

KARLOVA UNIVERZITA V PRAZE
FILOZOFICKÁ FAKULTA
ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ

Informační věda



Teze 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



Syllabus of the 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

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

Úvod	1
1 Data mining	3
2 Data mining v knihovnách	5
2.1 Použití data miningu ve fyzických knihovnách	5
2.2 Použití data miningu v digitálních knihovnách	6
3 Analýza dat z knihovního softwaru Clavius	7
3.1 Využití dat z knihovnického softwaru Clavius	7
3.1.1 Vyhodnocení – Data mining nad daty z Clavia	10
3.1.2 Analýzy – rozložení uživatelů	10
3.1.3 Analýzy – vzorce chování	13
Závěr	20

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 jsme se jej snažili představit.

Toto představení bude zahrnovalo vysvětlení pojmů, metodik, technologií a softwaru, které se k jeho provádění používají, dále metod, jejichž použití na data se za data mining označuje. Nakonec jsme probrali právní a etické otázky, kterými by se každý výzkumník měl zabývat.

Prvním cílem práce je analýza použití data miningu v knihovnách; zjištění, k čemu jej knihovny využívají. Metodou zpracování je průzkum literatury. Zdrojem literatury byly články z odborných databází 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, Knihovna a ITlib).

Druhým cílem bylo zjistit, jaké je použití data miningu na dostupných datech. Použili jsme data z knihovního softwaru Clavius, využívaného Univerzitní knihovnou Slezské univerzity v Opavě, dostupného pro účely této práce. Příklady pro použití data miningu jsme čerpali z prostudovaných knihovnických článků.

Posledním cílem bylo provedení data miningu na datech Univerzitní knihovny Slezské univerzity v Opavě, vycházející z jejich potřeb (a z možností softwaru Clavius). Protože má knihovna dlouhodobě problémy s ochotou zřizovatele jí

Obsah

naslouchat, zaměřili jsme se primárně na analýzy, které by knihovně mohly pomoci v manažerských rozhodnutích a podpořily by její pozici u zřizovatele.

Tohoto jsme se snažili dosáhnout 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, bylo 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

S termínem data mining těsně souvisí termín knowledge discovery in databases (dále jen KDD). Fayyad, Shapiro a Smyth vymezili ve svém článku v roce 1997 KDD jako proces hledání dříve neznámých a užitečných znalostí v datech; k tomuto cíli měl KDD využívat metod data miningu.¹ Data mining byl tedy chápán jako užší termín, který zahrnuje iteraci procesů k objevení konkrétních závislostí v datech. Mnoho autorů² 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 DM se pojí celá řada dalších pojmů, které jsme si v práci vyjmenovali a vymezili.

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. Uvedli jsme tři nejpopulárnější. Na platformě závislou metodiku A5 (od firmy SPSS), metodiku SEMMA (od firmy SAS) a na platformě nezávislou metodiku CRISP-DM (Cross-Industry Standard Process for Data Mining). V průběhu této práce jsme se drželi metodiky CRISP-DM, umožňující volnější pohyb mezi jednotlivými fázemi zpracování dat.

Pojednali jsme také o tom, který software se v souvislosti s data miningem používá, uvedli jsme si zástupce z placených softwarů i Open Source. 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. Zmínili a vysvětlili jsme také pojmy datový sklad, datové tržiště a OLAP.

Volba metody souvisí s úlohami, které chceme provést a zároveň souvisí s daty, na něž je hodláme aplikovat. Typy úloh, které můžeme data miningu zadat, jsou klasifikace, predikce, deskripce, hledání „zajímavých“ zjištění.³

Máme-li představu o úloze, kterou bychom nechali zpracovat, musíme se podívat na data, která jsou k dispozici. Berka⁴ řadí mezi metody KDD: rozhodovací

¹FAYYAD, Usama et al. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. Fall 1996, roč. 17. ISSN 0738-4602.

²Včetně Piatetsky-Shapira a dalších. (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))

³BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9, s. 18.

⁴BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9.

1 Data mining

stromy, asociační pravidla, rozhodovací pravidla, neuronové sítě, evoluční algoritmy, bayesovskou klasifikaci, strojové učení založené na analogiích a induktivní logické programování. Pod DM zařazuje ze statistiky hlavně regresní, diskriminanční a shlukovou analýzu.⁵ Jak je uvedeno v pramenech, rozdíl mezi statistikou a DM v použití statistických nebo jiných metod je v účelu použití.⁶ Použijeme-li statistické metody pro DM, jde o DM. Použijeme-li techniku užívanou k jinému účelu než je objevování závislostí v datech, o DM se nejedná.

Existuje několik právních a etických otázek, které jsme si museli ujasnit. Tyto se týkaly práva organizace ukládat údaje o uživatelích či zákaznících; tématu, kdo je vlastníkem databáze, a kdo k ní vlastní autorská práva; etických zásad výzkumníka a souhlasu knihovny se zpracováním a zveřejněním údajů o ní.

⁵BERKA, Petr. *Dobývání znalostí z databází*. 2003. ISBN 80-200-1062-9, s. 46–59.

⁶GREINER, 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>).

2 Data mining v knihovnách

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.

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 spojovat¹ a ukládat do datových skladů.

Průzkumem literatury byla témata rozdělena do těchto okruhů: Akvizice s využitím DM, Analýzy záznamů v katalogu, Systémy pro podporu rozhodování, Rozdělování financí, Služby a data mining, Evaluace OPACu a webu knihovny, Průzkum chování uživatelů pomocí DM a Analýzy dat ze sociálních sítí. Ke každému okruhu byla zmíněna řada studií; bylo pojednáno o tom, zda a jak jsou studie, obsahující aplikaci DM, nápomocné knihovnám.

Součástí tématu bylo také zmínění pojmu bibliomining a jeho užití. Zmínili jsme také aplikaci data miningu v českých a slovenských knihovnách. Uvedli jsme

¹GUENTHER, Kim. Applying Data Mining Principles to Library Data Collection. *Computers in Libraries*. 2000, roč. 20, č. 4.

dvě praktické práce,² dvě teoretické práce³ a jmenovali jsme také dvě prezentace na toto téma.⁴

2.2 Použití data miningu v digitálních knihovnách

Zde jsme uvedli pouze modelové studie, protože jsme se hodlali zaměřit na fyzické knihovny. V této části jsme rozdělili okruhy článků pouze na teoretické, praktické, a ty s použitím text miningu. Teoretické práce se zabývaly návrhy algoritmů, aplikací, popřípadě konceptů. Z praktických studií jsme uvedli nejvýznamnější práci z Taiwanu.

²OMASTOVÁ, 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; 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).

³SEDLÁČKOVÁ, Beáta. Data mining a jeho uplatnění při poznávání knihovnických jevů a zákonitostí. *Knížnica*. 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). ISSN 1335-7026.; 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>).

⁴ŘÍMANOVÁ, 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>); PILECKÁ, Věra. Zpráva z konference Využívanie informácií 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 teoretické části jsme potvrdili, že vytěžování dat lze v knihovnách využívat. Také jsme ukázali, jakým způsobem je možné knihovnám pomoci. Pro referenci jsme použili akademické a veřejné knihovny, které DM aplikovaly. Cílem praktické části práce pak bylo ověření, jaké jsou možnosti použití DM na datech z Clavia. Použití bylo ověřováno na příkladech z teoretické části. Dalším z cílů praktické části bylo zjistit, zda i přes absenci manažerského modulu lze data z knihovního softwaru stáhnout a vyhodnotit tak, aby bylo výsledky možné použít k manažerským rozhodnutím v knihovně.

Pro tento účel byly k dispozici data Univerzitní knihovny Slezské univerzity v Opavě (dále jen UKSU), používající systém Clavius. Popisovali jsme období od 1. 9. 2011 do 3. 8. 2015. Data byla předem anonymizována.¹

Tato knihovna slouží především akademickým pracovníkům a studentům Slezské univerzity v Opavě. Má k dispozici moduly Katalogizace, OPAC (Carmen), Revize fondu a Výpůjční protokol.² Pokud by měla jiná knihovna k dispozici data také z jiných modulů, bylo by možné, že by se její možnosti pro analýzu mohly lišit. Data jsme zpracovali s vazbou na akademické roky, nikoliv na roky kalendářní. Analýzy byly provedeny v softwaru R a Microsoft Excel. MS Excel byl použit převážně jako nástroj pro vizualizace.

3.1 Využití dat z knihovnického softwaru Clavius

Využití dat z knihovnického softwaru Clavius jsme se rozhodli rozdělit dle článků v teoretické části na akvizici, analýzu záznamů v katalogu, systémy pro podporu rozhodování, rozdělování financí, služby, průzkum chování uživatelů, analýzy dat ze sociálních sítí. Mnoho ze zmíněných studií vycházelo kromě dat z knihovny také z jiných dat, propojených následně s těmi knihovními. Studie například doplňkově využívaly data školy; v našem případě jsme data školy neměli k dispozici, a proto můžeme posuzovat pouze vhodnost dat, uložených v knihovním softwaru.

Probrané způsoby analýzy dat na knihovních datech, a ověření jejich použití na datech ze systému Clavius shrneme v Tabulce č. 3.1. Do červena zbarvené

¹V tabulkách je uveden jednoznačným identifikátor uživatele, díky kterému jej bylo možné přiřadit k jednotlivým činnostem.

²Emailová korespondence s Evou Lackovou. 7. 3. 2016. eva.lackova@slu.cz.

3 Analýza dat z knihovního softwaru Clavius

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 analýzu bez toho, aby k nim bylo cokoli přidáno. V mnohých případech by bylo 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 by bylo 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 strukturována³ a jestli by bylo nutné využít dalších znalostí.⁴

Dále jsme postupovali ve dvou krocích. Nejdříve jsme zjistili, jaká data lze z Clavia pro podporu managementu zanalyzovat. Data v Claviu byla vhodná k analýze pouze ve dvou kategoriích, pro analýzu o uživatelích a analýzu dat o fondu. My jsme se specializovali na data o chování uživatelů.

Pro přehlednost jsme ukázali příklady statistik, které Clavius sám nabízí. Zpracování statistik se nachází ve Výpůjčním modulu,⁵ přičemž je v něm možné generovat tiskové sestavy s konkrétním obsahem, například výpůjčky ve dnech v měsíci, statistiku MVS, návštěvnost v týdnu atd.

Při prověřování statistik UKSU jsme však narazili na problémy; naše čísla se dramaticky lišila od těch, nacházejících se ve výročních zprávách. Zjistili jsme, že údaje jsou zkreslovány tím, že se do celkového čísla ve statistikách Clavia počítají také prolongace, které systém někdy v datech označuje jako výpůjčky. Pro přesnější výsledky je třeba tato data odfiltrovat. Z těchto čísel si knihovna může nechat přímo systémem odstranit pouze statistiky MVS.

Dalším problémem je počet čtenářů, které Clavius počítá vždy pouze z těch, kteří mají zaplacenou registraci na daný rok. Samozřejmě, zaměstnanci univerzity poplatky neplatí. Když je takový čtenář automaticky převeden do dalšího roku, znamená to, že aktivně knihovnu používá? Dále knihovna nově zprostředkovává také elektronické knihy. Tyto je možné nalézt přes OPAC, kam se mohou uživatelé přihlásit kdykoliv i bez zaplacení registrace (to je řešeno přes jednotné přihlašování školy). Knihu si poté mohou půjčit v rozhraní Ebrary. Teoreticky tedy platí, že pokud uživatel využívá pouze OPAC a půjčuje si knihy z Ebrary, není počítán mezi čtenáře knihovny.

³Složitější struktura uložení značí vyšší obtížnost zpracování. Například tabulka `linear`.

⁴Například znalosti ve fuzzy logice nebo v oblasti neuronových sítí.

⁵Lanius. *Clavius*. c2016. [software]. verze 09.02.2016. Tábor.

3 Analýza dat z knihovního softwaru Clavius

Tabulka 3.1: Souhrn způsobů použití dat z Clavia

	Aplikace na datech UKSU	Použitelné tabulky	Průkaznost v souč. stavu	Náročnost provedení
Akvizice				
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)
Analýza bib. záznamů				
Denton	ano	linear	vysoká	nízká
Sys. pro podp. rozhodování				
Papavlasopoulos	ne (potřeba ext. dat)	kpujcky	X	v souč. stavu není možné zjistit
Rozděl. financí				
Walters	ne (potřeba ext. dat)	svazky	X	v souč. stavu není možné zjistit
Služby				
Tempelman-Kluit + Fontane + Fagan	ne (potřeba ext. dat)	ctenari	X	v souč. stavu není možné zjistit
Lacroix	ano	kpujcky, tituly, svazky	střední	střední
Evaluace OPACu a webu				
Blecic + Soskova	ne (špatné uložení)	data jsou neprovázatelná	X	střední
Průzkum chování				
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
Anal. dat ze soc. sítí				
Sewell	ne (potřeba ext. dat)	X	X	X

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 poměr z Fakulty veřejných politik (dále jen FVP), není ale možné zjistit.

3.1.1 Vyhodnocení – Data mining nad daty z Clavia

Vzhledem k situaci UKSU, problémům s prosazováním změn a nových rozhodnutí na celouniverzitní úrovni, jsme si pro analýzu vybrali data spojená s chováním uživatelů. Od výsledků si slibujeme lepší popis současné situace a podklady pro prosazení změn v knihovně, navázané zvláště na fakulty.

Zjistili jsme ovšem, že 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 jsme proto museli směřovat k analýze takových akcí uživatelů, které systém sleduje, a kterých by bylo možné využít k podpoře managementu.

3.1.2 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. Aby naše tvrzení co nejvíce odpovídala skutečnosti, definujeme aktivního uživatele; za aktivního uživatele považujeme každého uživatele, který si vypůjčil knihu, nebo se přihlásil do OPACu⁶ (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.⁷

Z výsledků jsme odvodili, že největší počet aktivních uživatelů z jednotlivých fakult má OPF, avšak v poměru k celkovému počtu potenciálních uživatelů

⁶Návštěva 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.

⁷Při 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.

z fakult není hodnota nad průměrem.⁸ 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.

Dalším tématem, které jsme pro knihovnu zpracovali, je analýza složení stálých uživatelů.⁹ Mezi stálé uživatele se dle našeho předpokladu řadí především doktorandi a zaměstnanci, tento předpoklad se potvrdil. 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 %.

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

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 v aktivitách vypůjčování a přihlašování do OPACu liší.

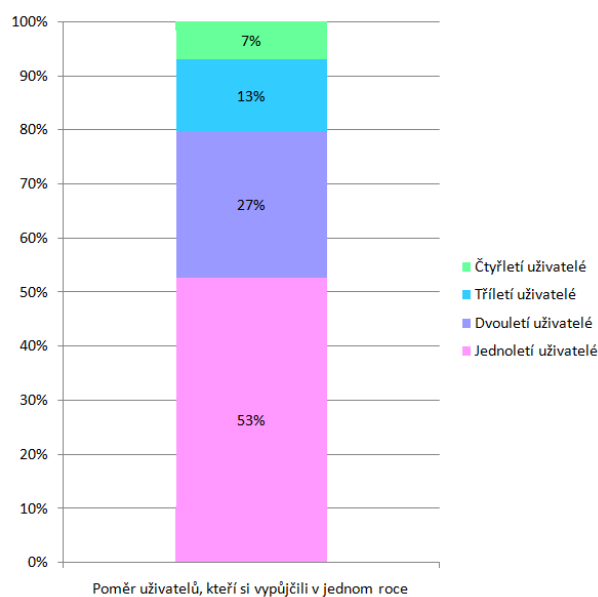
Celkem si v knihovně UKSU za čtyři roky vypůjčilo minimálně jednu knihu 5 278 uživatelů. Z těchto uživatelů více než 50 % knihovnu využívalo pouze v jednom roce. Celých 27 % uživatelů si knihy půjčovalo po dva roky; dále 13 % symbolizuje uživatele, kteří si dokumenty knihovny půjčovali po tři roky, a kolem 7 % uživatelů si knihy vypůjčovalo pravidelně; toto rozložení vyjadřuje Obrázek č. 3.1.

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. Tomuto zjištění by měla být přizpůsobena strategie služeb. Podobně z výsledků

⁸Jako potenciální uživatelé jsou označováni všichni, kteří působí na konkrétních fakultách. Hodnota nám pomáhá normovat výsledky k velikosti fakult.

⁹Za stálé uživatele považujeme ty, kteří byli aktivní po celé sledované období (čtyři roky).

3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.1: Poměr typů uživatelů, vypůjčujících si dokumenty v jednom roce.

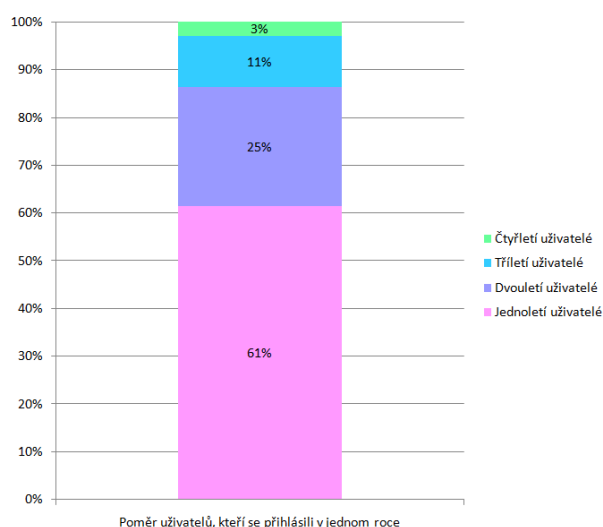
vyplývá, že se po dvou letech promění 80 % uživatelů knihovny. Cirkulace uživatelů je tedy poměrně vysoká, což klade nároky i na informační lekce.

Dále 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. Poměr jednotlivých typů vyjadřuje Obrázek č. 3.2.

Po rozčlenění uživatel na skupiny u výpůjční aktivity zjistíme, že procento tvoří prezenční studenti, pak zaměstnanci, distanční studenti a až nakonec doktorandi. Vzhledem k celkovému počtu studentů na univerzitě jsou procenta spíše orientační. Nicméně, z výsledků vyplývá, že každý šestý zaměstnanec univerzity si chodil do knihovny půjčovat dokumenty po celé sledované období. Ostatní hodnoty nejsou vzhledem k celkovým počtům překvapivé.

Ze stálých uživatelů, kteří se přihlašují do OPACu, jsou pouze 3 % zaměstnanci. Poměr doktorandů výrazně nezměnil a z výsledků 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

3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.2: Poměr typů uživatelů v jednom roce, kteří se přihlásili do OPACu.

jít o nezájem nebo o nízkou informovanost.¹⁰ Nebo je pro zaměstnance prostě jednodušší do knihovny zajít osobně.

3.1.3 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 jen informačně, protože je identifikován pouze IP adresou. Jinými v Carmen zaznamenanými interakcemi se nebudeme zabývat především z toho důvodu, že Clavius jako takový mnoho údajů neukládá. Ty, které ukládá, jsou nepropojitelné nebo mají příliš málo záznamů pro odvozování.

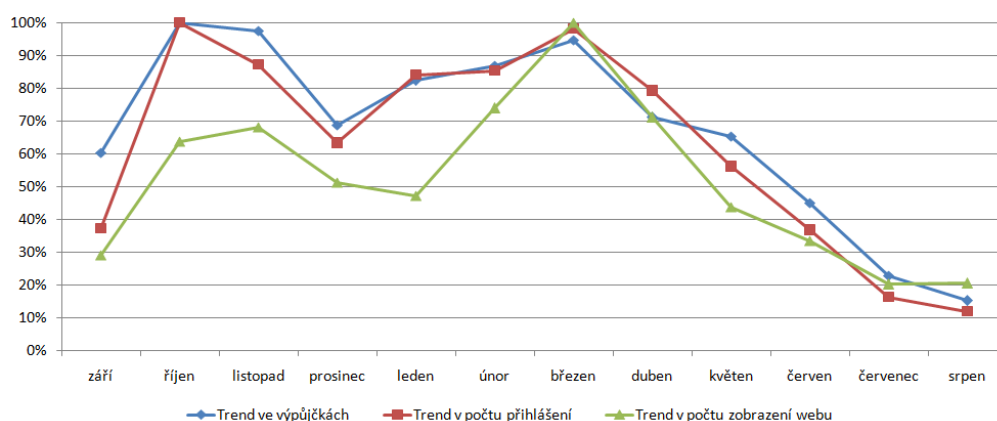
Na Obrázku č. 3.3 můžeme srovnat trendy ve vypůjčování publikací, přihlašování a zobrazování webových stránek, abychom zjistili, jaký průběh tyto činnosti mají. Můžeme odvodit, že mají podobný trend.¹¹ Nejvytíženějšími měsíci jsou říjen a listopad v zimním semestru, v letním potom březen. Trend ve vypůjčkách a trend v přihlášení mají velmi podobný průběh; zatímco trend v počtu zobrazení webu se v počátečních měsících liší.

Nejméně vytíženými měsíci v roce jsou červenec a srpen. Tyto výsledky nejsou překvapivé. Knihovna má v těchto měsících letní režim, který reprezentuje omezená otevírací doba. Pokud by knihovna plánovala změny v provozu, stěhování, školení

¹⁰Na rozdíl od studentů zaměstnanci žádné informační lekce neabsolvují.

¹¹Údaje v grafu byly normovány k maximu, aby byl lépe vidět průběh akademického roku.

3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.3: Trendy ve vypůjčování publikací, přihlašování a zobrazování webových stránek.

personálu, velký softwarový update nebo něco podobného, bylo by dobré provést změny v těchto měsících. 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.¹²

Provedli jsme také podrobnější analýzu přihlašování do OPACu (v týdnu, ve dnech a v hodinách), která v budoucnu knihovně pomůže stanovit otevírací dobu v nové budově.

Tyto závěry nelze odvodit z výpůjček (protože ty podléhají současným otevíracím hodinám) ani z přístupů na web (protože ty jsou zavádějící). Nejméně aktivity uživatelé vykazují v sobotu. Z výsledků analýzy je jasné vidět, jak velký důraz je kladen na začátek týdne, kdy jsou uživatelé zřejmě zvyklí intenzivně pracovat. Průběh všech tří dnů v různých akademických letech je srovnatelný. Díky možné uživatelské aktivitě bychom sobotu doporučovali pro spouštění týdenních statistik, čištění logů a jiné pravidelné procedury, protože je v tento den provoz nejmenší.

Z trendů také vyplývá, že uživatelé jsou zvyklí o víkendů pracovat dlouho. Stejně i na začátku týdne vydrží pracovat déle. Otevírací hodiny knihovny by měly určitě pokrýt dobu od deváté hodiny ranní každý den, kromě soboty a

¹²Data 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. Data o přihlašování by pro správnou interpretaci musela být data vyčištěna od chyb či útoků, které se vyskytly na přelomu roku 2013 a začátku roku 2014. Záznamy o výpůjčkách jsme vyčistili 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.

3 Analýza dat z knihovního softwaru Clavius

neděle, kdy je aktivita nejnižší. Knihovnu by bylo možné zavřít také v pátek odpoledne, když aktivita klesne ke 30 %. Navzdory trendům bychom však knihovně nedoporučovali zavřít knihovnu v pátek úplně; 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.

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, ve stejných časech. V této době by měla zpráva knihovny největší publikum. Zároveň lze předpokládat, že podle analýzy přístupů jednotlivých dní v týdnu budou uživatelé v pondělí až ve středu pravděpodobněji reagovat na emaily a upomínky ze strany knihovny.

Z průzkumu přihlašování do OPACu během let jsme zjistili, že ne vždy se největší počet zástupců z dané skupiny shoduje s nejvyšší aktivitou v přihlašování do OPACu v daném roce. Například, distanční studenti, kterých měla univerzita nejvíce v roce 2011/2012; byli nejvíce aktivní v roce 2012/2013. Celkově se u distančních studentů v prvních třech letech držela aktivita v OPACu držela nad 90 %. Z toho plyne, že ačkoliv distanční studenti nemají takový počet aktivních uživatelů jako jiné skupiny, pokud jsou čtenáři aktivní, přihlašují se poměrně často.

Zcela opačný trend zaznamenávají zaměstnanci. Jejich počty se dramaticky nemění ani v počtu zaměstnanců na univerzitě celkem, ani v počtu aktivních uživatelů v knihovně. Přesto počet jejich přihlášení do online katalogu knihovny klesá, konkrétně v letech 2013/2014 a 2014/2015. Výpůjční aktivita však zůstává stejná. Z výsledků plyne, že tomu tak nebylo vždycky, zájem o OPAC u zaměstnanců upadá až v posledních letech.

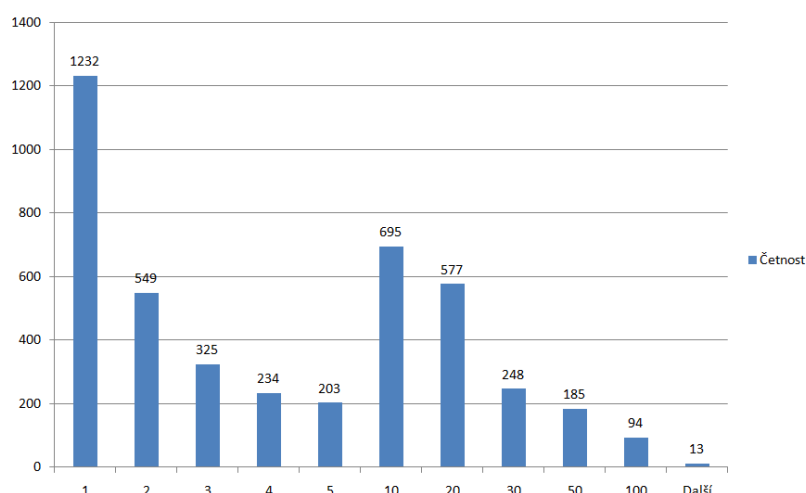
Z výsledků také vyplývá, že celkový počet aktivních uživatelů knihovny klesá.

Měřítkem, které vypovídá o aktivitě přihlašování do OPACu, je návratovost uživatelů. Z histogramu četnosti návratů zjistíme, že největší skupina uživatelů (1 232) navštívila OPAC pouze jednou. Celkem 549 uživatelů se vrátilo jedenkrát, 325 třikrát atd. (viz Obrázek č. 3.4).

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). Pokud bychom tyto sloupce barevně rozdělili dle kategorií, zjistili bychom, že největší zastoupenou skupinou jsou prezenční a následně distanční studenti.

Prezenční studenti se do OPACu přihlašovali nejčastěji v rozsahu 15 až 120 dní. Distanční studenti si udržovali variabilitu od 15 dní dále. Naopak zaměstnanci se

3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.4: Histogram návratovosti v přihlašování do OPACu.

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 výsledků plyne, že se uživatelé přihlašují do OPACu poměrně málo. Pokud se podíváme na statistiky nastavení uživatelského účtu v OPACu, všimneme si, že možnosti jeho nastavení využilo velmi malé množství uživatelů. Atraktivita OPACu pro uživatele je jedním z důležitých bodů, kterým by měla knihovna věnovat pozornost.

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.

Tabulka 3.2: 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

3 Analýza dat z knihovního softwaru Clavius

Pokud bychom chtěli spočítat hodnotu délky výpůjční doby pro jednotlivé kategorie, dostali bychom Tabulku č. 3.2. 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. Doktorandi a zaměstnanci mají standardní dobu výpůjčky nastavenou na jeden rok, proto se u nich doba vracení 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. 3.5. 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. 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.

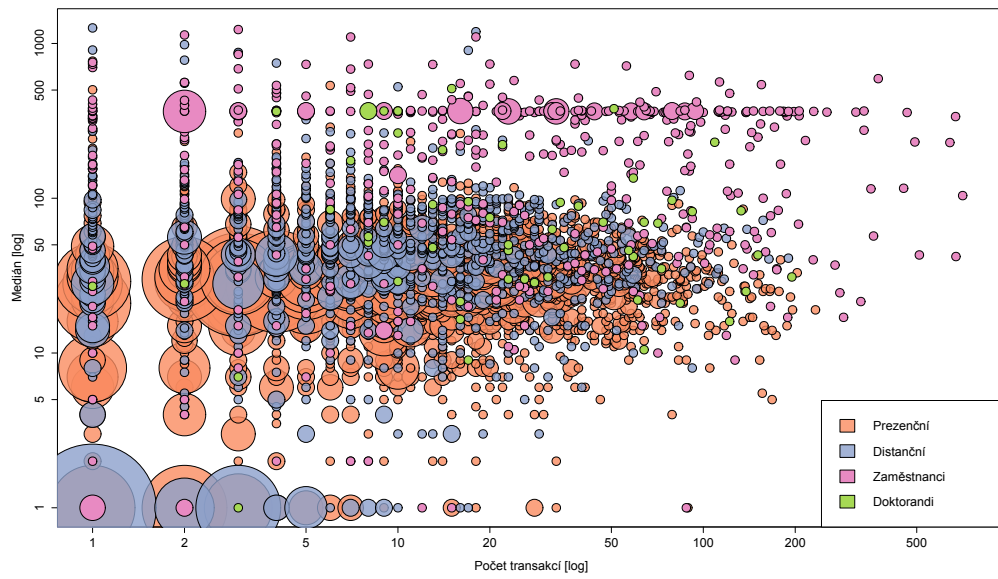
V posledním kroku jsme prověřili, jestli existuje nějaká korelace mezi oběma aktivitami (přihlašování a vypůjčování). V tomto smyslu jsme museli jinak zpracovat tabulku výpůjček. Nekorelovali jsme vypůjčení jako takové, ale aktivitu, kterou lze v tabulce výpůjček registrovat. V průběhu analýzy jsme zjistili, že přihlášení do OPACu a poté provedení akce (vrácení, prodloužení, rezervace, vypůjčení, objednání MVS) spolu koreluje s 95% relevancí při p -hodnotě 0,001, viz Obrázek č. 3.6.

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

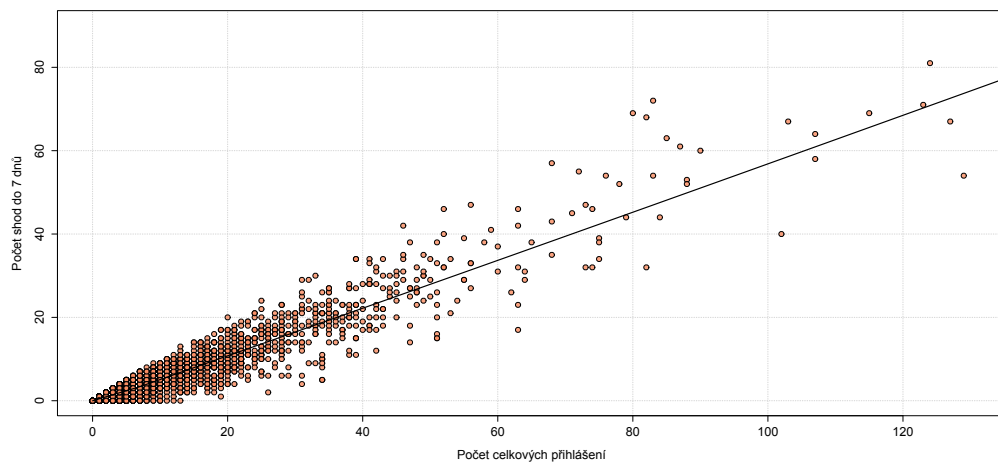
Výsledky jsme dále srovnali se studií Stephena Zweibela a Zachary Laneho.¹³ Jejich závěry jsou takové, že knihy nejdéle drží zaměstnanci. Také u UKSU se tento trend potvrdil. Dále Zweibel a Lane referují o tom, že celková aktivita uživatelů klesá. V UKSU je pokles celkové aktivity zapříčiněn poklesem potenciálních uživatelů, míra jejich aktivity se dramaticky nemění. Dokonce se v některých

¹³ZWEIBEL, 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.

3 Analýza dat z knihovního softwaru Clavius



Obrázek 3.5: Množství půjčených dokumentů vztahované k době výpůjčky a ke kategorii.



Obrázek 3.6: Korelace přihlášení a akce uživatele v tabulce výpůjčky.

3 Analýza dat z knihovního softwaru Clavius

letech objevil pokles počtu aktivních uživatelů, avšak celková aktivita stoupla (rok 2013/2014).

Zweibel a Lane ve své studii zjišťovali, jak se osvědčily změny ve výpůjční politice knihovny. Podobné změny, jako provedla Knihovna University of New York, bychom však UKSU nedoporučovali provést i když si určité množství studentů (distančních i prezenčních) půjčuje dokumenty až na 90 dní. Prodloužení výpůjční doby by u UKSU, podobně jako v Knihovně University of New York, mohlo přinést pokles výpůjček a rozmrzelost studentů.

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.¹⁴

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

¹⁴Naše 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 uložená (záznamy o návštěvách); nepropojitelná s jinými údaji (data o neúspěšném vyhledávání); případně je obtížnost propojení vysoká (obtížnost zpracování tabulky *linear*); interpretace dat je 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. Knihovna by případně mohla provést průzkum uživatelské spokojenosti s OPACem a funkcemi dostupnými pod přihlášením.

Potvrdili jsme trendy ve vypůjčování a vracení publikací. Zjistili jsme, že nejdéle knihy drží vyučující. Naše výsledky jsme poté srovnali s výsledky studie 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 customisation. *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>.
- 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.

Bibliografie

- 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=00000532&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=00000575&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.
- 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>.
- 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.
- 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.

Bibliografie

- 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).
- 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).
- 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.

Bibliografie

- 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/>).
- 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).
- 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/>).

Bibliografie

- 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).
- 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, [cit. 2015-06-18]. Dostupný z: (<http://www.wolfram.com/mathematica/new-in-9/social-network-analysis/>).
- 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).
- 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.
- 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.
- 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. *Knížnica*. 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. 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.
- 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.

Bibliografie

- 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 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).
- 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).
- 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/>).
- 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.

Bibliografie

- 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).
- 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/KnihovnaUniverzitySlezskuniverzity.pdf>).
- 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).
- 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).
- 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>).

Bibliografie

- 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-dlouhodobý-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.
- 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: ([2015-06-01http://www.oclc.org/research/activities/worldmap.html](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 ŠAVELKA. 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.