postupně klesá, viz Tabulka č. 3.3, ovšem pokud

<sup>97</sup>Lanius, c2016.

<sup>98</sup>Lanius, c2016.

<sup>99</sup>Lanius, c2016.

<sup>100</sup> *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2011* [online]. Slezská univerzita v Opavě, 2012, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-záměr/2011/vz-cinnost-2011.pdf>); *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2012* [online]. Slezská univerzita v Opavě, 2013, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-záměr/2012/vz-cinnost-2012.pdf>); *Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2013* [online]. Slezská univerzita v Opavě, 2014, [cit. 2015-06-15]. Dostupný z: (<http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodobý-záměr/2013/vz-cinnost-2013.pdf>).

### **Statistika čtenářů od počátku roku 2016**

Celkový počet návštěv čtenářů ve sledovaném období : 2

Věkové rozložení čtenářů		
rozmezí	počet	%
do 15-ti let -	0	0.00 %
16 až 19 let -	0	0.00 %
20 až 29 let -	0	0.00 %
30 až 39 let -	0	0.00 %
40 až 49 let -	1	50.00 %
50 až 59 let -	0	0.00 %
60 až 69 let -	0	0.00 %
70 až 79 let -	0	0.00 %
nad 80 let -	1	50.00 %
nevyplněno -	0	0.00 %

Pohlaví	počet	%
Muži -	2	100.00 %
Ženy -	0	0.00 %

Poplatek	počet	%
Platí -	2	100.00 %
Zdarma -	0	0.00 %

Bydliště	počet	%
Místní -	2	100.00 %
Ostatní -	0	0.00 %

Obrázek 3.5: Vzorová věková statistika čtenářů z programu Clavius.<sup>97</sup>

#### **Přehled návštěv čtenářů včetně časů od počátku roku 2016**

Odd.	Čas příchodu	Čas odchodu	Evidenční číslo	Jméno čtenáře	Datum narození
D	23.01.2016 06:37:14		1	Novák Antonín	16.12.1971
D	23.01.2016 06:37:52		2	Klůfa Petr	11.11.1911

Počet čtenářů v tomto seznamu : 2

Obrázek 3.6: Vzorová statistika návštěv z programu Clavius.<sup>98</sup>

### 3 Analýza dat z knihovního softwaru Clavius

#### Seznam čtenářů s aktuálními výpůjčkami

Návštěva	Odd	Číslo	Jméno čtenáře	Upomínka	Výpůjček
18.05.11	D	9	Deváták Stanislav	1. upom.	5 ks
05.03.13	D	7	Sedmý Lukáš Mgr.	1. upom.	3 ks
23.08.13	D	6	Havlíčkův Brod KKV	4. upom.	2 ks
19.06.15	D	8	Osmá Kateřina	4. upom.	4 ks
23.01.16	D	1	Novák Antonín	1. upom.	8 ks
23.01.16	D	2	Klůfa Petr	1. upom.	7 ks

Celkem čtenářů v tomto seznamu : 6

Obrázek 3.7: Vzorová statistika seznamu čtenářů s vypůjčenými dokumenty z programu Clavius.<sup>99</sup>

se podíváme na počty výpůjček v datech, propady nejsou tak dramatické, dokonce v posledním roce počet výpůjček stoupl. Údaje jsou zkreslovány tím, že se do celkového čísla ve statistikách Clavia počítají také prolongace,<sup>101</sup> rezervace, statistiky MVS atd.<sup>102</sup> Z těchto čísel si knihovna může nechat přímo systémem odstranit pouze statistiky MVS. Pro získání čistého počtu výpůjček by byla potřeba další práce s daty.<sup>103</sup>

Dalším problémem je počet čtenářů, které Clavius počítá vždy pouze z těch, kteří mají zaplacenou registraci na daný rok. Samozřejmě, zaměstnanci univerzity poplatky neplatí. Registraci čtenář platí vždy po jejím vypršení při prvním příchodu do knihovny, bez ní nemohou být čtenáři poskytovány služby. Zatím se nám může zdát, že je problém jasně definovaný a jeho řešení je správné. Ovšem knihovna nově zprostředkovává také elektronické knihy, tyto je možné nalézt přes OPAC, kam se mohou uživatelé přihlásit kdykoliv i bez zaplacení registrace (je to řešeno přes jednotné přihlašování školy). Knihu si poté mohou půjčit v rozhraní Ebrary. Teoreticky tedy platí, že pokud uživatel využívá pouze OPACu a půjčuje si knihy z Ebrary, není počítán mezi čtenáře knihovny. Se zvyšujícím se zájmem o elektronické dokumenty se bude knihovna muset v budoucnu tímto problémem zabývat.<sup>104</sup>

Z uvedených vzorových příkladů statistik vyplývá, že knihovna je schopna pomocí statistik zjistit základní čísla o svém provozu, ale není schopna vidět poměry a srovnávat. Je možné zjistit, jaký poměr čtenářů je z Filozoficko-přírodovědecké fakulty (dále jen FPF) a jaký je z Fakulty veřejných politik (dále jen FVP), ale

<sup>101</sup>System má zvláštní značení prolongací a někdy je v datech označuje jako výpůjčky. Pro přesnější výsledky je třeba tato data odfiltrovat.

<sup>102</sup>Výpůjční doba u prezenčních studentů je 30 dní, u distančních je to 45 a prolongaci mohou provést osobně nebo online přes OPAC. U vyučujících ale platí jiná pravidla a prolongace se provádějí automaticky.

<sup>103</sup>Ideální by byl dotaz SQL do databáze Clavia.

<sup>104</sup>Podobně jako monitoringem přístupů do EIZ.

není schopna například zjistit, jaký je poměr mezi kategoriemi u fakult, nebo jaké jsou vzorce chování, které se v datech opakují.

#### 3.5.4 Analýzy – rozložení uživatelů

Aby knihovna získala lepší pozici a také měla lepší přehled, bylo by pro ni výhodné vědět, v jakém poměru využívají knihovnu prezenční či distanční studenti, zaměstnanci, a ze kterých fakult pocházejí. Bohužel, tabulka s daty uživatelů nezaznamenává údaje o tom, kdy se člověk stal čtenářem knihovny, ani kdy svůj statut obnovil. Aktuální počet čtenářů knihovny se u statistik počítá z tabulky, která eviduje platby. Podle vybraného období se pak spočítá, kdo měl v danou dobu zaplacený čtenářský poplatek. Vyučující ale čtenářský poplatek neplatí, jsou tedy za uživatele bráni automaticky. Znamená to ale, že knihovnu aktivně využívají?

Aby naše tvrzení co nejvíce odpovídala skutečnosti, definujeme pojem aktivní uživatel; za aktivního uživatele považujeme každého uživatele, který si vypůjčil knihu, nebo se přihlásil do OPACu<sup>105</sup> (přihlášení je vyžadováno například u objednávání MVS, rezervace knihy, sestavování seznamu literatury, atd.). Z tabulky uživatelů (tato obsahuje všechny studenty a vyučující na SU díky systému jednotného přihlášení) jsme vybrali všechny, kteří si půjčili knihu, nebo se přihlásili. Celkový přehled je možné najít na Obrázku č. 3.8. Při zpracování jsme z dat vynechali pracovníky rektorátu, veřejnost (celkem 32 uživatelů) a uživatele z Matematického ústavu (celkem 78). Jejich záznamy jsou totiž v tabulce uživatelů uloženy nekonzistentně.

Na Obrázku č. 3.8 vidíme, že největší počet aktivních uživatelů měla knihovna z Obchodně-podnikatelské fakulty (dále jen OPF) – a to jak z hlediska prezenčních, tak z hlediska distančních<sup>106</sup> studentů. Zajímavé také je, že ze zaměstnanců jsou nejvíce aktivní zaměstnanci Filozoficko-přírodovědecké fakulty (dále jen FPF). Poměr prezenčních a distančních studentů Fakulty veřejných politik (neboli FVP) je poměrně vyrovnaný. Nejvíce aktivních doktorandů je z Obchodně podnikatelské fakulty.

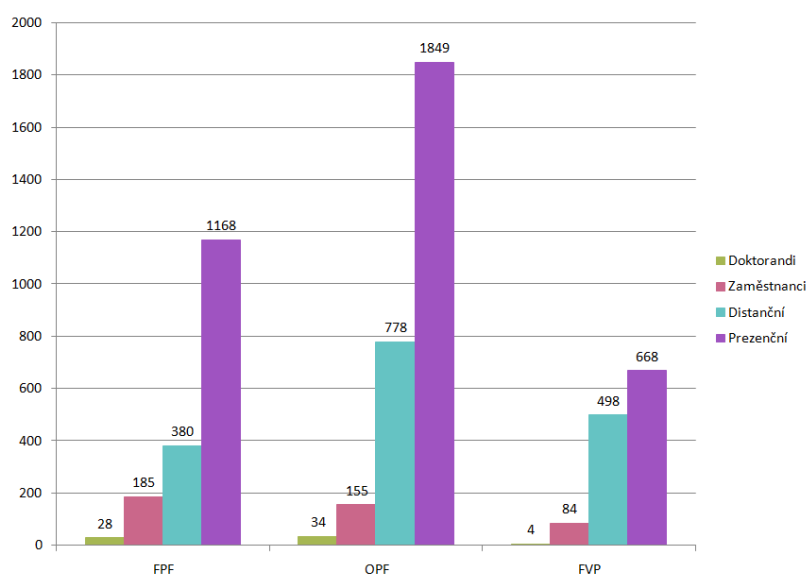
Tyto výsledky však nezohledňují velikost fakult a počty jejich studentů. Dáme-li počet studentů fakulty do poměru s počtem aktivních uživatelů, viz Obrázek č. 3.9, zjistíme například, že ačkoliv OPF vedla v počtu studentů, v procentuálním přepočtu ji u prezenčních studentů předběhla FVP. Počet distančních studentů OPF v grafu na Obrázku č. 3.8 byl také vyšší než u všech ostatních fakult. V poměru k celkovému počtu se však jejich počet pohybuje na velice podobné

---

<sup>105</sup>Návštěva u nás není brána jako aktivita a to proto, že systém ukládá pouze datum poslední návštěvy uživatele. Statistiky návštěvnosti generuje z logů, které nejsou zpracovatelné bez pomoci firmy Lanisus.

<sup>106</sup>Jako distanční studenti jsou systémem označováni všichni studenti, kteří nestudují na SU v prezenčním studiu, tj. distanční a kombinovaní.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.8: Celkový počet aktivních uživatelů na kategorii a fakultu.

úrovni s ostatními fakultami. Zajímavý je vyšší poměr aktivních doktorandů OPF k celkovému počtu, který nebyl v předchozím grafu tak patrný. Zajímavý je také nízký poměr využívání knihovny zaměstnanci FVP.

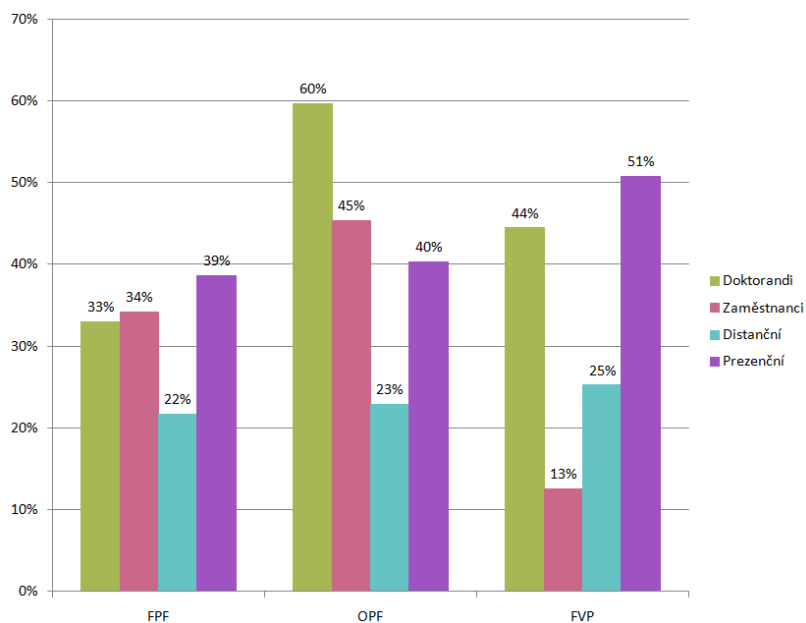
Průměrná hodnota využívání knihovny k celkovému počtu studentů a zaměstnanců se pohybuje kolem 33 %. To znamená, že celkově využívá knihovnu 33 % z potenciálních uživatelů všech fakult dohromady. Průměrná hodnota pro jednotlivé fakulty je: FPF 33 %, FVP 32 % a OPF 34 %. Průměr je u všech fakult velice vyrovnaný. Pokud bychom rozložili tyto analýzy do jednotlivých let, zjistíme, že se rozložení dat v jednotlivých akademických letech nemění.

Tyto informace poskytují knihovně poměrně dobrý přehled o tom, kdo jsou její čtenáři a do jaké míry se jí daří oslovovat jednotlivé kategorie uživatelů na fakultách.

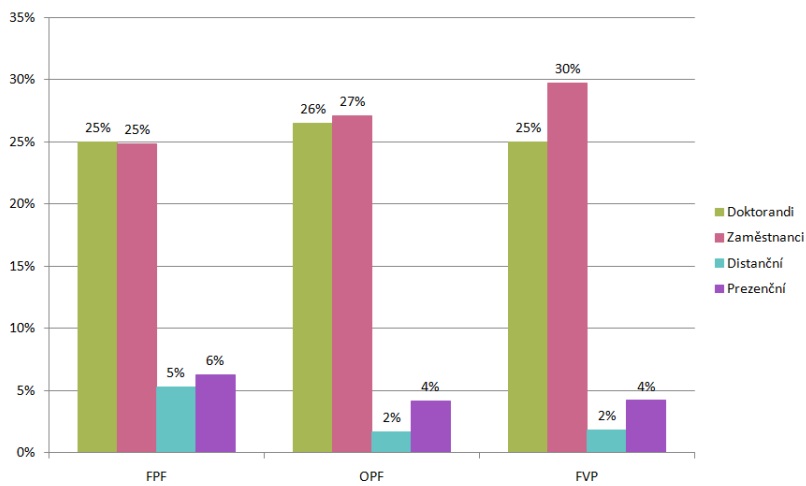
Dalším problémem, který by knihovnu mohl zajímat, jsou stálí uživatelé. Za stálé uživatele považujeme ty, kteří byli aktivní po celé sledované období (čtyři roky). Existuje předpoklad, že by mezi stálé uživatele měli patřit hlavně doktorandi a zaměstnanci, protože se na univerzitě nacházejí delší dobu. Na Obrázku č. 3.10 je patrné, že tento předpoklad byl správný. Pokud celkový počet 350 stálých uživatelů rozložíme mezi fakulty a kategorie, pak zaměstnanci tvoří podíl cca 27 % a doktorandi přibližně 25 %. To znamená, skoro každý třetí doktorand či zaměstnanec na fakultě, který využívá knihovnu, ji využívá dlouhodobě. Naopak, ze všech studentů, využívajících knihovnu (ze všech fakult i typů studia), je dlouhodobých uživatelů pouhých 6 %.



### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.9: Poměr aktivních uživatelů k velikosti fakulty a ke kategorii.

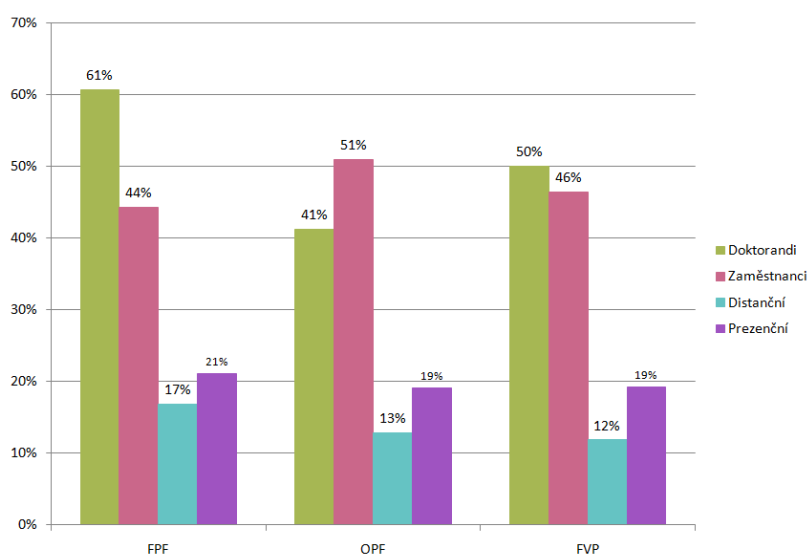


Obrázek 3.10: Poměr stálých uživatelů z jednotlivých fakult na kategorii.

### 3 Analýza dat z knihovního softwaru Clavius

Protože doporučená doba studia u bakalářských oborů bývá tři roky (ne čtyři, které jsme sledovali my), rozhodli jsme se vysledovat aktivitu uživatelů ve třech letech, abychom zjistili, jak moc se změnil poměr studentů. Jinými slovy, chtěli jsme zjistit, jaký poměr z aktivních studentů chodí do knihovny po tři roky. Výsledný graf naleznete na Obrázku č. 3.11.

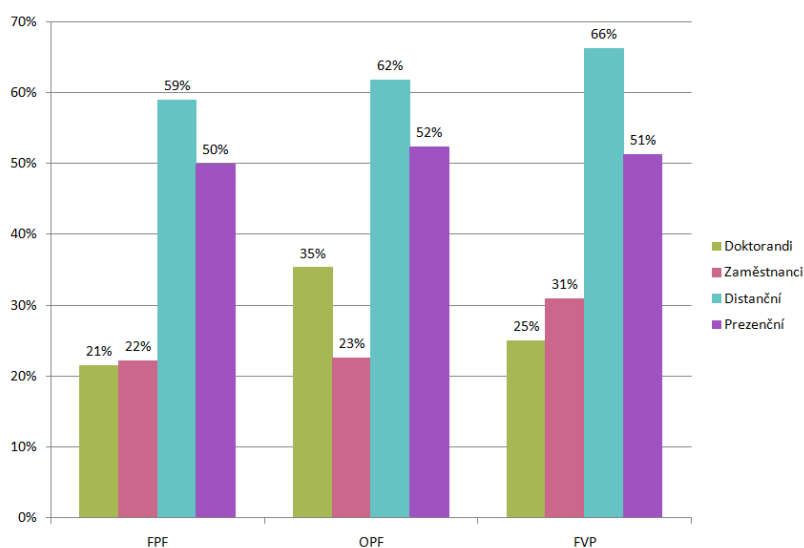
Nejvýrazněji se proměnil počet doktorandů, pak poměr zaměstnanců a až na třetím místě poměr prezenčních studentů. Poměr dálkových studentů se změnil nejméně. Z výsledků vyplývá, že průměrně polovina ze všech aktivních doktorandů využívá knihovnu tři a více let, dále průměrně 20 % z aktivních prezenčních studentů využívá knihovnu tři a více let; asi 47 % z aktivních zaměstnanců využívá knihovnu tři a více let, a pouze 13 % z aktivních distančních studentů využívá knihovnu tři a více let. Kromě vyššího počtu aktivních doktorandů z FPF, kteří se od jiných doktorandů liší až o 11 %, není mezi fakultami významný rozdíl.



Obrázek 3.11: Poměr uživatelů aktivních po tři roky z jednotlivých fakult a v kategoriích.

Dalším zajímavým údajem pro knihovnu by byla zpráva o tom, jak uživatelé odpadají po jednom roce využívání knihovny (příčemž pro záznam v jednom roce stačí, když se čtenář jednou přihlásí do OPACu nebo si půjčí knihu). Na Obrázku č. 3.12 je vidět, jak tato analýza dopadla. Po jednom roce přestane být uživateli knihovny průměrně 29 % aktivních doktorandů (na OPF ještě více), dále 29 % aktivních prezenčních studentů přestane být uživateli knihovny po jednom roce, a 63 % aktivních distančních studentů přestane být uživateli knihovny po jednom roce. Průměrná hodnota pro zaměstnance je 24 % přičemž o něco hůře jsou na tom zaměstnanci FVP.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.12: Poměr uživatelů aktivních pouze jeden rok z jednotlivých fakult a v kategoriích.

Kromě trendů aktivních uživatelů v jednotlivých letech můžeme na grafech pozorovat podobné rysy ve využívání knihovny fakultami. Až na výjimky a lokální odlišnosti využívají všechny fakulty knihovnu podobným způsobem. Z toho plyne ujištění, že pro budoucí jednání o finanční podpoře knihovna může požádat o příspěvek rovným dílem každou z fakult. Nelze argumentovat, že studenti nebo vyučující z dané fakulty knihovny nevyužívají nebo jí využívají méně než jiné fakulty.

Také je možné zjistit, jaké je složení uživatelů, každoročně využívajících knihovních služeb. Jednotlivé skupiny se rozdělí dle chování; následně se vyhodnotí rozdíly mezi aktivitou vypůjčování a přihlašování do OPACu. Obě aktivity ukážeme ve dvou oddělených grafech.

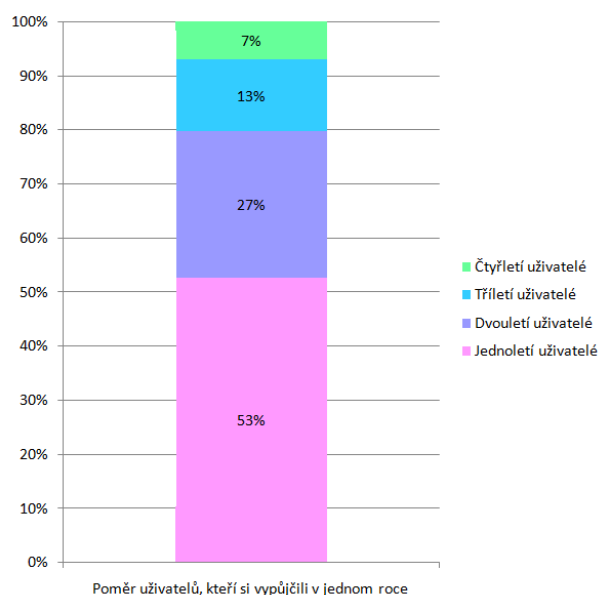
Nejdříve se podíváme na aktivity ve vypůjčování. Celkem si v knihovně UKSU za čtyři roky vypůjčilo minimálně jednu knihu 5 278 uživatelů. Z těchto uživatelů více jak 50 % knihovnu využívalo pouze v jednom roce.<sup>107</sup> Celých 27 % uživatelů si knihy půjčovalo po dva roky.<sup>108</sup> Dále 13 % symbolizuje uživatele, kteří si dokumenty knihovny půjčovali po tři roky<sup>109</sup> a kolem 7 % uživatelů si knihy vypůjčovalo pravidelně; toto rozložení vyjadřuje Obrázek č. 3.13.

<sup>107</sup>Samozřejmě toto číslo zahrnuje studenty, kteří studují po jednom roce zanechali.

<sup>108</sup>Tato hodnota v sobě zahrnuje také studenty magisterských studií, jejichž studium trvá dva roky.

<sup>109</sup>Číslo zahrnuje studenty, kteří studovali na SU bakalářské studium, které trvá tři roky.

### 3 Analýza dat z knihovního softwaru Clavius



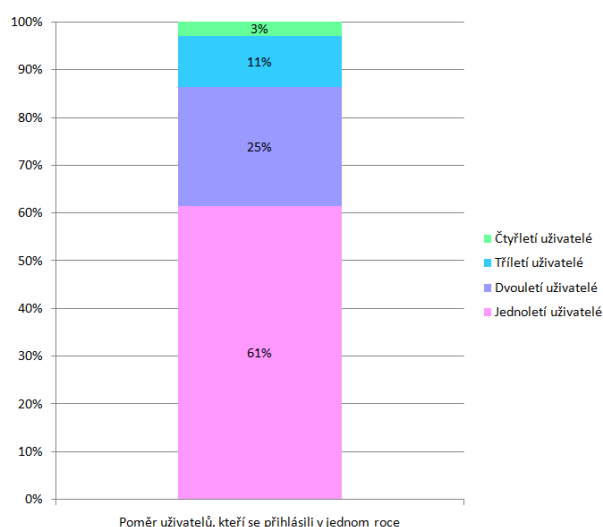
Obrázek 3.13: Poměr typů uživatelů, vypůjčujících si dokumenty v jednom roce.

Z výsledků vyplývá, že má knihovna vždy poměrně omezenou dobu na to, aby své uživatele oslovila. Jestliže této šance nevyužije, uživatel se už nemusí vrátit. Tomu by měla být přizpůsobena strategie služeb. Podobně z výsledků vyplývá, že po dvou letech se promění 80 % uživatelů knihovny. Cirkulace uživatelů je tedy poměrně vysoká, což klade nároky i na informační lekce.

Dále se podíváme na rozložení čtenářů, přihlašujících se do OPACu po celé sledované období. Poměr jednotlivých typů vyjadřuje Obrázek č. 3.14. Za sledované období se do OPACu alespoň jednou přihlásilo 4 355 uživatelů. Oproti předchozí skupině (výpůjčky dokumentů) je číslo nižší; navíc se skupiny zcela nekryjí. To znamená, že uživatelé, kteří se přihlásili, si nemuseli nic půjčit (a obráceně). Asi 61 % uživatelů využívalo OPAC knihovny pouze jeden rok (hodnota zahrnuje i jednou přihlášené uživatele). Pouze 25 % využívalo OPAC dva roky; 11 % uživatelů tři roky, a pouhé 3 % celé čtyři roky. Po dvou letech se v přihlašování do OPACu obmění 86 % uživatelů.

Po rozčlenění uživatelů na skupiny se podíváme na ty stálé, využívající katalog či výpůjční služby po celé sledované období. Existuje předpoklad, že by v poměru k celkovému počtu v obojím měli vést zaměstnanci nebo doktorandi. Nejprve se podíváme na výpůjčky, jejichž zhodnocení nalezneme na Obrázku č. 3.15. Z něj je možné vyčíst, že největší procento tvoří opět prezenční studenti, pak zaměstnanci, distanční studenti a až nakonec doktorandi. Vzhledem k celkovému počtu studentů na univerzitě jsou procenta spíše orientační. Skupina má asi 350 jednotlivců;

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.14: Poměr typů uživatelů v jednom roce, kteří se přihlásili do OPACu.

přičemž víme, že zaměstnanců je na univerzitě cca 600; z toho akademických pracovníků je asi polovina. Třicet dva procent zaměstnanců z 350 znamená, že každý šestý zaměstnanec univerzity si chodil do knihovny půjčovat dokumenty po celé sledované období. Ostatní hodnoty, vzhledem k celkovým počtům, nejsou překvapivé.

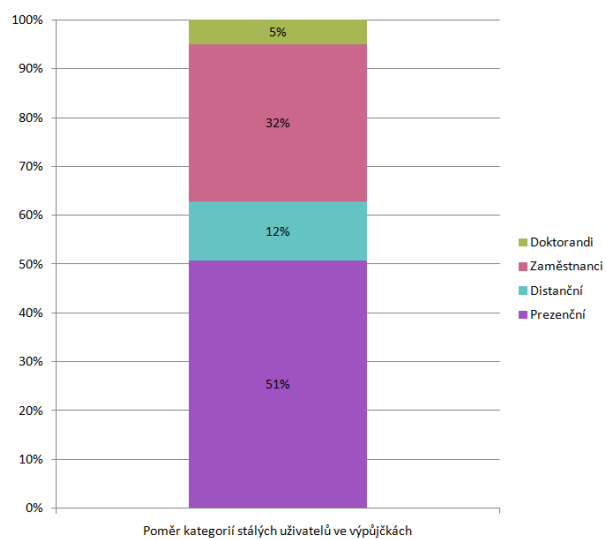
Nyní se podíváme, jak jsou na tom stálí uživatelé vzhledem k přihlašování do OPACu. Zhodnocení vidíme na Obrázku č. 3.16. Poměrně drasticky se snížilo procento zaměstnanců, kteří se do online katalogu přihlašovali po celé čtyři roky. Poměr doktorandů se naopak výrazně nezměnil. Z uvedeného vyplývá, že distanční a prezenční studenti jsou k přihlašování do OPACu ochotnější. Bylo by dobré zjistit například dotazníkovým průzkumem, jaký je postoj k OPACu ze strany zaměstnanců. Může jít o nezáměr nebo o nízkou informovanost.<sup>110</sup> Nebo je pro zaměstnance prostě jednodušší do knihovny zajít osobně.

Pokud bychom chtěli zjistit, zda využívají knihovnu spíše muži, nebo ženy, snažili bychom se nejprve vztáhnout data k příslušnosti k fakultám. To proto, že na různých fakultách studuje různé množství žen a mužů v závislosti na oborech (logicky by měla mít větší počet mužů fakulta FPF, se studijními obory fyzika a informatika, oproti fakultě FVP, kde lze studovat ošetřovatelství a sociálně zaměřené disciplíny). Poměr počtu mužů a žen na fakultách celkem je zřejmý na Obrázku č. 3.17.

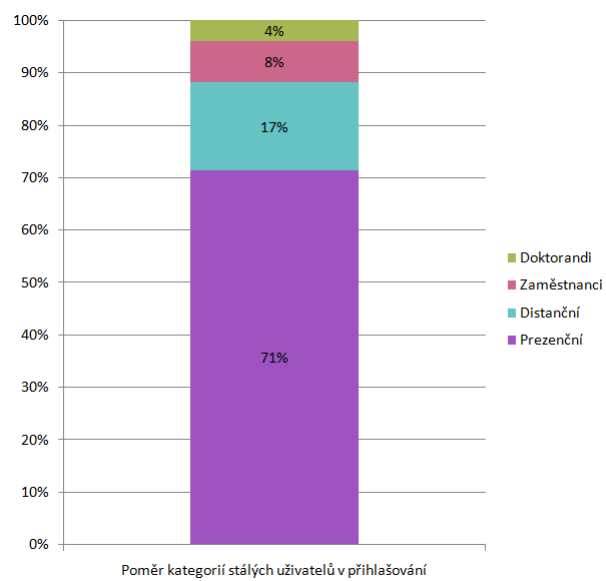
Poměr ve prospěch žen je zřetelný; nejvyváženější poměr má FPF, o něco větší procento žen se nachází na OPF, nejvyšší poměr žen k mužům má FVP.

<sup>110</sup>Na rozdíl od studentů zaměstnanci žádné informační lekce neabsolvují.

### 3 Analýza dat z knihovního softwaru Clavius

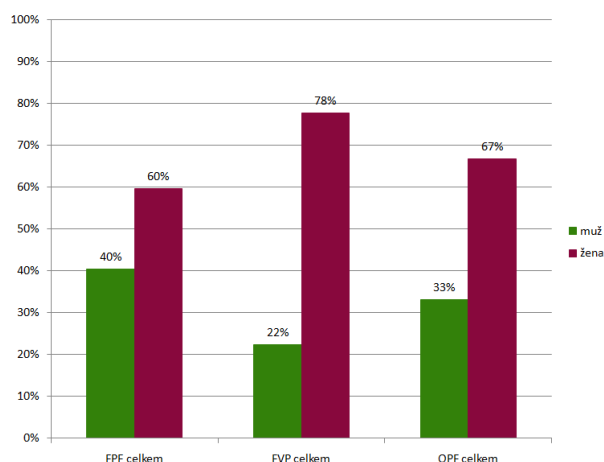


Obrázek 3.15: Poměr kategorií stálých uživatelů ve výpůjčkách.



Obrázek 3.16: Poměr kategorií stálých uživatelů v přihlašování do OPACu.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.17: Rozložení žen a mužů na fakultách.

Tyto informace však vzhledem ke studovaným oborům nejsou překvapivé. Obrázek č. 3.18 pak ukazuje poměr aktivních uživatelů podle pohlaví na fakultách vzhledem k potenciálním uživatelům. Z obrázku je patrné, že ženy na FPF jsou aktivnější; naopak na FVP je procento aktivních čtenářů zastoupeno převážně muži. Poměr aktivních žen a mužů na OPF je vyrovnaný.

Z analýzy vyplývá, že zájem mužů, studujících na FPF je oproti FVP nižší. Bylo by zajímavé prozkoumat, proč tomu tak je vzhledem k oborům, nabízeným na fakultách.

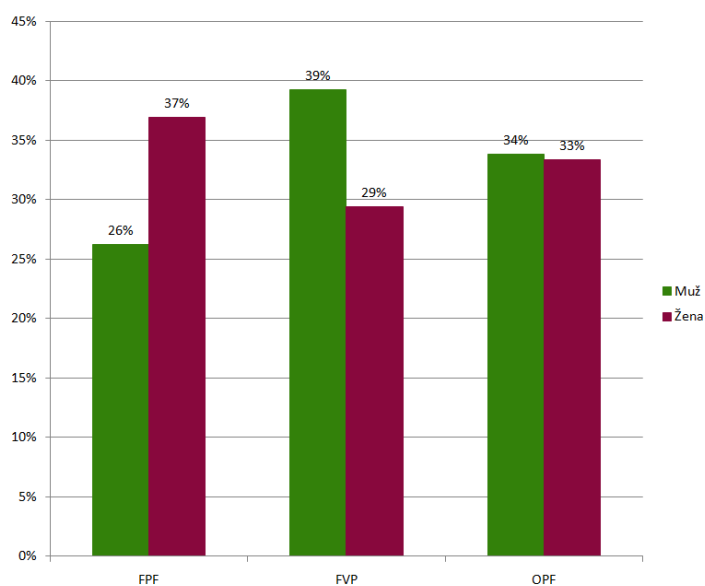
Smyslem této analýzy bylo co nejlépe demograficky popsat složení uživatelů, abychom věděli, kdo UKSU nejvíce využívá a v jakém poměru. Vynechali jsme zatím bližší popis aktivit; tomu se budeme věnovat v následující části.

#### 3.5.5 Analýzy – vzorce chování

Identifikace vzorců chování (nebo také trendů) nám pomůže vymezit pro knihovnu nejvíce a nejméně náročná období v roce, vytížení Clavia uživateli v konkrétních časech a ukáže vztahy, které nejsou na první pohled zřejmé. Budeme vycházet primárně z těchto aktivit – vypůjčování publikací a přihlašování do OPACu. Přístup na webové stránky budeme brát pouze informačně, protože identifikace přístupu se odehrává pouze pod IP adresou. Stejnou IP adresu může mít několik uživatelů, kteří mají stejného internetového poskytovatele. Pro nás je však důležité spojit aktivitu s konkrétním uživatelem, a proto analýzu přístupů do OPACu provádět nebudeme.

S uživatelským účtem lze vyhledávání v katalogu spojit pouze částečně. V tabulce `car_cloud` se nachází vyhledávané termíny, časový údaj a u přihlášených

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.18: Rozložení aktivních žen a mužů vztahený k celkovému počtu potenciálních uživatelů.

uživatelů také jejich čtenářské číslo, nicméně ne všichni uživatelé vyhledávají pod přihlášením. A dále když je vyhledávání neúspěšné, je ukládáno do tabulky `car_expression_not_found`. V této tabulce se nenalézá jediný identifikátor (kromě časového údaje), dle kterého by bylo možné spojit uživatele a jeho vyhledávání. Pokud v online katalogu vyhledává naráz více uživatelů, systém je dle této tabulky nerozliší. Není možné rozlišit přihlášeného a anonymního uživatele a pokud uživatel zadá do vyhledávání postupně více termínů s tím, že je časově posune třeba o pár minut, jakýkoli způsob spojení v datech je ztracen. Proto se vyhodnocením činnosti vyhledávání nebudeme primárně zabývat, protože je nejde jednoznačně identifikovat jako činnost konkrétního uživatele.

S uživatelským účtem na druhou stranu lze spojit aktivity označování publikací, sestavování seznamů literatury a nastavování uživatelského účtu. Bohužel tyto aktivity provozovalo poměrně málo uživatelů na to, aby se z nich dalo něco usuzovat. Údaje z nich budeme používat spíše pro doplnění.

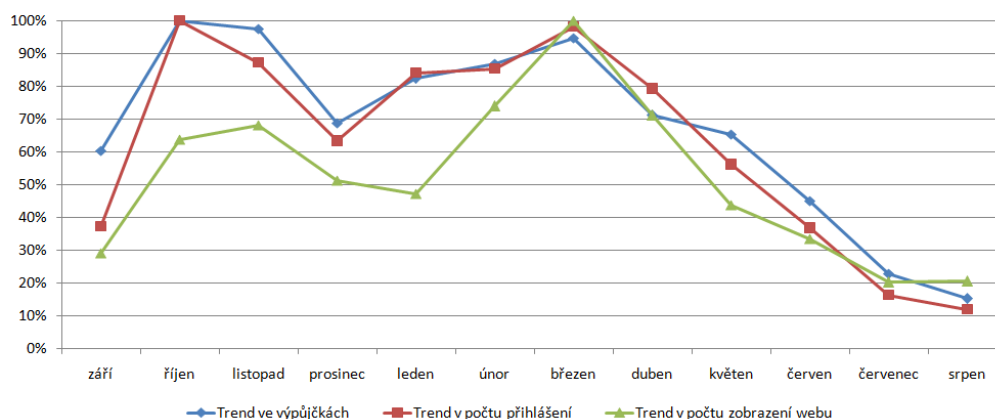
Nejdříve srovnáme trendy ve vypůjčování publikací, přihlašování a zobrazování webových stránek, abychom zjistili, jaký průběh tyto činnosti mají, a zda je u všech aktivit podobný. Srovnání je uvedeno na Obrázku č. 3.19, můžeme odvodit, že trend podobný je.<sup>111</sup> Nejvytíženějšími měsíci jsou říjen a listopad v zimním semestru, a pak březen v letním semestru. Trend ve vypůjčkách a trend v přihlášení mají

<sup>111</sup>Údaje v grafu byly normovány k maximu, aby byl lépe vidět průběh akademického roku.



### 3 Analýza dat z knihovního softwaru Clavius

velmi podobný průběh; zatímco trend v počtu zobrazení webu se v počátečních měsících liší.



Obrázek 3.19: Trendy ve vypůjčování publikací, přihlašování a zobrazování webových stránek.

Nejméně vytíženými měsíci v roce jsou červenec a srpen.<sup>112</sup> Pokud by knihovna plánovala změny v provozu, stěhování, školení personálu, velký softwarový update nebo něco podobného, bylo by dobré provést změny v těchto měsících.<sup>113</sup> Naopak nejširší zájem by u uživatelů vzbudilo pořádání informačních lekcí v prvních měsících akademického roku, konkrétně v říjnu a listopadu, nebo potom v březnu (na začátku letního semestru).

Trendy by bylo možné spočítat také z údajů z Clavia, ovšem výsledek by nebyl tak spolehlivý. Data o zobrazení webu by pro správnou interpretaci měla být vyčištěna od záznamů indexovacích robotů jednotlivých vyhledávačů. V této tabulce však systém identifikuje jednotlivé uživatele pouze pod IP adresou, a identifikace dle ní není vždy jednoznačná. Například, velké světové vyhledávače si platí datová centra v Evropě, aby indexovací roboti mohli běžet na jejich strojích; tyto stroje pak mají různé IP adresy.

Dále by pro správnou interpretaci dat o přihlašování musela být data vyčištěna od chyb či útoků, které se vyskytly na přelomu roku 2013 a začátku roku 2014. Jako útoky či chyby charakterizujeme pokusy o přihlášení, které proběhly v krátkém intervalu po sobě a jejichž počet překročil 1 000 v průběhu dvou hodin. Tato „neplatná“ přihlášení by nebylo možné odfiltrovat, kdybychom neměli přístup k datům jako takovým (pokud bychom spoléhali jen na statistiky v Carmen).

<sup>112</sup>Tyto výsledky nejsou překvapivé. Knihovna má v těchto měsících letní režim, který reprezentuje omezená otevírací doba. Také koleje jsou přes léto z většiny zavřené a studenti, kteří jsou většinou skupinou uživatelů knihovny, tráví léto doma.

<sup>113</sup>Například, pokud by se rozhodla provést update v jiný čas, mohla by pomocí trendů odvodit, že září je vhodnější než duben a do jaké míry.

Systém sám samozřejmě žádný problém v datech nerozezná, pro něj jsou všechny pokusy o přihlášení platné. Nicméně by možná bylo lepší, kdyby firma Lanius (jako producent Clavia) zařadila do Carmen systém pro CAPTCHA.<sup>114</sup>

Poslední údaje v grafu na Obrázku č. 3.19, které jsme upravili, byly záznamy o výpůjčkách. Vyčistili jsme je od prodlužování, automatického prodlužování, rezervací a jiných systémových záznamů. Umožnilo nám to srovnat čistě interakce uživatele s knihovnou a jejím systémem v co nejryzejší podobě.

Víme, že momentálně se staví nová budova knihovny, všechna pracoviště v Opavě pak budou pod jednou střechou. Znamená to, že rozložení knihovníků bude daleko příznivější. Pokud by knihovna uvažovala o stanovení nové otevírací doby, bylo by dobré znát trendy v chování uživatelů nejen v průběhu akademického roku, ale také v průběhu týdne, a třeba jednotlivých hodin v konkrétních dnech. Tyto trendy však nelze odvodit z výpůjček, protože ty podléhají současným otevíracím hodinám. Nejspolehlivěji by o způsobech práce uživatelů v čase vypovídala tendence přihlašování do OPACu. Jsme si vědomi toho, že tímto odfiltrujeme všechny uživatele, kteří s katalogem pracují anonymně, ovšem v přístupech do OPACu bez přihlášení nemáme jak jednotlivé uživatele od sebe odlišit.<sup>115</sup> Přihlašování je jedna z mála akcí v OPACu, která má slušné množství záznamů, a pod kterou lze zároveň jednotlivé uživatele jednoznačně identifikovat.

Nyní se podíváme na průběh akademického roku v přihlašování do OPACu a porovnáme k sobě jednotlivá léta, abychom viděli, jestli je průběh trendu stálý. Porovnání naleznete na Obrázku č. 3.20.<sup>116</sup>

Průběh akademických let je normován, aby bylo možné jednotlivé roky jednodušeji srovnávat. Až na některé odchylky je možné říci, že je průběh akademického roku v přihlašování do OPACu poměrně stálý, hlavně v počátečních a ve finálních měsících. Průběh zimního zkouškového období se lehce liší. V roce 2014/2015 aktivita obecně mírně klesla. Pokud ovšem knihovna bude plánovat revizi či období pro nutné uzavření knihovny, ideálními měsíci jsou červenec a srpen.

Podobně se můžeme v datech podívat na průběh týdne v akademických letech. Ověříme, zda se průběh týdne dramaticky nemění, a zjistíme, jaké vzorce chování čtenáři v týdnu vykazují. Z Obrázku č. 3.21 je jasné, že je průběh týdne poměrně stálý ve všech akademických letech; nejvytíženějšími dny v přihlašování do OPACu jsou pondělí až středa.<sup>117</sup> Nejméně aktivity uživatelé vykazují v sobotu. Z hlediska případné nové otevírací doby je vidět, jak velký důraz je kladen na začátek

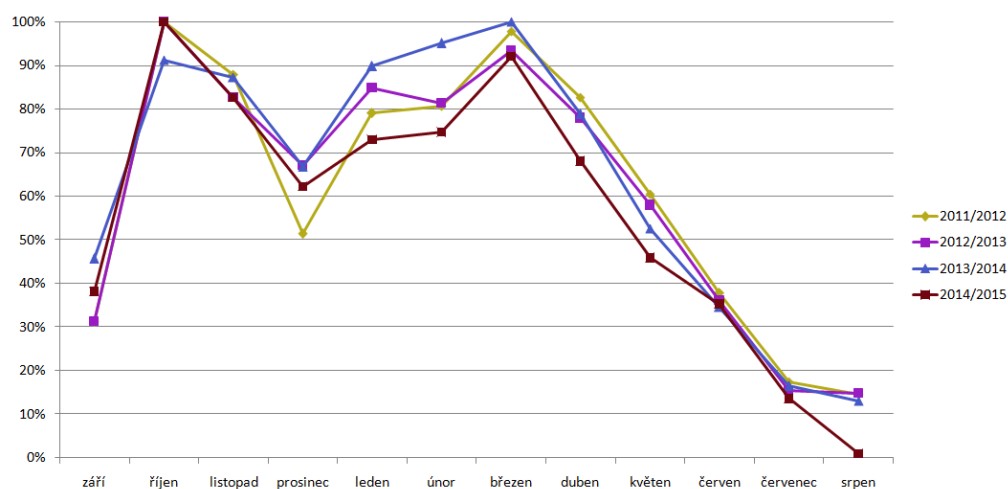
<sup>114</sup>CAPTCHA je program ochraňující webovou stránku proti robotům procházejícím síť. Jedná se o test, vytvořený programem pro rozlišení robota od člověka. (*CAPTCHA* [online]. Carnegie Mellon University, c2000–2010, [cit. 2016-02-15]. Dostupný z: (<http://www.captcha.net/>))

<sup>115</sup>Například všichni uživatelé z Opavy, kteří by přistupovali do OPACu anonymně a jejichž poskytovatelem internetu je firma OpavaNet (tj. jeden z největších poskytovatelů připojení v Opavě) by byli skryti pod jednou IP adresou jako jeden uživatel.

<sup>116</sup>Podobný graf je možné získat taky analýzou dat z Carmen, ovšem naše data jsou vyčištěná od chyb či útoků, které byly zaznamenány v prosinci 2013 až březnu 2014.

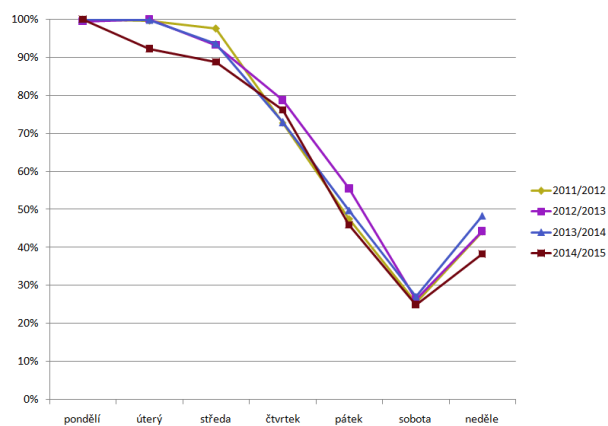
<sup>117</sup>Pravděpodobně také v souvislosti s rozvrhem.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.20: Trendy v přihlašování do OPACu v jednotlivých akademických letech.

týdne, kdy jsou uživatelé zřejmě zvyklí intenzivně pracovat. Průběh všech tří dnů v různých akademických letech je srovnatelný. Pro přesnější specifikaci otevíracích hodin se podíváme ještě na Obrázek č. 3.22. Díky možné uživatelské aktivitě bychom sobotu doporučovali pro spuštění týdenních statistik, čištění logů a jiné pravidelné procedury, protože je v tento den provoz nejmenší.<sup>118</sup>



Obrázek 3.21: Trendy v přihlašování do OPACu v týdnu v ak. letech.

<sup>118</sup>Také k podobným výsledkům lze dojít na základě dat z Clavia, ovšem jak jsme již uvedli, data by potřebovala vyčistit.

### 3 Analýza dat z knihovního softwaru Clavius

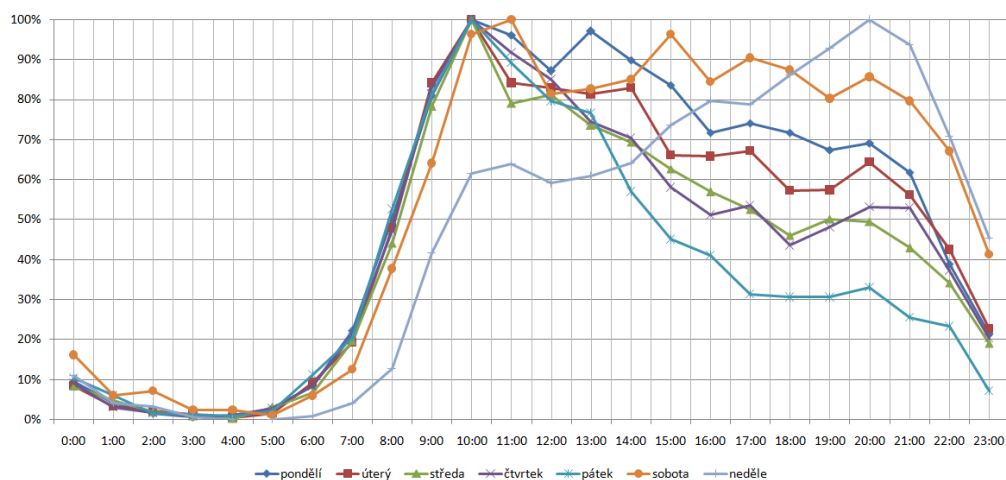
Průběh dní v týdnu po hodinách nalezneme na Obrázku č. 3.22. Tento nám prozradí trendy práce s OPACem v závislosti na dnech (data jsou normována k maximu, aby mohly být trendy porovnány mezi sebou). Vidíme, že průběh práce uživatelů s katalogem se v závislosti na dni v týdnu poměrně mění. Velice podobný je začátek práce, který probíhá od sedmé hodiny ránní a kulminuje kolem desáté dopoledne. Výjimku tvoří neděle, kdy uživatelé začínají s OPACem pracovat kolem osmé a trend vrcholí kolem jedenácté. Od pondělí do soboty pak v různých křivkách počet přihlašování klesá. V pondělí a v úterý se křivka drží kolem 60–70 % až do osmé hodiny večerní, pak teprve strměji klesá. Středa a čtvrtek se od sebe v průběhu výrazně neliší, v trendu je znatelný pokles aktivity od desáté hodiny dopolední až do šesté hodiny večerní, kdy se aktivita v přihlašování o něco zvýší a pak strmě klesne. Pátek vykazuje podobný průběh jako středa a čtvrtek, s tím rozdílem, že propad v průběhu přihlašování od desáté dopolední k šesté odpolední hodině je daleko strmější. Trend se u šesté večerní pouze zastaví a od osmé dál klesá nejstrměji ze všech dní v týdnu. V sobotu se průběh práce s OPACem drží nad 80 % celý den. Pokles nastává až po osmé hodině večerní. V neděli trend práce s OPACem celý den stoupá až k osmé hodině a pak teprve klesá.

Z toho vyplývá, že uživatelé jsou zvyklí o víkendu pracovat dlouho. Také ze začátku týdne vydrží pracovat déle. Otevírací hodiny knihovny by měly určitě pokrýt dobu od deváté hodiny ránní každý den kromě soboty a neděle, kdy je aktivita dle Obrázku č. 3.21 nejnižší. Knihovnu by bylo možné zavřít také v pátek odpoledne, když aktivita klesne ke 30 %. Navzdory trendům na Obrázku č. 3.21 bychom však knihovně nedoporučovali zavřít knihovnu v pátek úplně, protože v tento den mají některé obory výuku kombinovaných studentů. Protože mnoho z nich nemusí pocházet z Opavy, bylo by jim využívání služeb knihovny podstatně ztíženo.

Kromě toho z trendů na Obrázku č. 3.22 a Obrázku č. 3.21 vyplývá, že spouštění týdenních statistik a čištění logů by se mělo provádět v sobotu, po druhé hodině ránní, kdy je provoz v OPACu minimální, jak jsme již uvedli výše. Pokud by byla potřeba udělat údržbu serveru, update softwaru nebo něco podobného, opět doporučujeme tuto dobu.

Dále by výsledků těchto analýz mohla knihovna využít, pokud by potřebovala udělat mezi čtenáři reklamu nějaké službě na svých stránkách. Pokud uživatelé chodí do OPACu přes web knihovny (tuto skutečnost lze ověřit například přes statistiky Google Analytics), potom tento trend částečně pokrývá také provoz na stránkách knihovny. Z výsledků vyplývá, kdy by bylo dobré upozornění na stránky knihovny vložit – pro zimní semestr v říjnu či listopadu v rozmezí pondělí až středa, přesněji od deváté hodiny dopoledne do osmé hodiny večer. Pro letní semestr v březnu, ve stejných dnech v týdnu, a ve stejných časech. V této době by měla zpráva knihovny největší publikum. Zároveň lze předpokládat, že podle analýzy přístupů ve dnech v týdnu budou uživatelé v pondělí až ve středu pravděpodobněji reagovat na emaily a upomínky ze strany knihovny.

### 3 Analýza dat z knihovního softwaru Clavius



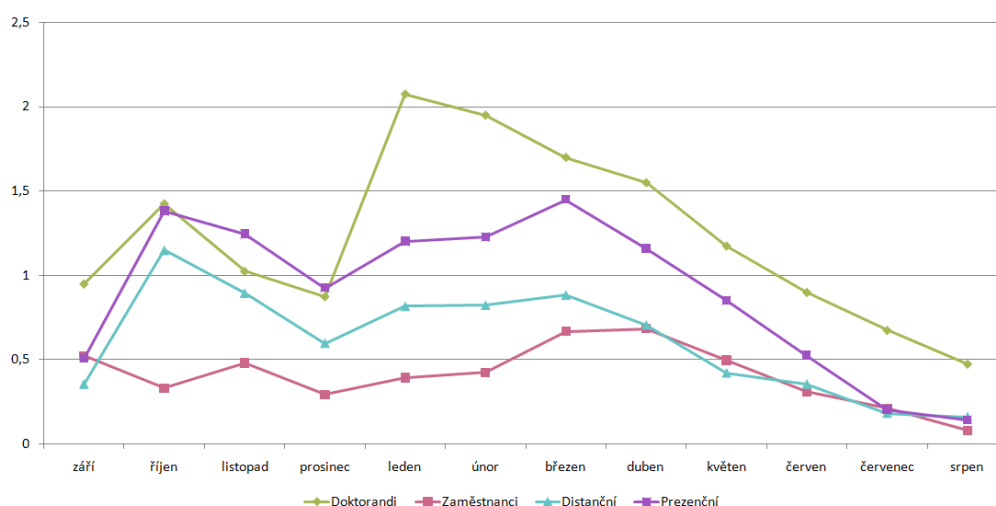
Obrázek 3.22: Trendy v přihlašování do OPACu v hodinách ve dnech v týdnu.

Kdybychom přihlašování do OPACu v jednotlivých dnech v týdnu a hodinách, jako na Obrázku č. 3.22, rozložili dle akademických let, došli bychom k závěru, že průběh je u všech velmi podobný; z toho vyplývá, že se tyto trendy zřejmě nebudou v průběhu let příliš měnit.

Pokud bychom chtěli vycházet z výsledků o přihlašování do OPACu při propagaci služeb knihovny jednotlivým skupinám uživatelů, měli bychom vědět, jaké jsou trendy v přihlašování do OPACu u jednotlivých skupin. Nejdříve se podíváme na přihlašování kategorií čtenářů v akademickém roce. Protože jsou kategorie jinak početně zastoupeny (prezenčních studentů je nejvíce a nejméně je doktorandů), nebylo by možné je spolu porovnávat. Přepočítali jsme proto počty přihlášení na osobu, viz Obrázek č. 3.23.

Z výsledků se zdá, že neaktivnější skupinou v přihlašování jsou doktorandi (výsledky jsou platné za celé sledované období), druhou nejvíce aktivní skupinou jsou prezenční studenti, pak distanční studenti a nejméně aktivní skupinou v přihlašování do OPACu jsou zaměstnanci. Průběh akademického roku je u studentů, ať už distančních nebo prezenčních, podobný. Doktorandi se od těchto dvou skupin liší především ve zkuškovém období; od začátku letního semestru (březen) však vykazují podobný trend. Přihlašování do OPACu v případě zaměstnanců je poměrně nízké, což si vysvětlujeme tak, že na rozdíl od ostatních skupin jsou jejich výpůjčky prodlužovány automaticky na rok. Každý rok pak UKSU zasílá zaměstnancům zprávu s výpisem jejich výpůjček a žádá je, aby je v případě, že knihy již nebudou potřebovat, vrátili. Pokud zaměstnanci po roce knihy nevrátí, jsou jim prodlouženy na další rok. Zvýšené přihlašování v dubnu a březnu tak může znamenat, že si zaměstnanci kontrolují, jaké knihy mají na kontě vlastně načteny.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.23: Průměrné přihlášení kategorií uživatelů v průběhu roku.

Také předchozí výsledky naznačily, že zaměstnanci neprojevují o přihlašování do OPACu přílišný zájem (zda využívají OPAC anonymně, nemáme jak zjistit).

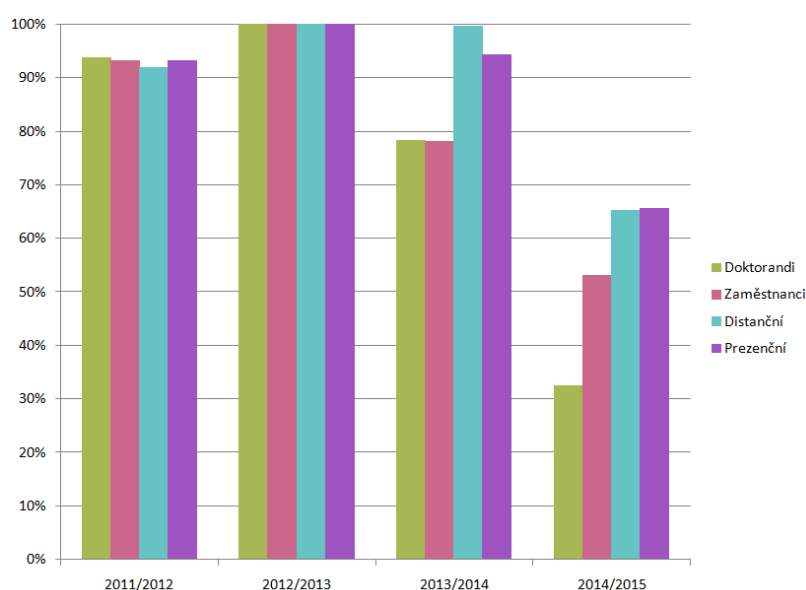
Knihovna by na základě těchto závěrů mohla uvažovat o tom, zda provést průzkum mínění, kterým by mohla zjistit u vybraných uživatelů, užívajících OPAC, jestli jim nějaké funkce nechybí. Pak by mohla společnosti Lanius navrhnout případné vylepšení funkcionality online katalogu.

Nyní by bylo dobré prověřit, jak se proměňuje přihlašování do OPACu během let v daných kategoriích. Z Obrázku č. 3.20 víme, že celkový počet přihlašování do OPACu v akademickém roce 2014/2015 mírně klesl. Na následujícím Obrázku č. 3.24 si rozložíme do akademických let přihlašování v kategoriích. Z tohoto grafu vyplývá, že nejméně aktivním akademickým rokem v přihlašování do online katalogu ve všech skupinách byl akademický rok 2012/2013. Pokud data porovnáme s Tabulkou č. 3.7,<sup>119</sup> zjistíme, že ne vždy se největší počet zástupců z dané skupiny shoduje s nejvyšší aktivitou v přihlašování do OPACu v daném roce.

Například, největší počet distančních studentů měla univerzita v roce 2011/2012; nejvíce aktivních čtenářů v této kategorii ale bylo v roce 2012/2013; pokud bychom se podívali na Obrázku č. 3.24 u distančních studentů na první tři roky, zjistíme, že se aktivita v OPACu držela nad 90%. Z toho plyne, že ačkoliv distanční studenti nevyužívali knihovnu v takové míře jako jiné skupiny (viz Obrázek č. 3.8 a Obrázek č. 3.9), pokud jsou distanční studenti aktivní, přihlašují se poměrně často.

<sup>119</sup>Řádky s hodnotou celkem vyjadřují počet dané skupiny uživatelů na univerzitě v konkrétním akademickém roce. Řádky s označením aktivní symbolizují počet čtenářů v dané kategorii, kteří daný akademický rok využívali knihovnu, tj., byli aktivní. Za aktivního čtenáře je označen čtenář, když si v konkrétním roce vypůjčí publikaci, nebo se přihlásí do OPACu.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.24: Průměrný počet přihlášení na kategorii v akademických letech.

Tabulka 3.7: Počet potenciálních uživatelů celkem k počtu aktivních uživatelů v akademických letech.

	2011/2012	2012/2013	2013/2014	2014/2015
Celkem distančních	3 056	2 743	2 684	2 341
Aktivních distančních	688	722	657	479
Celkem doktorandů	269	253	219	183
Aktivních doktorandů	41	40	47	35
Celkem prezenčních	5 133	4 918	4 641	4 005
Aktivních prezenčních	1 752	1 713	1 589	1 328
Celkem zaměstnanců	621	647	650	604
Aktivních zaměstnanců	249	296	262	252

Zcela opačný trend zaznamenávají zaměstnanci. Jejich počty se dramaticky nemění ani v počtu zaměstnanců na univerzitě celkem, ani v počtu aktivních uživatelů v knihovně, viz Tabulka č. 3.7. Přesto můžeme pozorovat na Obrázku č. 3.24, že počet jejich přihlášení do online katalogu knihovny klesá, konkrétně v letech 2013/2014 a 2014/2015. Výpůjční aktivita však zůstává stejná. To dokládají také Obrázky: Obrázek č. 3.14 a Obrázek č. 3.15. Avšak z Obrázku č. 3.24 plyne, že tomu tak nebylo vždycky, zájem o OPAC u zaměstnanců upadá až v posledních letech.

### 3 Analýza dat z knihovního softwaru Clavius

Počet prezenčních studentů na univerzitě během let klesl asi o 1 100 jedinců. Zajímavé je, že početnější ztráty u aktivních uživatelů knihovny se vyskytují až v posledních dvou sledovaných letech, a ještě zajímavější je, že aktivita prezenčních studentů v přihlašování do OPACu v roce 2013/2014 předčila aktivitu z roku 2011/2012, z čehož vyplývá, že se v roce 2013/2014 přihlašováním aktivně projevovala skupina prezenčních studentů, která buď školu v roce 2014/2015 opustila, nebo o přihlašování ztratila zájem, nebo začala využívat anonymního přístupu.

Poslední skupinou z grafu na Obrázku č. 3.24 jsou doktorandi. Ačkoliv počet doktorandů na univerzitě postupně klesal, v počtu aktivních doktorandů se to projevilo až v posledním roce; viz Tabulka č. 3.7. Podobně na Obrázku č. 3.24 je patrný propad v přihlašování do OPACu ve skupině doktorandů v roce 2014/2015. Propad může znamenat odchod ze školy u skupiny, která se do OPACu často přihlašovala, nebo zvýšení zájmu o anonymní přístup do online katalogu.

Největší pokles v zájmu o přihlašování do OPACu dle Obrázku č. 3.24 zaznamenali tedy doktorandi a zaměstnanci. Pokles v počtu aktivních uživatelů není tak výrazný oproti poklesu zájmu o přihlášení do online katalogu.

Z Tabulky č. 3.7 také vyplývá, že i celkový počet aktivních uživatelů knihovny klesá. Ovšem, pouze klesající počty potenciálních uživatelů klesající využívanost OPACu pod přihlášením nezpůsobují.

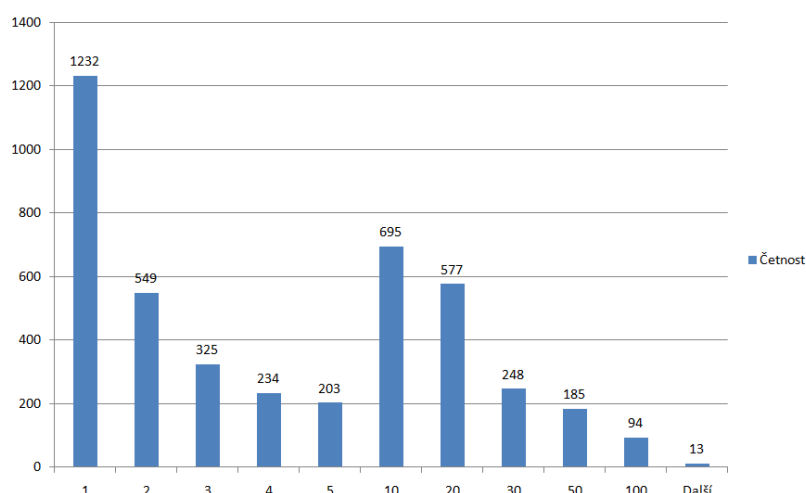
Měřítkem, které vypovídá o aktivitě přihlašování do OPACu, je návratovost uživatelů. Pro porovnání návratovosti uživatelů jsme sestavili histogram, který rozdělil uživatele dle četnosti jejich návratů. Dle Sturgesova pravidla bychom měli stanovit počet tříd na 13 a velikost tříd na 26, ovšem toto pravidlo předpokládá normální rozdělení, které se v datech nevyskytuje. Ani podobná statistická pravidla nedopadla v rozdělení příliš dobře. Proto jsme granulovali výsledky intuitivně sami. Histogram četností můžete vidět na Obrázku č. 3.25. Z dat jasně vyplývá, že největší skupina uživatelů (1 232) navštívila OPAC pouze jednou. Celkem 549 uživatelů se vrátilo jedenkrát, 325 třikrát atd.

Z Obrázku č. 3.25 vyplývá, že téměř třetina uživatelů, kteří se do OPACu přihlásí, se už nikdy nevrátí. Asi dvě třetiny se přihlásí až 10 krát a pouze jedna třetina uživatelů jsou pravidelní návštěvníci OPACu (s přihlášením).

Histogram četnosti nám ale neřekne nic o tom, jak často se uživatelé vrací. Pro zobrazení těchto trendů nám poslouží Obrázek č. 3.26. Barvy reprezentují jednotlivé kategorie, tvary obrazců a velikost pak množství zástupců ve skupině (čtverec znamená přes sto uživatelů, jeden bod označuje jedince a poloměr kruhu udává velikost skupiny až do sta). Škála je v obou případech logaritmická. Osa  $x$  reprezentuje vážený aritmetický průměr návratů do OPACu a osa  $y$  počet sezení, které uživatel zahájil. Jinými slovy, osa  $x$  vyjadřuje, v jakých intervalech dní se vracel, a osa  $y$  zase kolikrát se vrátil. Křivka označuje pravidelnost návratů. Například pokud na ose  $x$  najdeme číslo 5 a vertikálně protneme tuto křivku, najdeme poblíž tohoto místa prezenčního studenta. Jeho pozice na křivce (i v okolí) znamená, že se tento konkrétní student přihlašoval do OPACu každých 5 dní po celé



### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.25: Histogram návratovosti v přihlašování do OPACu.

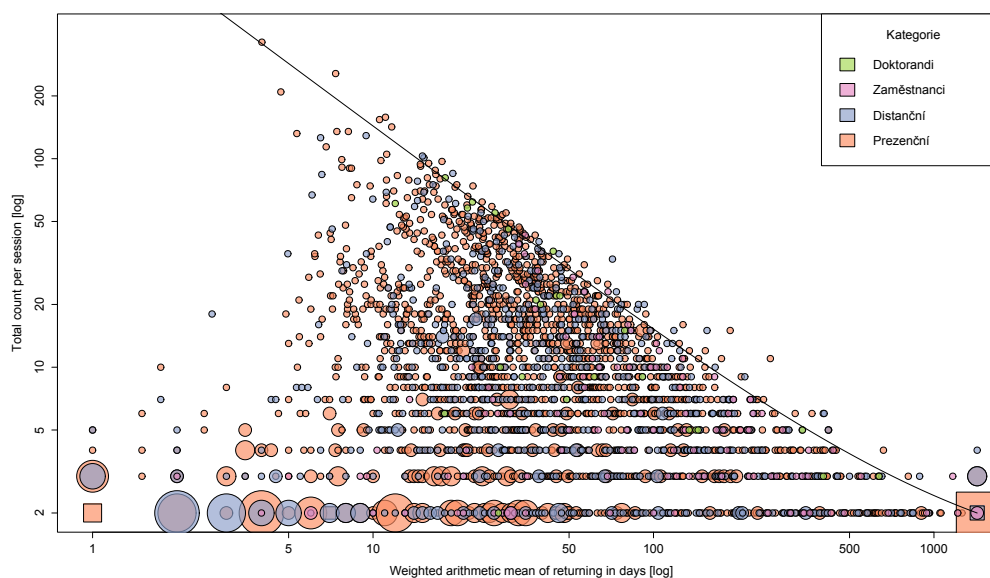
sledované období. Počet jeho přihlášení do online katalogu je maximem v intervalu pěti dní, kterého mohl dosáhnout. Jedinci nad křivkou se přihlašovali vícekrát za den.<sup>120</sup>

Na Obrázku č. 3.26 můžeme vidět, že žádná ze skupin netvoří přímo shluky. Růžové body, reprezentující zaměstnance, se nacházejí spíše ve spodní části grafu. Mnoho z nich se nachází do pěti a v podstatě všichni do počtu padesát přihlášení. Distanční a prezenční studenti (modré a červené body) jsou reprezentováni velkými skupinami u levého dolního rohu (které se v OPACu objevily, a pak už se nikdy nevrátily) a roztržitěnou skupinou individuí vprostřed grafu a výše (tito OPAC využívají často – určité procento z nich velmi pravidelně). Někteří se dokonce přihlásili stokrát a vícekrát. Světle zelené body symbolizují doktorandy, jejichž počet je pouze v jednotkách; ovšem všimněme si, že se jejich chování blíží křivce vpravo, čili, je poměrně pravidelné. Tato kategorie variuje od pěti do sedmdesáti pěti přihlášení (s jednou výjimkou).<sup>121</sup>

Protože Obrázek č. 3.26 zobrazuje všechny přihlášené uživatele najednou, je poměrně málo přehledný. Nyní se pokusíme data zobrazit jinak. Pokud vezmeme dobu trvání databáze jako spojitou proměnnou a rozsekáme ji pro každého uživatele zvlášť na období, po kterých se uživatel do OPACu mohl znovu přihlásit, vydělíme to pravděpodobností výskytu jevu pro daný rozsah a dostaneme Obrázek č. 3.27. Jednotlivé sloupce reprezentují rozsahy dní, závorky pak určují, jestli číslo je nebo

<sup>120</sup>Skupina prezenčních studentů označená čtvercem úplně v pravé části grafu označuje skupinu, která se přihlásila jednou na začátku sledovaného období a jednou na jeho konci. Mohlo jít o uživatele, kteří se přihlásili ve svém prvním semestru a pak v posledním, aby zkontrolovali své uživatelské konto.

<sup>121</sup>Skript ke zpracování přihlašování uživatelů se nachází v Příloze č. 1.



Obrázek 3.26: Poměr návratovosti v přihlašování do OPACu.

není obsahem intervalu. Každý sloupec je navíc rozčleněn barevně dle toho, jaké kategorie uživatelů do něj spadají.

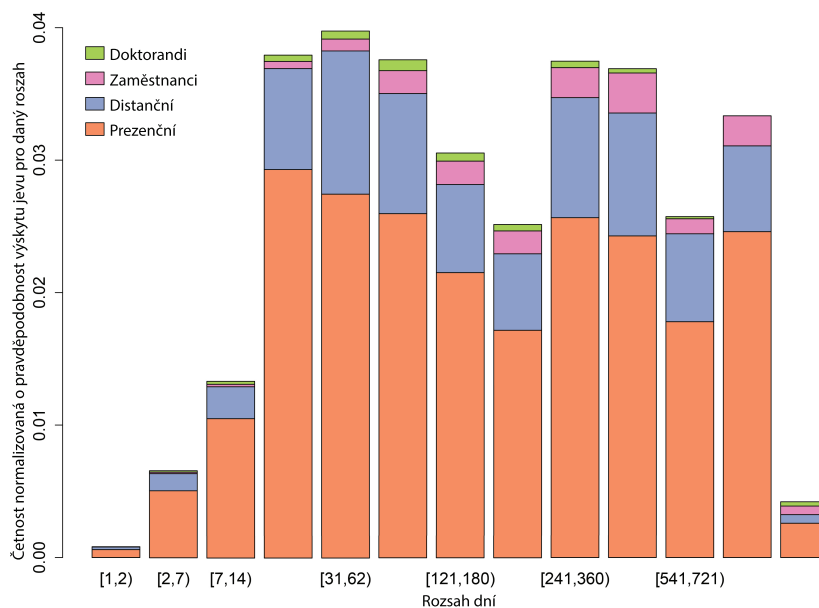
Na Obrázku č. 3.27 tedy vidíme, v jakých intervalech se které kategorie čtenářů do OPACu přihlašovaly nejčastěji. Prezenční studenti se přihlašovali nejvíce v rozsahu 15 až 120 dní. Distanční studenti si udržovali variabilitu od 15 dní dále. Naopak zaměstnanci se přihlašovali spíše po delších intervalech a nejčastěji po roce. Doktorandi aktivitu vykazovali nejvíce po 63 dnech, rozložení jejich celkové aktivity je mezi 15 až 540 dny. Z obrázku je tedy možné vyčíst trend, ve kterém se uživatelé do OPACu přihlašují. Z pravděpodobnosti výskytu, ilustrované škálou vlevo plyne, že se uživatelé přihlašují do OPACu poměrně málo.<sup>122</sup>

Atraktivita OPACu pro uživatele je jedním z důležitých bodů, kterým by měla knihovna věnovat pozornost. Využívání možností nastavení účtu a sestavování seznamu literatury nám mohou pomoci demonstrovat, jak se uživatelé v Carmen orientují, a jestli využívají jejich funkcí. Za celé sledované období bylo zařazeno do seznamu literatury pouze 1 142 různých knih a pouhými 329 uživateli z celkových 4 355 (tj. 13 %).

Pokud se zároveň podíváme na statistiky nastavení uživatelského účtu v OPACu, všimneme si, že možnosti jeho nastavení využilo velmi malé množství uživatelů a to 320. Jedním z doporučení pro knihovnu je, aby více propagovala možnosti

<sup>122</sup>Zřejmě je dostatečné množství funkcí OPACu přístupné anonymně.

### 3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.27: Návrátovost v přihlašování do OPACu.

svého online katalogu a školení s jeho zacházením zařadila také do informačních lekcí.

Další sledovanou aktivitou, dle které zařazujeme uživatele mezi aktivní čtenáře, je vypůjčování dokumentů. Jak jsme předeslali, Clavius umožňuje generovat statistiky k výpůjčkám i s možností zobrazit si průběh dne. Pokud se spokojíme s tím, že údaje mohou zahrnovat různé transakce, a ne pouze výpůjčky, pak nám běžné statistiky postačí. Nyní zkusíme opustit sumarizace, které systém zvládá, a podíváme se na výpůjčky sofistikovanějším způsobem.

Průměrná hodnota výpůjční doby je 93 dní. Tuto hodnotu jistě silně ovlivňují zaměstnanci, kteří mohou držet dokumenty v podstatě po neomezenou dobu. Tuto domněnku potvrzuje rozložení kvantilů, do kterých rozdělíme výpůjční dobu: První kvantil je jeden den. Druhý, obsahující 25 % záznamů, je dlouhý 18 dní. Třetí kvantil je 40 dní a obsahuje 50 % záznamů. Čtvrtý kvantil je 89 dní se 75 % záznamů. A pátý kvantil je 1 420. Průměrná hodnota se tedy nachází kolem čtvrtého kvantilu pro celý datový soubor.

Pokud bychom chtěli spočítat hodnotu délky výpůjční doby pro jednotlivé kategorie, dostali bychom Tabulku č. 3.8. Z tabulky lze usuzovat na chování ve výpůjčkách a rozdíly mezi kategoriemi. Například, většina doktorandů nevrátí knihu dříve než za 12 dní. Třetí kvantily u distančních a prezenčních studentů korelují se standardní dobou výpůjčky, která pro první skupinu činí 45 dní a pro druhou 30 dní.

Tabulka 3.8: Rozložení kvantilů v délce výpůjční doby pro kategorie.

Kategorie	Doktorandi	Zaměstnanci	Distanční	Prezenční
1. kvantil	12	1	1	1
2. kvantil	48	108	29	25
3. kvantil	99	235	44	35
4. kvantil	170	352	63	49
5. kvantil	600	1 135	1 260	864

Doktorandi a zaměstnanci mají standardní dobu výpůjčky nastavenou na jeden rok, proto se u nich doba vrácení prodlužuje. Nutno podotknout, že doktorandi jsou v držení dokumentů daleko ukázněnější; naproti tomu 75 % zaměstnanců potřebuje ke vrácení publikace jeden rok.

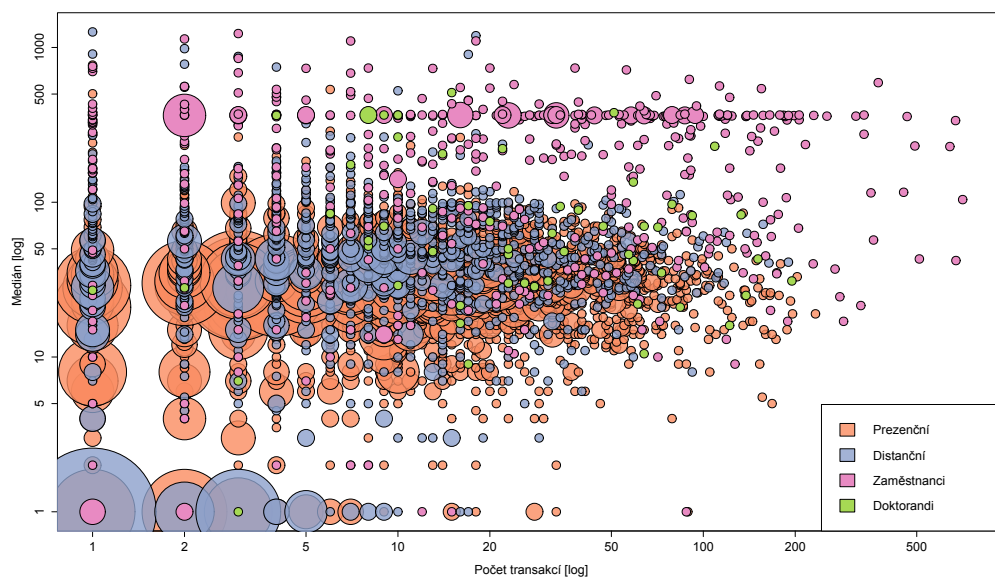
Nyní se podíváme na rozložení délky a počtu výpůjček mezi kategorie, viz Obrázek č. 3.28. Osa  $x$  reprezentuje počet transakcí, čím více doprava, tím větší množství knih si uživatel půjčil. Osa  $y$  reprezentuje medián dní, po které měl uživatel dokumenty vypůjčeny. Obě osy jsou pro přehlednost v logaritmické škále.<sup>123</sup> Na obrázku vidíme jasnou růžovou linku, symbolizující roční výpůjčky zaměstnanců a několika doktorandů. Zajímavé je, že nad 200 výpůjček mají pouze zaměstnanci. Kolem třiceti dní se pohybuje výpůjční doba prezenčních studentů a kolem 45 dní distančních; což není překvapivé. Maximální doba prodloužení je 90 dní; tomuto číslu odpovídají také záznamy. Pouze výjimečně se v grafu nachází několik jedinců, kteří knihy vrací později. Zajímavá je oblast v grafu kolem středu a doprava, kde se nachází část prezenčních a distančních studentů, kteří si pravidelně knihy půjčují ve větším množství. Doktorandi tvoří volně rozprostřenou skupinu mezi ostatními. Z grafu lze usoudit, že si knihy půjčují velmi často a drží je delší dobu.

Ke stejným závěrům lze dojít také na základě Obrázku č. 3.29, který rozčleňuje výpůjční dobu do různých rozsahů v závislosti na četnosti období. Na obrázku lze jasně pozorovat, že výpůjční doba prezenčních studentů nejčastěji trvá 14 až 24 dní. Naopak u distančních studentů je výpůjční doba rozložena do delšího intervalu 14 až 89 dní. Doktorandi jsou jakýmsi mezikrokem mezi zaměstnanci a studenty, délka výpůjční doby u nich variuje. Naopak zaměstnanci si knihy nechávají dlouho, 360 až 540 dní.

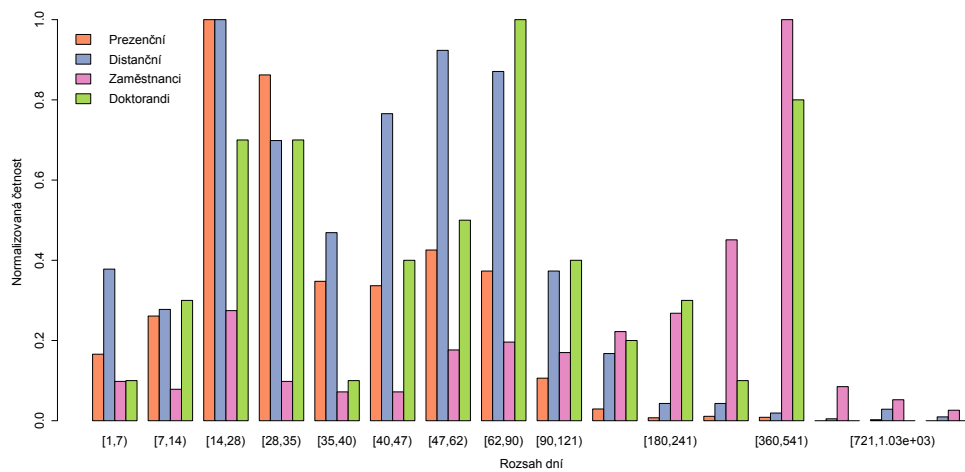
Posledním krokem, který nám zbývá udělat, je prověření, jestli existuje nějaká korelace mezi oběma aktivitami (přihlašování a vypůjčování). V tomto smyslu jsme museli jinak zpracovat tabulku výpůjček. Nekorelovali jsme vypůjčení jako takové, ale aktivitu, kterou lze v tabulce výpůjček registrovat. V průběhu analýzy jsme zjistili, že přihlášení do OPACu a poté provedení akce (vrácení, prodloužení,

<sup>123</sup>Skript ke zpracování této analýzy se nachází v Příloze č. 3.

### 3 Analýza dat z knihovního softwaru Clavius



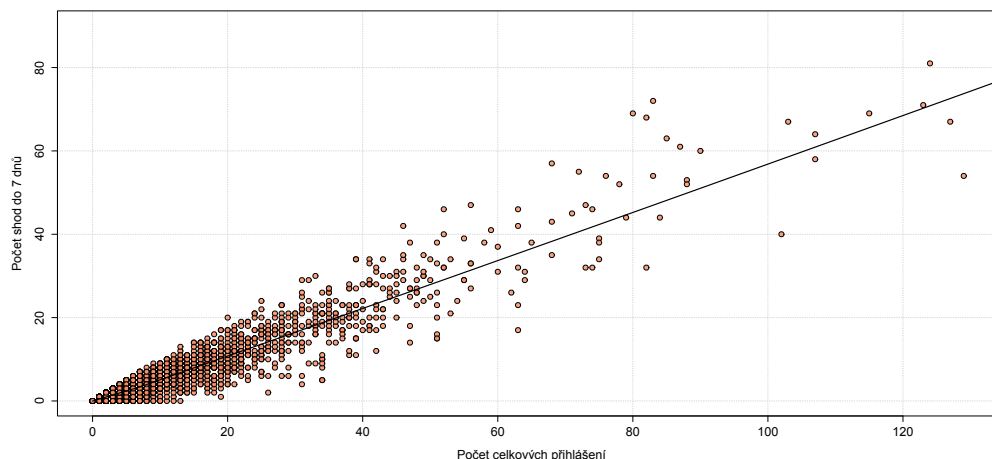
Obrázek 3.28: Množství půjčených dokumentů vztahované k době výpůjčky a ke kategorii.



Obrázek 3.29: Histogram půjčování dokumentů.

### 3 Analýza dat z knihovního softwaru Clavius

rezervace, vypůjčení, objednání MVS) spolu koreluje s 95% relevancí při p-hodnotě 0,001, viz Obrázek č. 3.30.<sup>124</sup>



Obrázek 3.30: Korelace přihlášení a akce uživatele v tabulce výpůjčky.

Z toho plyne, že si uživatelé chodí do OPACu pod přihlášením hlavně kontrolovat výpůjční dobu, prodlužovat, rezervovat publikace a objednávat dokumenty přes MVS.

Obecné trendy pak může management knihovny uvádět do kontextu s výsledky z jiných knihoven. Například Stephen Zweibel a Zachary Lane zpracovávali studii, ověřující výpůjční politiku knihovny. Snažili se také nalézt v datech závislosti a trendy. Během průzkumu přišli na to, že nejdéle drží dokumenty zaměstnanci (první závěr), že celková aktivita uživatelů klesá (druhý závěr), a že se zvyšuje počet online transakcí (třetí závěr).<sup>125</sup>

Knihovna by mohla výsledky naší analýzy porovnat s výsledky Stephena Zweibela a Zachary Laneho ve dvou ze tří bodů. První závěr platí také pro UKSU. Nejdéle drží dokumenty zaměstnanci, (v UKSU) do takové míry, že značně ovlivňují průměrnou hodnotu výpůjční doby až na 93 dní.<sup>126</sup>

Zadruhé v případě UKSU můžeme potvrdit, ubývající počet aktivních uživatelů knihovny (viz Tabulka č. 3.7), což je však dáno poklesem potenciálních uživatelů. Poměr aktivních a neaktivních osob v rámci potenciálních uživatelů se sice mění, ale

<sup>124</sup>Skript ke zpracování korelace se nachází v Příloze č. 2.

<sup>125</sup>ZWEIBEL, Stephen et al. Probing the Effects of Policy Changes by Evaluating Circulation Activity Data at Columbia University Libraries: Edited by Rick J. Block. *The Serials Librarian*. 2012, roč. 65, č. 1, s. 17–27. ISSN 0361-526X.

<sup>126</sup>Víme, že doktorandi mají v tomto smyslu také svůj vliv, ale z Obrázku č. 3.28 jasně vyplývá, že zaměstnanci mají vliv o mnoho větší.

### 3 Analýza dat z knihovního softwaru Clavius

ne dramaticky (v rozsahu maximálně 5–6 % u každé skupiny).<sup>127</sup> Kromě aktivních uživatelů jsme samozřejmě zaznamenali úbytek v přihlašování do OPACu.

Poslední závěr o zvýšených online transakcích nelze z dat UKSU potvrdit ani vyvrátit. Celkový trend přihlašování do katalogu klesl, ovšem to není jediná aktivita, kterou uživatelé mohou v OPACu provádět. Některé tabulky z Carmen bohužel nejde na činnost (konkrétních) uživatelů navázat, takže v tomto bodě nemůžeme o snížení či zvýšení rozhodnout.

Z Obrázku č. 3.28 vyplývá, že si velké množství studentů (distančních i prezenčních) půjčuje dokumenty až na 90 dní. Prodloužení výpůjční doby u UKSU by bohužel, podobně jako v Knihovně University of New York, mohlo přinést pokles výpůjček a rozmrzelost studentů.

Studenti, na které se již výtisk nedostal, si museli opatřovat dokument jinde nebo čekat zbytek semestru. Proto celkového počtu výpůjček ubylo<sup>128</sup> a lze odhadovat, jaký to mělo vliv na spokojenost studentů.

Možnost srovnat změny v politice knihovny šetří peníze, čas i námahu. Pokud knihovna zná dobře situaci, která se v její knihovně nachází, je schopna čerpat ze zkušenosti jiných a vylepšovat svou pozici na univerzitě, oblíbenost u studentů i fungování.

---

<sup>127</sup>U studentů prezenčního studia zůstává poměr zcela překvapivě na 33–34 procentech.

<sup>128</sup>Zweibel, Stephen, 2012.

## Závěr

V teoretické části práce jsme se věnovali popisu pojmu data mining a s ním související problematiky. Dále jsme navázali analýzou literatury s cílem zjistit, k jakým účelům knihovny data mining používají. V rámci plnění prvního cíle práce jsme rozdělili literaturu do tématických skupin dle využití ve fyzických knihovnách. Tyto skupiny popisovaly akvizici a analýzy bibliografických záznamů v katalogu; rozebíraly DM jako proces a jako systém podpory rozhodování. Dále jsme DM ukázali jako prostředek k rozdělování financí a nástroj pro knihovnické služby. Také jsme poukázali na možnosti DM v oblasti evaluace OPACu a knihovního webu a s tím souvisejícím průzkumem uživatelského chování (rozčleněného v tomto případě na průzkum vlivu knihovny na retenci studentů, analýzu využívání fondu a jeho reorganizaci). V neposlední řadě jsme také poukázali na teoretickou možnost DM jakožto prostředku pro analýzu dat ze sociálních sítí.

V průběhu zpracování jsme zjistili, že knihovny DM používají ve velmi širokém záběru. Výsledky jim mohou sloužit jako podklad pro každodenní činnosti (akvizice), pro získání přehledu (analýzy bibliografických záznamů), pro podporu rozhodování a evaluaci činností, k ověřování politiky knihovny, objevování závislostí mezi mnohými činnostmi, a podobně.

Získané znalosti jsme zúročili v praktické části, kde jsme nejdříve zjišťovali, jaké jsou možnosti použití DM na datech z knihovního softwaru Clavius. Ačkoli se nám v rámci tohoto cíle povedlo rozdělit použití do několika tématických skupin, zjistili jsme, že pro provedení mnoha DM postupů bohužel v Claviu chybí data.<sup>129</sup>

Postupy, které by bylo možné v rámci současného stavu dat UKSU provést, se týkají akvizice (analýza dle Kitajimy), analýzy bibliografických záznamů (dle Dentona), služeb (dle Lacroix), průzkumu chování (dle Zweibela, Goswamiho a Botera). V ostatních případech by bylo třeba data knihovny doplnit daty z externího zdroje. V případě použití dat dle Kitajimy, Dentona, Zweibela, Goswamiho a Botera by byla průkaznost výsledků vysoká. Ovšem zopakování Goswamiho použití dat by vyžadovalo speciální znalosti v oblasti síťových grafů.

Obtížně proveditelné analýzy stejného charakteru na datech UKSU pak představují ty, podle studií výzkumníků – Kitajimy, Goswamiho a Botera. Střední obtížnost pak u Lacroix, Zweibela a nízká u Dentona. Z toho plyne, že data, která má UKSU ze systému Clavius k dispozici jsou v aplikaci DM omezena. Z poměrně velkého počtu způsobů zpracování dat (dle jednotlivých studií), uvedených

---

<sup>129</sup>Naše závěry se samozřejmě týkají jen UKSU a knihoven, které mají k dispozici stejné moduly.



v teoretické části práce, se data z Clavia k mnoha druhům analýz nehodí. Je to z toho důvodu, že systém některá data neukládá (např. příslušnost k oborům či oddělením); neeviduje činnosti (jako třeba referenční služby); data jsou mnohdy nevhodně uložena (záznamy o návštěvách); nepropojitelná s jinými údaji (data o neúspěšném vyhledávání); případně je obtížnost propojení vysoká (obtížnost zpracování tabulky *linear*); podobně jako je interpretace dat obtížná (systém nemá v evidenci např. zápis do studia). Samozřejmě víme, že platnost výsledků je omezena pouze na data UKSU.

Posledním cílem práce bylo provést data mining na datech z Clavia, aby bylo možné výsledek využít k podpoře manažerských rozhodnutí v knihovně. Tento cíl sestával ze dvou kroků. Nejdříve jsme zjistili, že data z Clavia lze účinně podpořit rozhodnutí managementu v oblasti uživatelů a v oblasti obsahu fondu. K prvnímu by sloužila analýza chování uživatelů, ke druhému pak analýza bibliografických záznamů.

S cílem co nejlépe podpořit management knihovny, jsme provedli DM nad data, souvisejícími s činností uživatelů. DM bylo možné provést pouze nad takovým typem dat, který bylo možné propojit, a uživatele v něm jednoznačně rozeznat. Analýza tak byla provedena nad uživatelskou tabulkou, tabulkou o přihlášení do OPACu, a nad tabulkou výpůjček. Ukázali jsme také způsob, jakými vyhodnocuje data Clavius, abychom uvedli analýzy do kontextu.

První série analýz se týkala složení uživatelů; má knihovně pomoci blíže identifikovat skupiny, které ji využívají.

Nejprve jsme si definovali aktivní uživatele, abychom byli schopni rozeznat, kteří uživatelé aktivně knihovnu využívají. Systém evidence v Claviu je řešen pouze přes poplatky. My jsme však zjistili, že se zcela nekryjí uživatelé, kteří se hlásí do OPACu, a kteří jsou aktivní ve výpůjčkách.

Ze záznamů jsme vytvořili přehled o tom, v jakých cyklech uživatelé knihovnu využívají, a v jakém poměru jsou vzhledem k fakultám a ke kategoriím. Srovnali jsme výsledky také poměrově; ukázali jsme, jaká je každoroční skladba uživatelů knihovny, potom též jaké procento uživatelů dle kategorie ročně odpadá. Toto jsme provedli pro obě aktivity (přihlašování do OPACu i výpůjčování).

Zavedli jsme pojem stálý uživatel (tj. čtenář byl uživatelem po celé sledované období) a ukázali jsme, jaký je jejich poměr mezi čtenáři.

Výsledky ukázaly, že započítáme-li velikost fakult, je jejich míra ve využívání knihovny velmi podobná; pohybuje se kolem 33%. Poměrově nejaktivnější skupinou byli doktorandi z Obchodně podnikatelské fakulty, kdy 60% z nich knihovnu navštěvovalo. Mezi nejstálejší uživatele patřili zaměstnanci a doktorandi. Zjistili jsme však také, že je velký rozdíl u zaměstnanců mezi přihlašování do OPACu a mezi půjčováním. U zaměstnanců byl poměr půjčování po celé sledované období 32%, kdežto u přihlašování to bylo pouze 8%.

Zjistili jsme, že zhruba 80% uživatelů knihovny se obmění každé dva roky. Což klade velké nároky jak na služby a například také informační lekce.

### 3 Analýza dat z knihovního softwaru Clavius

Druhá série analýz se týkala vzorců chování uživatelů. Určili jsme trend využívání knihovny v akademickém roce, v týdnu a v jednotlivých dnech. Tyto trendy mohou knihovně například pomoci stanovit ideální otevírací dobu. Mimo to, tyto mohou pomoci naplánovat údržbu, softwarové updaty nebo třeba stěhování.

Také se nám podařilo určit, že neaktivnější skupinou v přihlašování jsou dokto-  
randi a prezenční studenti. Také ochota zaměstnanců se do OPACu přihlašovat v posledních několika letech dramaticky klesla. Dále je zřejmá klesající aktivita v celkových počtech, ovšem vesměs je to dáno poklesem počtu potenciálních uživatelů.

Z výsledků také vyplývá, že není velmi populární využívat OPAC knihovny pod přihlášením. Třetina lidí, kteří se do OPACu přihlásili 1–5 krát se už nikdy znovu nepřihlásili. Dodatečné funkce v něm využívá asi 13% uživatelů. Proto bychom mohli doporučit knihovně, aby zacházení s ním zařadila do informačních lekcí a zkusila jej více propagovat. Případně aby provedla průzkum mezi uživateli o tom, jak jim OPAC, a jeho funkce dostupné pod přihlášením, vyhovují.

Potvrdili jsme trendy ve vypůjčování a vracení publikací a zjistili jsme, že nejdéle knihy drží vyučující. Naše výsledky jsme pak mohli srovnat s jednou ze jmenovaných studií, konkrétně autorů Zweibela a Laneho.

Nakonec jsme při hledání vzájemné korelace potvrdili vazbu mezi přihlašováním do OPACu a akcí v systému (vypůjčení, vrácení, rezervace, objednávka MVS), která proběhne do 7 dní poté.

Z výsledků jsme odvodili doporučení, kterými se knihovna může řídit; tyto se nacházejí v praktické části.

## Bibliografie

- 3.0 Česká republika. *Opendata.cz* [online]. [cit. 2016-01-28]. Dostupný z: (<http://opendata.cz/cs/node/29>).
- Adresáře institucí: statistika: knihovna. *Odborný útvar NIPOS* [online]. Ministerstvo kultury ČR, [cit. 2016-02-15]. Dostupný z: (<http://www.nipos-mk.cz/?cat=88>).
- AHMAD, Pervaiz, Mark BROGAN, Michael N. JOHNSTONE. The e-book power user in academic and research libraries: Deep log analysis and user customization. *Australian Academic & Research Libraries*. 2014, roč. 45, č. 1, s. 35–47. ISSN 0004-8623.
- Audience Level. In: *OCLC Research* [online]. OCLC, c2015, [cit. 2015-06-01]. Dostupný z: (<http://www.oclc.org/research/activities/audience.html>).
- BANERJEE, Kyle. Is Data Mining Right for Your Library? *Computers in Libraries*. November/December 1998, s. 28–31.
- BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9.
- BLECIC, Deborah D., Nirmala S. BANGALORE, Josephine L. DORSCH, Cynthia L. HENDERSON, Melissa H. KOENIG, Ann C. WELLER. Using transaction log analysis to improve OPAC retrieval results. *College and Research Libraries*. 1998, roč. 59, č. 1, s. 39–50. ISSN 0010-0870.
- BOTER, Jaap, Michel WEDEL. User categorization of public library collections. *Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. Dostupný také z: (<http://dx.doi.org/10.1016/j.lisr.2005.01.004>). ISSN 0740-8188.
- BOUCHET, Emilie. Autorské právo a digitalizace v archivech, muzeích a knihovnách. *Knihovna*. 2008, roč. 19, č. 1, s. 48–55. Dostupný také z: (<http://oldknihovna.nkp.cz/knihovna81/81048.htm>). ISSN 1801-3252.
- BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *J. Med. Libr. Assoc.* 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>).
- BREEDING, Marshall. Mining Data for Library Decision Support. *Computers in Libraries*. 2013, roč. 33, č. 5, s. 23–25. ISSN 1041-7915.
- Business Intelligence. In: *Technopedia* [online]. Janalta Interactive, c2015, [cit. 2015-06-13]. Dostupný z: (<http://www.techopedia.com/definition/345/business-intelligence-bi>).

## Bibliografie

- CAPTCHA [online]. Carnegie Mellon University, c2000–2010, [cit. 2016-02-15]. Dostupný z: <http://www.captcha.net/>.
- CELBOVÁ, Iva. Dezideratum. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: NK ČR, c2014 [cit. 2016-02-01]. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001182&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001182&local_base=KTD).
- CELBOVÁ, Ludmila. hit. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003– [cit. 2016-02-15]. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000532&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000532&local_base=KTD).
- CELBOVÁ, Ludmila. Z39.50. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)*. Praha: Národní knihovna ČR, 2003–. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000575&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000575&local_base=KTD).
- Clavius – základní informace. *Lanius* [online]. Lanius, [cit. 2016-02-15]. Dostupný z: <http://www.lanius.cz/>.
- CLEYLE, Su, Scott NICHOLSON. Approaching librarianship from the data: using bibliomining for evidence-based librarianship. *Library hi tech*. 2006, roč. 24, č. 3, s. 369–375. ISSN 0737-8831.
- Co umějí Big Data? Třeba pošlou Formuli 1 do boxů nebo poradí fotbalistům. In: *Technet.cz* [online]. Mafra, c2015, [cit. 2015-06-16]. Dostupný z: [http://technet.idnes.cz/big-data-c53-/veda.aspx?c=A141010\\_115149\\_veda\\_kuz](http://technet.idnes.cz/big-data-c53-/veda.aspx?c=A141010_115149_veda_kuz).
- CONDIT FAGAN, Jody. The effects of reference, instruction, database searches, and ongoing expenditures on full-text article requests: an exploratory analysis. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 3–4, s. 264–274. ISSN 0099-1333.
- CULLEN, Kevin. Delving into data. *Library Journal*. 2005, roč. 130, č. 13, s. 30–32. ISSN 0363-0277.
- ČESKÁ REPUBLIKA. Předpis č. 257/2001 Sb. Zákon o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb: knihovní zákon. In: *Sbírka zákonů ČR*. ročník 2001. 98/2001.
- Data Analysis. In: *BusinessDictionary.com* [online]. WebFinance, c2015, [cit. 2015-06-12]. Dostupný z: <http://www.businessdictionary.com/definition/data-analysis.html>.
- Database dump. *Technopedia* [online]. [cit. 2016-02-15]. Dostupný z: <https://www.techopedia.com/definition/23340/database-dump>.

## Bibliografie

- Datová tržiště. In: *IBM Knowledge Center* [online]. IBM, listopad 2012, [cit. 2016-01-20]. Dostupný z: ([http://www-01.ibm.com/support/knowledgecenter/SSGU8G\\_12.1.0/com.ibm.acc.doc/ids\\_acc\\_012.htm?lang=cs](http://www-01.ibm.com/support/knowledgecenter/SSGU8G_12.1.0/com.ibm.acc.doc/ids_acc_012.htm?lang=cs)).
- DECKER, Reinhold, Antonia HERMELBRACHT. Planning and evaluation of new academic library services by means of web-based conjoint analysis. *The Journal of academic librarianship*. 2006, roč. 32, č. 6, s. 558–572. ISSN 0099-1333.
- DENTON, William. On dentographs, a new method of visualizing library collections. *Code4Lib Journal*. 2012, roč. 16, s. 171–179. Dostupný také z: (<http://journal.code4lib.org/articles/6300>). ISSN 1940-5758.
- DICKEY, Timothy J. Books as Expressions of Global Cultural Diversity. *Library Resources and Technical Services*. 2011, roč. 55, č. 3, s. 148–162.
- Digitální knihovna K4* [online]. [cit. 2016-02-15]. Dostupný z: (<http://kramerius.slu.cz/search/#intro6>).
- Emailová korespondence s Evou Lackovou. 7. 3. 2016. [eva.lackova@slu.cz](mailto:eva.lackova@slu.cz).
- EMMONS, Mark, Frances C. WILKINSON. The academic library impact on student persistence. *College and Research Libraries*. 2011, roč. 72, č. 2, s. 128–148. ISSN 0010-0870.
- FAYYAD, Usama, Gregory PIATETSKY-SHAPIRO, Padhraic SMYTH. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. Fall 1996, roč. 17. ISSN 0738-4602.
- FINNELL, Joshua, Walt FONTANE. Reference question data mining: A systematic approach to library outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.
- GOSWAMI, Sumit, Anandarup MUKHERJEE, Mansi KHARBANDA, Abhinav GUPTA, Pulkit SONI. Visualisation of Relationships Among Library Users Based on Library Circulation Data. *DESIDOC Journal of Library & Information Technology*. 2010, roč. 30, č. 2, s. 26–39. ISSN 0974-0643.
- GREINER, Lynn. What is Data Analysis and Data Mining? In: *Database Trends and Applications* [online]. c2015, [cit. 2015-06-13]. Dostupný z: (<http://www.dbta.com/Editorial/Trends-and-Applications/What-is-Data-Analysis-and-Data-Mining-73503.aspx>).
- GUENTHER, Kim. Applying Data Mining Principles to Library Data Collection. *Computers in Libraries*. 2000, roč. 20, č. 4.
- Guidelines for Conducting a Focus Group [online]. Office of Assessment, Duke Trinity College of Arts and Sciences, c2005, [cit. 2016-02-09]. Dostupný z: ([https://assessment.trinity.duke.edu/documents/How\\_to\\_Conduct\\_a\\_Focus\\_Group.pdf](https://assessment.trinity.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf)).

## Bibliografie

- Harvard Library. Library Analytics Toolkit. *Harvard Library Lab* [online]. Date Funded: February 1, 2011, [cit. 2016-02-01]. Dostupný z: <https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>.
- HAVLOVÁ, Jaroslava, Miroslav RESSLER. Web 2.0. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003–, [cit. 2016-02-15]. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000014569&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000014569&local_base=KTD).
- HRABINOVÁ, Světlana. Úvod k Open Bibliographic Data aneb i knihovny mají svá skrytá místa! *Knihovna plus*. 2013, roč. 1. Dostupný také z: <http://oldknihovna.nkp.cz/knihovnaplus131/hrabin.htm>.
- CHANG, Chan-Chine, Ruey-Shun CHEN. Using data mining technology to solve classification problems: A case study of campus digital library. *The Electronic Library*. 2006, roč. 24, č. 3, s. 307–321. Dostupný také z: <http://dx.doi.org/10.1108/02640470610671178>. ISSN 0264-0473.
- CHEN, Chia-Chen, An-Pin CHEN. Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*. 2007, roč. 25, č. 6, s. 711–724. Dostupný také z: <http://dx.doi.org/10.1108/02640470710837137>. ISSN 0264-0473.
- JACKSON, Mary Ellen. *Changing Global Book Collection Patterns in ARL Libraries*. 2006.
- JAFAR, Mohamed, R. SIVAKUMAR. Ant-based clustering algorithms: A brief survey. *International Journal of Computer Theory and Engineering*. 2010, roč. 2, č. 5, s. 787–796. Dostupný také z: <http://www.ijcte.org/papers/242-G730.pdf>. ISSN 1793-8201.
- JOHNSON, Doug. A data mining primer and implications for school library media specialists. *Knowledge Quest*. 2004, roč. 32, č. 5, s. 32–35.
- JONÁK, Zdeněk. TEXTQUEST: software pro obsahovou analýzu. *Ikaros*. 2000, roč. 4, č. 5. Dostupný také z: <http://ikaros.cz/textquest-software-pro-obsahovou-analyzu>. ISSN 1212-5075.
- KAO, S.-C., H.-C. CHANG, C.-H. LIN. Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management*. 2003, roč. 39, č. 1, s. 133–147. ISSN 0306-4573.
- KITAJIMA, Mitsuro, Kensuke BABA, Toshiro MINAMI. An Evaluation of Book Selection in a University Library by Loan Record Analysis. *International Journal of Information and Education Technology*. 2015, roč. 5, č. 10, s. 728–731. ISSN 2010-3689.
- Knihovny. *SU Aleph* [online]. SU Aleph: Neformální sdružení uživatelů knihovního systému Aleph, [cit. 2016-02-15]. Dostupný z: <http://www.sualeph.cz/clients/>.

## Bibliografie

- KRČÁL, Martin. Sběr a využití statistických dat v českých knihovnách. *ProInflow*. 2015, roč. 7, č. 2, s. 41–52. Dostupný také z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/1163>.
- KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online]. Praha: Národní knihovna ČR, 2003–, [cit. 2016-02-15]. Dostupný z: [http://aleph.nkp.cz/F/GJMY8FQJM38NRR8S2KTFEJTBDU47HR8YJX62AUNC9VRSU516-53-47429?func=file&file\\_name=find-b](http://aleph.nkp.cz/F/GJMY8FQJM38NRR8S2KTFEJTBDU47HR8YJX62AUNC9VRSU516-53-47429?func=file&file_name=find-b).
- LACROIX, Eve-Marie, Maria Elizabeth COLLINS. Interlibrary loan in US and Canadian health sciences libraries 2005: update on journal article use. *Journal of the Medical Library Association*. 2007, roč. 95, č. 2, s. 189–194. ISSN 1536-5050.
- LALUHOVÁ, Katarína. Bibliomining – data mining pre knižnice. *ITlib: Informačné technológie a knižnice*. 2007, roč. 4. Dostupný také z: <http://www.cvtisr.sk/itlib/itlib074/laluhova.htm>.
- LANIUS. *Clavius*. c2016. [software]. verze 09.02.2016. Tábor.
- LEONARD, Michelle F., Stephanie C. HAAS, Vernon N. KISLING. Metrics and science monograph collections at the Marston Science Library, University of Florida. *Issues in Science and Technology Librarianship*. 2010, č. 62, s. 1. Dostupný také z: <http://dx.doi.org/10.5062/F4PC308T>. ISSN 1092-1206.
- LEVINE-CLARK, Michael. Developing a Multiformat Demand-Driven Acquisition Model. *Collection Management*. 2010, roč. 35, č. 3–4, s. 201–207. Dostupný také z: <http://dx.doi.org/10.1080/01462679.2010.486965>.
- LI, Ying, Dick R. MILLER, Mary BUTTNER. Bibliographic data mining: automatically building component part records for e-journal articles on the Internet. *Journal of internet cataloging*. 2002, roč. 5, č. 1, s. 29–41. ISSN 1091-1367.
- Machine Learning Group at the University of Waikato. *Weka 3: Data Mining Software in Java* [online]. University of Waikato, [cit. 2016-01-20]. Dostupný z: <http://www.cs.waikato.ac.nz/ml/weka/>.
- Ministerstvo kultury ČR. *Koncepce rozvoje knihoven ČR na léta 2011–2015 včetně internetizace knihoven* [online]. [cit. 2016-02-01]. Dostupný z: [http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce\\_rozvoje\\_knihoven\\_2011-2015.pdf](http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce_rozvoje_knihoven_2011-2015.pdf).
- MITCHELL, Anna. Data Driven Acquisition in the Library. *YouTube* [online]. Publikováno 9. 11. 2015, [cit. 2016-01-02]. Dostupný z: <https://youtu.be/CKbtmRCRemc>.
- New in Mathematica 9: Social Network Analysis. In: *Wolfram Mathematica* [online]. Wolfram, c2015, [cit. 2015-06-18]. Dostupný z: <http://www.wolfram.com/mathematica/new-in-9/social-network-analysis/>.

## Bibliografie

- NIELSEN, Jakob. Thinking Aloud: The #1 Usability Tool. *Nielsen Norman Group* [online]. January 16, 2012, [cit. 2016-02-09]. Dostupný z: ([https://assessment.trinity.duke.edu/documents/How\\_to\\_Conduct\\_a\\_Focus\\_Group.pdf](https://assessment.trinity.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf)).
- NICHOLSON, Scott. A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*. 2004, roč. 60, č. 2, s. 164–182. ISSN 0022-0418.
- NICHOLSON, Scott. A framework for Internet archeology: Discovering use patterns in digital library and Web-based information resources. *First Monday*. 2005, roč. 10, č. 2.
- NICHOLSON, Scott. Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works. *Journal of the American Society for information science and technology*. 2003, roč. 54, č. 12. ISSN 1081-1090.
- NICHOLSON, Scott. *Bibliomining.com* [online]. [cit. 2015-05-20]. Dostupný z: (<http://www.bibliomining.com/>).
- NICHOLSON, Scott. The bibliomining process: Data warehousing and data mining for library decision making. *Information technology and libraries*. 2003, roč. 22, č. 4, s. 146. ISSN 2163-5226.
- NICHOLSON, Scott, Jeffrey M. STANTON. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. 2003, s. 247–262.
- O NTK. *Národní technická knihovna* [online]. NTK, c2006–2016, [cit. 2016-02-15]. Dostupný z: (<https://www.techlib.cz/cs/82794-o-ntk>).
- OCLC. WorldShare Collection Evaluation. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<https://www.oclc.org/collection-evaluation.en.html>).
- OCLC. WorldShare License Manager. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<https://www.oclc.org/license-manager.en.html>).
- OCLC. WorldShare *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<http://www.oclc.org/worldshare.en.html>).
- OCLC. WorldShare Report Designer. *OCLC.org* [online]. [cit. 2016-02-05]. Dostupný z: (<http://www.oclc.org/worldshare-report-designer.en.html>).
- OMASTOVÁ, Pavlína, Radka ŘÍMANOVÁ, Ctibor ŠKUTA. Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety. In: *Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety*. 2011, s. 101–109. ISBN 978-80-86249-62-9.



## Bibliografie

- PAPAVLASOPOULOS, Sozon, Marios POULOS. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, roč. 33, č. 4/5, s. 261–271. ISSN 0143-5124.
- PIATETSKY, Gregory. *CRISP-DM, still the top methodology for analytics, data mining, or data science projects*. In: *KDnuggets* [online]. Oct 28, 2014, [cit. 2016-01-20]. Dostupný z: <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>.
- PIATETSKY-SHAPIRO, Gregory. *Lesson: Data Mining, and Knowledge Discovery: An Introduction*. In: *KDnuggets* [online]. [cit. 2015-05-20]. Dostupný z: [http://www.kdnuggets.com/data\\_mining\\_course/x1-intro-to-data-mining-notes.html](http://www.kdnuggets.com/data_mining_course/x1-intro-to-data-mining-notes.html).
- PILECKÁ, Věra. Zpráva z konference Využívání informací v informační společnosti. *Ikaros*. 2006, roč. 10, č. 11. Dostupný také z: <http://ikaros.cz/node/12255>. ISSN 1212-5075.
- Pracoviště Opava: Elektronické informační zdroje *Slezská univerzita v Opavě* [online]. Slezská univerzita v Opavě, c2011, [cit. 2016-02-15]. Dostupný z: [http://www.slu.cz/slu/cz/univerzita-knihovna/pracoviste-opava/fondy/elektronicke\\_informacni\\_zdroje](http://www.slu.cz/slu/cz/univerzita-knihovna/pracoviste-opava/fondy/elektronicke_informacni_zdroje).
- R Studio* [online]. R Studio, c2015, [cit. 2016-01-20]. Dostupný z: <https://www.rstudio.com/>.
- Reference knihovního systému Clavius v ČR. *Lanius* [online]. Lanius, [cit. 2016-02-15]. Dostupný z: <http://www.lanius.cz/>.
- RENAUD, John et al. Mining Library and University Data to Understand Library Use Patterns. *The Electronic Library*. 2015, roč. 33, č. 3. Dostupný také z: <http://dx.doi.org/10.1108/EL-07-2013-0136>.
- ŘÍMANOVÁ, Radka. Bibliomining – dobývání znalostí z knihovních databází a systémů [online]. 3. 11. 2008, [cit. 2016-02-15]. Dostupný z: <http://uisk.ff.cuni.cz/detail.do?articleId=6569>.
- SAMSON, Sue. Usage of e-resources: Virtual value of demographics. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 6, s. 620–625. ISSN 0099-1333.
- SEDLÁČKOVÁ, Beáta. Data mining a jeho uplatnění při poznávání knihovnických jevů a zákonitostí. *Knižnica*. 2007, roč. 8, č. 11–12, s. 21–22. Dostupný také z: [http://www.snk.sk/images/snk/casopis\\_kniznica/2007/november-december/21.pdf](http://www.snk.sk/images/snk/casopis_kniznica/2007/november-december/21.pdf). ISSN 1335-7026.
- SEWELL, Robin R. Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*. 2013, roč. 31, č. 1, s. 160–170. Dostupný také z: <http://dx.doi.org/10.1108/07378831311303994>. ISSN 0737-8831.

## Bibliografie

- SHIEH, Jiann-Cherng. From website log to findability. *The Electronic Library*. 2012, roč. 30, č. 5, s. 707–720. ISSN 0264-0473.
- SCHULMAN, Sandy. Data mining: Life After Report Generators. *Information Today*. 2003, roč. 15, č. 3, s. 52. ISSN 0363-0277.
- SILIPIGNI CONNAWAY, Lynn, Larry OLSZEWSKI. *A Geographical Representation of WorldCat Resources: A Decision-Making Tool for Acquisitions and Collection Management* [online]. Charleston Conference, November 10, 2006, [cit. 2015-06-01]. Dostupný z: (<http://www.oclc.org/content/dam/research/presentations/connaway/20061110-charleston-worldmap.pdf>).
- SILIPIGNI CONNAWAY, Lynn, Edward T. O'NEILL, Chandra PRABHA. Last Copies: What's at Risk? *College and research libraries*. 2006, roč. 67, č. 4, s. 370–379. Dostupný také z: (<http://www.oclc.org/research/publications/archive/2006/connaway-cr107.pdf>). ISSN 0010-0870.
- SORIA, Krista M., Jan FRANSEN, Shane NACKERUD. Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*. 2014, roč. 40, č. 1, s. 84–91. ISSN 0099-1333.
- SOŠKOVÁ, Michaela. *Analýza a vyhodnocení činností uživatelů souborného online katalogu Masarykovy univerzity*. Brno, 2006. Bakalářská práce. Kabinet knihovnictví, Ústav české literatury a knihovnictví, Filozofická fakulta, Masarykova univerzita. Dostupný také z: ([http://is.muni.cz/th/108801/ff\\_b/Bakalarka\\_prace\\_Soskova.doc](http://is.muni.cz/th/108801/ff_b/Bakalarka_prace_Soskova.doc)).
- SOUKOPOVÁ, Jana. Nákladově-výstupové metody hodnocení (CMA, CEA, CUA). In: *Veřejné zakázky a veřejné projekty a jejich hodnocení* [PDF]. Masarykova univerzita, 2013 [cit. 9. 2. 2016]. Dostupný také z: ([https://is.muni.cz/el/1456/jaro2013/MPV\\_VZVP/um/33148301/Studijni\\_text\\_nakladove\\_vystupove\\_metody\\_CMA\\_CEA\\_CUA.pdf](https://is.muni.cz/el/1456/jaro2013/MPV_VZVP/um/33148301/Studijni_text_nakladove_vystupove_metody_CMA_CEA_CUA.pdf)).
- SPSS Statistics: Put the power of advanced statistical analysis in your hands* [online]. IBM, [cit. 2016-01-20]. Dostupný z: (<http://www-01.ibm.com/software/analytics/spss/products/statistics/buy-now.html>).
- STAMPER, Jason. Open source BI firms refute SAS CEO's criticism. In: *Computer Business Review* [online]. Digital Media Group, 15:33, November 25, 2010, [cit. 2016-01-20]. Dostupný z: (<http://www.cbronline.com/news/open-source-bi-firms-refute-sas-ceos-criticism>).
- Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2011. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/22951/download/>).

## Bibliografie

- Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2012. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/29843/download/>).
- Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2013. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/34251/download/>).
- Statistická ročenka školství: Zaměstnanci a mzdové prostředky 2014. *MŠMT* [online]. MŠMT, c2013-2016, [cit. 2016-02-20]. Dostupný z: (<http://www.msmt.cz/file/36358/download/>).
- TEMPELMAN-KLUIT, Nadaleen, Alexa PEARCE. Invoking the User from Data to Design. *College and Research Libraries*. 2014, roč. 75, č. 5, s. 616–640. Dostupný také z: (<http://dx.doi.org/10.5860/crl.75.5.616>). ISSN 0010-0870.
- Text Mining. In: *Statistics – Textbook* [online]. Dell, c2015, [cit. 2015-06-13]. Dostupný z: (<http://documents.software.dell.com/Statistics/Textbook/Text-Mining>).
- The DataCamp Team. What is the best statistical programming language? Infograph. In: *DataCamp* [online]. June 3rd, 2014, [cit. 2016-01-20]. Dostupný z: (<https://www.datacamp.com/community/tutorials/statistical-language-wars-the-infograph>).
- The R Project for Statistical Computing* [online]. R Foundation, [cit. 2015-06-20]. Dostupný z: (<http://www.r-project.org/>).
- Timeline Univerzitní knihovny SU v Opavě. *Facebooková stránka Univerzitní knihovny SU v Opavě* [online]. [cit. 2016-02-15]. Dostupný z: ([https://www.facebook.com/UniverzitaSUvOpav%C3%83%C2%AD-knihovna-SU-v-Opav%C3%84%C2%9B-212446575439142/timeline?ref=page\\_internal](https://www.facebook.com/UniverzitaSUvOpav%C3%83%C2%AD-knihovna-SU-v-Opav%C3%84%C2%9B-212446575439142/timeline?ref=page_internal)).
- TRIPATHI, Manorama, V. K. J. JEEVAN. A selective review of research on e-resource usage in academic libraries. *Library Review*. 2013, roč. 62, č. 3, s. 134–156. ISSN 0024-2535.
- TSAI, Hsu-Hao. Global data mining: An empirical study of current trends, future forecasts and technology diffusions. *Expert Systems with Applications*. 2012, roč. 39, č. 9, s. 8172–8181. Dostupný také z: (<http://dx.doi.org/10.1016/j.eswa.2012.01.150>).
- Univerzitní knihovna *Slezská univerzita v Opavě* [online]. Slezská univerzita v Opavě, c2011, [cit. 2016-02-15]. Dostupný z: (<http://www.slu.cz/slu/cz/univerzita-knihovna>).
- Univerzitní knihovna Slezské univerzity v Opavě. *Knihovní řád Univerzitní knihovny Slezské univerzity v Opavě* [online]. [cit. 2016-01-28]. Dostupný z: (<http://www.slu.cz/slu/cz/univerzita-knihovna/pracoviste-opava/o-nas/KnihovnaUniverzita-knihovnySlezskuniverzity.pdf>).

## Bibliografie

- VODIČKOVÁ, Hana. KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV). In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. [cit. 2016-02-15]. Praha: Národní knihovna ČR, 2003-. Dostupný z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001236&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001236&local_base=KTD).
- Výkonové ukazatele: VŠ studenti podle formy a typu studijního programu: podle vysoké školy/fakulty. *MŠMT Odbor analyticko-statistický* [online]. MŠMT, 2016, [cit. 2015-06-15]. Dostupný z: [http://dsia.uiv.cz/vystupy/vu\\_vs\\_f2.html](http://dsia.uiv.cz/vystupy/vu_vs_f2.html).
- Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2011* [online]. Slezská univerzita v Opavě, 2012, [cit. 2015-06-15]. Dostupný z: <http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodoby-zamer/2011/vz-cinnost-2011.pdf>.
- Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2012* [online]. Slezská univerzita v Opavě, 2013, [cit. 2015-06-15]. Dostupný z: <http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodoby-zamer/2012/vz-cinnost-2012.pdf>.
- Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2013* [online]. Slezská univerzita v Opavě, 2014, [cit. 2015-06-15]. Dostupný z: <http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodoby-zamer/2013/vz-cinnost-2013.pdf>.
- Výroční zpráva o činnosti Slezské univerzity v Opavě za rok 2014* [online]. Slezská univerzita v Opavě, 2015, [cit. 2016-02-20]. Dostupný z: <http://www.slu.cz/su/slu/cz/dokumenty/vyrocní-zpravy-a-dlouhodoby-zamer/2014/vz-cinnost-2014.pdf>.
- WALTERS, William H. A regression-based approach to library fund allocation. *Library Resources and Technical Services*. 2006, roč. 51, č. 4, s. 263–278. ISSN 0024-2527.
- WampServer* [online]. [cit. 2015-06-20]. Dostupný z: <http://www.wampserver.com/en/>.
- Web Mining. In: *Technopedia* [online]. Janalta Interactive, c2015, [cit. 2015-06-13]. Dostupný z: <http://www.techopedia.com/definition/15634/web-mining>.
- WEINER, Sharon. The Contribution of the Library to the Reputation of a University. *The Journal of Academic Librarianship*. 2009, roč. 35, č. 1, s. 3–13. ISSN 0099-1333.
- WHITMIRE, Ethelene. Academic library performance measures and undergraduates' library use and educational outcomes. *Library and information science research*. 2002, roč. 24, č. 2, s. 107–128. ISSN 0740-8188.

## Bibliografie

- WILL, Nicole. Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*. 2006, roč. 73, č. 8, s. 1045–1050. ISSN 0040-1625.
- WorldMap. In: *OCLC Research* [online]. OCLC, c2015, [cit. 2015-06-01]. Dostupný z: <http://www.oclc.org/research/activities/worldmap.html>.
- WU, Chien-Hsing. Data mining applied to material acquisition budget allocation for libraries: design and development. In: *Expert Systems with Applications*. 2003, roč. 25, č. 3, s. 401–411. ISSN 0957-4174.
- WU, Chien-Hsing, Tzai-Zang LEE, Shu-Chen KAO. Knowledge discovery applied to material acquisitions for libraries. In: *Information Processing and Management*. 2004, roč. 40, č. 6, s. 709–725. ISSN 0306-4573.
- YANG, Shih-Ting. An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*. 2012, roč. 10, č. 2, s. 163–173. ISSN 1741-5063.
- Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition: opinio: stanovisko k právní otázce.* Objednatel Národní technická knihovna. Zhotovitel Masarykova univerzita. Právnická fakulta. Zpracovatel Radim POLČÁK, Matěj MYŠKA. Jaromír ŠA-VELKA. Brno, 2012. Dostupný z: <http://www.techlib.cz/files/download/id/3157/open-bibliographic-data-ntk-studie-2012.pdf>.
- ZWEIBEL, Stephen, Zachary B. LANE. Probing the Effects of Policy Changes by Evaluating Circulation Activity Data at Columbia University Libraries: Edited by Rick J. Block. *The Serials Librarian*. 2012, roč. 65, č. 1, s. 17–27. ISSN 0361-526X.