
Použití data miningu ve „fyzických“ knihovnách

Usage of data mining in „real“ libraries

Anna Janíková

Ústav informatiky Filozoficko-přírodovědecké fakulty, Slezská univerzita v Opavě

Abstrakt:

Článek obsahuje přehledovou studii na téma použití data miningu ve „fyzických“ knihovnách. Téma se vymezuje slovem fyzický vůči knihovněm digitálním, které mají své vlastní možnosti a omezení při použití vytěžování z dat. Z tohoto důvodu nejsou digitální knihovny v tomto článku zahrnuty. Studie byla provedena nad odbornými články z oblasti Knihovnictví a informační vědy a shrnuje poznatky do několika témat, rozvádějících použití data miningu v knihovnách v zahraničí. Většina studií, o nichž článek pojednává, pochází z akademických knihoven. Samostatná část, pokrývající vybraná periodika, je věnována pracím o data miningu v Čechách a na Slovensku. Časové období sběru a zpracování je ukončeno rokem 2015.

Klíčová slova: použití, data mining, knihovna, odborné články

Abstract:

Text deals with data mining used in libraries. It focuses on „physical“ libraries as they can use data mining in many processes but chooses its scope opposite to digital libraries where data mining has different variety in usage. This is why papers concerning digital libraries are excluded from this summarizing study. All the papers were selected based on concern with Library and Information Science and were divided into several groups. Most of the studies come from academic environment. Papers from czech and slovak environment are described separately. Time limitation for inclusion of the papers is up to year 2015.

Keywords: usage, data mining, library, papers

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.

O DM se knihovnický svět začal zajímat na konci 90. let 20. století; začaly se objevovat první články, které ukazovaly, že DM je metoda použitelná také v knihovnách. Konceptem data miningu je využívání dat, sesbíraných a skladovaných z jiných důvodů, k získání nových znalostí. 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.

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 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 závěry¹.

Je důležité také zmínit, že název naší studie se nesnaží zavést termín „fyzická“ knihovna. „Fyzická“ pro nás znamená knihovna fyzicky reprezentovaná fondem, budovou, ve které je umístěna, pracovníky, kteří v ní působí a službami, které se z části odehrávají ve fyzickém, z části v digitálním světě.

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 knihovny² oproti tomu umožňují analyzovat data o veškeré elektronické interakci; má tedy více možností. Vzhledem k výše zmíněným omezením nás nepřekvapí, že procento zastoupení digitálních knihoven v problematice DM analýz je větší.³ Předmětem našeho zájmu budou v rámci tohoto článku i přesto výlučně knihovny „fyzické.“

Vytěžování z dat lze provádět nad různými typy dat. Ta mohou pocházet z turniketů, ze sociálních sítí, webových logů, selfchecků, z výpůjčního protokolu atd. Pro DM analýzy se tato data mohou spojovat⁴ a ukládat do datových skladů. Pro účely popisu využití DM ve „fyzických“ knihovnách je nutné studie rozdělit do několika celků. Rozdělení podle typu dat nebo metod není vhodné. V některých člancích se použití metod různí, popřípadě se navzájem kombinují. Proto nám přijde nejvhodnější je dělit podle využití. Články v této studii proto rozdělíme do těchto kategorií: Akvizice s využitím DM, Analýzy záznamů v katalogu,

¹ Je žádoucí, aby tato osoba úzce spolupracovala s knihovníkem, který mu pomáhá orientovat se v procesech v knihovně a formulovat pro praxi zajímavé otázky.

² Jejich provoz často souvisí s existencí fyzické knihovny, ale ne nevyhnutelně.

³ 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.)

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

Systémy pro podporu rozhodování, Rozdělování financí, Služby a data mining, Evaluace OPACu a webu knihovny, Průzkum chování uživatelů pomocí DM (tuto kategorii dále rozčleníme na Průzkum vlivu knihovny na retenci studentů a Analýzu využívání fondu) a jako poslední Zpracování fondu. Některé články by bylo možné zařadit do více okruhů, ovšem dáváme přednost specifickému před obecnějším a konkrétní aplikaci před vyšším smyslem.

Zdroji, ze kterých budeme čerpat, jsou dílčí výzkumné studie zahrnuté 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, Knižnica a ITlib).⁵ Českým a slovenským textům je věnována samostatná část našeho pojednání.

Akvizice s využitím DM

K akvizici, provozované pomocí DM se vztahuje termín „daty řízená akvizice“^{6, 7} Se snižováním rozpočtu knihovnám se zvýšil tlak na akviziční činnost a to v Čechách⁸, podobně jako v zahraničí. Zahraniční knihovny tak při hledání ideálního způsobu, jak obohacovat svůj fond, stále častěji sahají k analýze dat.

S termínem „daty řízená akvizice“ souvisí také dříve známý termín uživatelem řízená akvizice, (v angličtině patron-driven acquisition, user-driven acquisition či demand-driven acquisition),⁹ kdy uživatelé sami navrhnou knihovně koupit dokumentu.¹⁰ „Daty řízená akvizice“ 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 tematickými 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 model slouží pouze jako podpora rozhodování, umožňuje lépe získávat potřebné informace. Zároveň 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.¹¹ O rok později v jiném článku doplnil svůj koncept o použití asociačních pravidel, které by usnadňovaly objevení nových závislostí.¹²

⁵ Zpracované jazyky jsou angličtina, čeština a slovenština. Výběr článků dle klíčových slov byl omezen pojmy „data mining“ a „knihovna“.

⁶ Ustálený český termín k pojmu „data-driven acquisition“ zatím neexistuje.

⁷ MITCHELL, Anna. Data Driven Acquisition in the Library. *YouTube* [online]. Publikováno

9. 11. 2015, [cit. 2016-01-02]. Dostupný z: <https://youtu.be/CKbtmRCRemci>

⁸ Ministerstvo kultury ČR. *Koncepce rozvoje knihoven ČR na léta 2011–2015 včetně internetizace knihoven* [PDF]. [cit. 2016-02-01]. Dostupný z: http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce_rozvoje_knihoven_2011-2015.pdf, s. 1.

⁹ MITCHELL, Anna. Data Driven Acquisition in the Library. *YouTube* [online]. Publikováno

9. 11. 2015, [cit. 2016-01-02]. Dostupný z: <https://youtu.be/CKbtmRCRemci>

¹⁰ Česky se pro tyto dokumenty vžil termín dezideráta. (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

¹¹ 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.

¹² WU, Chien-Hsing, Tzai-Zang LEE and 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.

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 tematickými 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 model slouží pouze pro podporu rozhodování při akvizici, ne jako univerzální měřítko kvality akvizice.

Podobný systém jako Wu navrhuje také Kao, z databáze extrahuje data o výpůjčkách, které následně dělí do kategorií. V druhé fázi akademičtí pracovníci pomocí dotazníků hodnotí 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.¹³

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 vypůjčovnosti dle sledovaných období od roku 2000 do roku 2013 a od roku 2012 do roku 2013. Dle vypůjčovnosti pak rozhodoval, jak se mění vypůjčovnost a akviziční činnost v závislosti na oborech, a jestli jsou nějaké rozdíly mezi časovými obdobími. Podobností mezi tématy se nezabýval, ale zjistil, že složení fondu ve velké míře odpovídá frekvenci vypůjčovnosti publikací. Potvrdil zvyky ve vypůjčování knih jednotlivými obory (např. sociální a přírodní vědy) a doporučil doplnit fond o publikace v konkrétních kategoriích, které nebyly v knihovně tak bohatě zastoupeny.¹⁴

Nevýhody předchozích pojetí postrádá Yangův na webu založený systém pro doporučování nákupu knih. Ten využívá slov z vyhledávání, která nevrátila žádné hity.¹⁵ Tato slova pak řadí dle frekvence a podle společného výskytu sestavuje tezaurus slov. Slova pak spojuje s kategoriemi knih a popisy od prodejců; automaticky tak sestavuje seznamy doporučených knih ke koupi.¹⁶

Podobně jako Wuův 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.¹⁷

Analýzy záznamů v katalogu

Průzkum Williama Dentona se zaměřuje na analýzu dat z katalogu. Snaží se vizualizovat informace, schované v záznamech milionů knih, kterými knihovny disponují. Použil data těchto knihoven: San Francisco Public Library, Toronto Public Library, University of Toronto,

¹³ KAO, S.-C., H.-C. CHANG and C.-H. LIN. Decision support for the academic library acquisition budget allocation via circulation database mining. In: *Information Processing and Management*. 2003, roč. 39, č. 1, s. 133–147. ISSN 0306-4573.

¹⁴ KITAJIMA, Mitsuro, Kensuke BABA and 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.

¹⁵ 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)

¹⁶ YANG, Shih-Ting. An active recommendation approach to improve book-acquisition process. In: *International Journal of Electronic Business Management*. 2012, roč. 10, č. 2, s. 163–173. ISSN 1741-5063.

¹⁷ 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.

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.¹⁸

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 samozřejmé, 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 a době. Hledal zajímavé výkyvy v datech (v počtu publikací, počtu svazků, počtu překladů a počtu knihoven, které jsou členy OCLC). Vybral si šest zemí, Bolívii, Chile, Německo, Polsko, Jihoafrickou republiku a Thajsko. Na první pohled jde o velmi rozdílné země, což to bylo od počátku záměrem. Autor se zaměřil na diverzitu a jedinečnost jednotlivých národních sbírek, nikoliv na statistické zobecnění. 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.¹⁹

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, jsou ekonomicky nákladné, častým užíváním se poškozují, nebo je nutné je zachovat pro další zkoumání.²⁰ Stejným způsobem jako ve jmenované studii by bylo možné promítnout tato hlediska do kritérií pro klasifikaci²¹ záznamů knih v katalogu knihovny.

Systemy pro podporu rozhodování

Existuje několik odborných článků, ve kterých se autoři snaží vymezit, s jakými úkony může DM vedoucím knihovny pomoci.^{22, 23} Autorem jednoho z nich je Kevin Cullen,²⁴ 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

¹⁸ DENTON, William. On Dentographs, A New Method of Visualizing Library Collections. *Code4Lib Journal*. 2012, roč. 16, s. 171–179. ISSN 1940-5758. Dostupný také z: <http://journal.code4lib.org/articles/6300>

¹⁹ DICKEY, Timothy J. Books as Expressions of Global Cultural Diversity. *Library Resources and Technical Services*. 2011, roč. 55, č. 3, s. 148–162.

²⁰ SILIPIGNI CONNAWAY, Lynn, Edward O'NEIL and Chandra PRABHA. Last Copies: What's at Risk?. *College and research libraries*. 2006, roč. 67, č. 4, s. 370–379. ISSN 0010-0870. Dostupný také z: <http://www.oclc.org/research/publications/archive/2006/connaway-crl07.pdf>

²¹ Myšleno provedení klasifikace jako data miningového zadání.

²² SCHULMAN, Sandy. Data mining: Life After Report Generators. *Information Today*. 2003, roč. 15, č. 3, s. 52. ISSN 0363-0277.

²³ NICHOLSON, Scott and Jeffrey M. STANTON. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In: H. NEMATİ and C. BARKO eds. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey (PA): Idea Group Publishing, 2003, s. 247–262.

²⁴ CULLEN, Kevin. Delving into data. *Library Journal*. 2005, roč. 130, č. 13, s. 30–32. ISSN 0363-0277.

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.²⁵ 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.²⁶

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“.²⁷ 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 navrhuje, 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 uvádí rychlost, nízkou cenu a možnost rozšíření výzkumu o mnohá hlediska. Jako nevýhodu pak vidí problém zachování soukromí uživatelů.²⁸

Také Nicholson hovoří o DM v knihovnách jako o efektivní podpoře rozhodování²⁹ a to zvláště při strategickém plánování.³⁰ Navrhuje sjednotit všechna data, která má knihovna k dispozici, do datového skladu. Nad tímto skladem pak navrhuje provádět DM k získání přehledu o využívání knihovny a k zajištění finanční podpory ze strany zřizovatele pro stávající i nové služby.

Nicholson navrhuje vytěžování z dat použít k zjišťování chování uživatelů a jejich preferencí, k doporučování literatury pomocí analýzy nákupního košíku,³¹ k průzkumu chyb v systému pro vyladění služeb, k nalézání vzorů chování při ztrátě drahých či vzácných dokumentů pro úpravu výpůjčního řádu, k předvídání 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.³²

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. Podobně vyhodnocení málo půjčovaných

²⁵ 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.

²⁶ CULLEN, Kevin. Delving into data. *Library Journal*. 2005, roč. 130, č. 13, s. 30–32. ISSN 0363-0277.

²⁷ Neboli „evidence-based librarianship“ přesný český termín zatím chybí.

²⁸ 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.

²⁹ NICHOLSON, Scott. The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries*. ALA, 2003, no. 4. ISSN 2163-5226.

³⁰ NICHOLSON, Scott and 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.

³¹ Tím se myslí zjišťování, které zboží zákazníci kupují nejčastěji spolu. V našem případě to může znamenat, které dokumenty si uživatelé půjčují dohromady.

³² NICHOLSON, Scott and 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.

dokumentů může ukazovat na to, že byly třeba špatně zařazeny či popsány při katalogizaci.³³ Povzbuzuje také 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.³⁴ Podrobný metodický popis k procesu DM v knihovnách, který by vedl k lepšímu rozhodovacímu procesu, popisuje Nicholson v dalším ze svých článků.³⁵

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 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 group³⁶ a think-aloud³⁷ 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.³⁸

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;³⁹ 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.⁴⁰ Tento systém je samozřejmě výhodný, protože ukazuje celistvý obraz organizace a jejích služeb, může být však málo přehledný; zvláště, pokud porovnává více subjektů.

³³ Podobně může DM analýza posloužit při relokaci knih do skladů a jejich odepisování.

³⁴ NICHOLSON, Scott and 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.

³⁵ NICHOLSON, Scott. The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries*. ALA, 2003, no. 4. ISSN 2163–5226.

³⁶ 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* [online]. Office of Assessment, Duke Trinity College of Arts and Sciences, c2005 [cit. 9. 2. 2016]. Dostupné z:

https://assessment.trinity.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf)

³⁷ 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 vyžadují, 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)

³⁸ 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.

³⁹ 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/el/1456/jaro2013/MPV_VZVP/um/33148301/Studijni_text_nakladove_vystupove_metody_CMA_CEA_CUA.pdf)

⁴⁰ 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.

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 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.⁴¹

Konkrétní příklad aplikace DM do rozhodovacího procesu představuje práce dvojice autorů Arise Meletiou a Anthi Katsirikou, kteří s úmyslem vylepšit strategické plánování knihovny provedli vytěžování nad daty o publikacích a uživateli s úmyslem najít korelace, které by pomohly knihovně v jejím rozvoji.⁴²

Je jisté, že některé články z této části bychom mohli přeradit také do jiných sekcí, například do části „Služby a data mining“ nicméně, byly zařazeny do části o systémech pro podporu rozhodování, především pro jejich globální povahu.

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, uživateli 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:⁴³

- 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.⁴⁴

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

⁴¹ PAPAVALASOPOULOS, Sozon and Marios POULOS. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, vol. 33, iss. 4/5, s. 261–271. ISSN 0143-5124.

⁴² MELETIOU, Aristeidis and Anthi KATSIRIKOU. Methodology of analysis and interrelation of data about quality indexes of library services by using data and knowledge mining techniques. *Library Management*. 2009, vol. 30, no. 3. s. 138-147. ISSN 0143-5124. DOI 10.1108/01435120910937311.

⁴³ 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.

⁴⁴ 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.

ohled na situaci akademické knihovny. Zajímavost jeho přístupu tkví v tom, že ke stanovení rozdělení využívá jedné jednoduché rovnice.⁴⁵

Služby a data mining

Využití DM v oblasti služeb je poměrně široké téma a 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.⁴⁶ Podařilo se jí tak poskytnout univerzitě zcela unikátní službu.

Druhou oblastí, kde je DM využíván, je zlepšování služeb. V tomto smyslu použila Nadaleen Tempelman-Kluit spojení vytěžování z dat s tvorbou person, které jsou nástrojem user-centered design.⁴⁷ 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. Těmto byly přiřazeny čtyři osoby⁴⁸ reprezentující uživatele dané knihovny. Potenciál výsledků netkví v lepší orientaci knihovníků ve skupinách uživatelů, ale zvláště ve zlepšování výukových kurzů, webových stránek a nástrojů pro federativní vyhledávání dle konceptů user-centered design.⁴⁹

Zajímavou aplikaci DM metod provedli Fontane a Finnell,⁵⁰ 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, jež se na univerzitě vyučují). Snažili se také zjistit, jaké jsou vzorce využívání fondu, zda uživatelé, kteří se ptají, také doporučené zdroje využívají.⁵¹

Během průzkumu autoři objevili mnohé závislosti mezi vyučovanými kurzy a referenčními dotazy, přicházejícími 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.⁵²

⁴⁵ 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.

⁴⁶ 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.

⁴⁷ „Na uživatele zaměřený design“ označuje takové postupy ve webovém designu, které zajišťují jeho použitelnost.

⁴⁸ Takzvané „data-driven personas“.

⁴⁹ TEMPELMAN-KLUIT, Nadaleen and Alexa PEARCE. Invoking the User from Data to Design. *College and Research Libraries*. 2014, roč. 75, č. 5, s. 616–640. ISSN 0010-0870. Dostupný také z: <http://dx.doi.org/10.5860/crl.75.5.616>

⁵⁰ FINNELL, Joshua and Walt FONTANE. Reference Question Data Mining: A Systematic Approach to Library Outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.

⁵¹ FINNELL, Joshua and Walt FONTANE. Reference Question Data Mining: A Systematic Approach to Library Outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.

⁵² FINNELL, Joshua and Walt FONTANE. Reference Question Data Mining: A Systematic Approach to Library Outreach. *Reference and User Services Quarterly*. 2010, č. 3, s. 278–286.

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.⁵³

Příkladem explorační studie je článek Eve-Marie Lacroix a Marie Elizabeth Collins.⁵⁴ 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.^{55,56}

Přestože členů používajících DOCLINE přibývalo, 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 problematičtější se dostat, proto počet požadavků na MVS klesá (v celkových číslech), avšak zároveň stoupá stáří požadovaných dokumentů.⁵⁷ Takto se autorkám podařilo odhalit zajímavé trendy ve využívání služeb knihoven.

Evaluace OPACu a webu knihovny

Transakční analýzy OPACu (podobně webové analýzy) mohou sloužit analýze uživatelského chování v elektronickém prostředí katalogu (či webových stránek knihovny), k evaluaci jeho funkcí a zlepšování designu. Tento přístup je nasnadě hlavně proto, že pohyb uživatele v katalogu je vždy automaticky zaznamenáván knihovním softwarem.⁵⁸ 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 proklíkává cestu katalogem, a běžně se

⁵³ 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.

⁵⁴ LACROIX, Eve-Marie and 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.

⁵⁵ LACROIX, Eve-Marie and 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.

⁵⁶ Tento systém používají lékařské a nemocniční knihovny v USA a v Kanadě pro zadávání požadavků na MVS. Jako typ dokumentu jsou nejvíce požadovány odborné články. (LACROIX, Eve-Marie and 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.)

⁵⁷ LACROIX, Eve-Marie and 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.

⁵⁸ Samozřejmě vždy záleží na konkrétním typu softwaru, jaké údaje zaznamenává. V případě, že jde o knihovní software, který je distribuovaný jako open source, lze doplnit trekovací kódy, které budou chování uživatelů sledovat. V případě, že to není možné, musí knihovna přizpůsobit analýzy situaci.

potkává s chybovými hláškami. Následkem tohoto zjištění bylo zjednodušení jazyka vyhledávacího rozhraní, úvodních obrazovek i designu katalogu. Změnilo se i pořadí vyhledávacích možností. 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.⁵⁹

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, webové stránky), 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), připravil koncept nové služby zpřístupňování dokumentů za poplatek mimo univerzitu. Dále navrhoval vycházet z výsledků analýzy při redesignu stránek knihovny. Přestavění mělo způsobit, že se uživatelé rychleji dostanou k obsahu, který hledají. Dále změnil umístění některých málo užívaných služeb a hlavně vyhledávacího boxu.⁶⁰

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.⁶¹

Hsiao-Tieh Pu a Chyan Yang navrhuje DM využít ke zlepšení předmětového vyhledávání v OPACu knihovny a to nalezením asociací klasifikačních tříd. Tyto asociace se snaží vyhledat na základě znaků selekčního jazyka dokumentů, které byly vypůjčeny společně jedním uživatelem. Jde o to, že ne vždy jsou třídy systematického selekčního jazyka (v článku je jmenováno Deweyho desetinné třídění a Čínské desetinné třídění), které vystihují obsah dokumentu, uspořádány hierarchicky. Oba autoři se snaží přistupovat k problému takovým způsobem, aby eliminovali chyby způsobené různým počtem dokumentů v jednotlivých třídách, oblíbenost či neoblíbenost některých tříd a případně malou vypůjčovanost materiálů malým počtem uživatelů. Z výsledku experimentu pak vycházejí při konstrukci schématu, zlepšujícího OPAC a jeho předmětové vyhledávání.⁶²

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. Zmiňované 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.

⁵⁹ 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.

⁶⁰ BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *Journal of the Medical Library Association*. 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>

⁶¹ SHIEH, Jiann-Cherng. From website log to findability. *The Electronic Library*. 2012, roč. 30, č. 5, s. 707–720. ISSN 0264-0473.

⁶² PU, Hsiao-Tieh and Chyan YANG. Enriching user-oriented class associations for library classification schemes. *The Electronic library*. 2003, vol. 21, no. 2. ISSN 0264-0473. DOI 10.1108/02640470310470507.

Publikace ve většině případů pocházejí od knihovníků z akademických pracovišť. 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.⁶³

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.⁶⁴ 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.⁶⁵

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. Zahrnuti 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 nenaplnuje svým fondem dostatečně jejich informační požadavky (v dané knihovně bylo zaznamenáno nejvíce žádostí o MVS), a proto v ní studenti tráví méně času.⁶⁶

Kromě těchto skutečností bylo objeveno ještě několik dalších závislostí, které nebyly očekávány a vyžadovaly by hlubší prozkoumání.⁶⁷ 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).

⁶³ JOHNSON, Doug. A Data Mining Primer and Implications for School Library Media Specialists. *Knowledge Quest*. 2004, roč. 32, č. 5, s. 32–35.

⁶⁴ 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>

⁶⁵ 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>

⁶⁶ 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.

⁶⁷ 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.

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.⁶⁸ 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 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.⁶⁹

Autorem poslední studie v této podkapitole je Sharon Weiner. Zabývala se otázkou, zda je knihovna spoluvůrcem 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 spoluvůrcem reputace univerzity. Jak však dále sama uvádí, oproti jiným součástem bohužel silně podfinancovaným.⁷⁰

Analýza využívání fondu

Tato část nese název Analýza využívání fondu z jednoduchého důvodu, knihovny disponují kromě tištěných dokumentů také přístupem k odborným databázím, případně půjčují elektronické knihy. Část níže zmíněných článků provedla analýzy právě na těchto datech, protože elektronický přístup samozřejmě generuje větší množství dat.

První z článků, jehož autory jsou Stephen Zweibel a Zachary Lane, analyzuje data z výpůjček fyzických knih a elektronických interakcí. Data byla analyzována za účelem nalezení vzorců chování a ověření výpůjční politiky knihovny. Podařilo se nalézt tři hlavní trendy ve vypůjčování dokumentů. Zaprvé, na nejdělsí 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 prodloužování výpůjček (který se dal očekávat), ale také pokles výpůjční aktivity celkově.

⁶⁸ SORIA, Krista M., Jan FRANSEN and 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.

⁶⁹ EMMONS, Mark and 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.

⁷⁰ 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.

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.^{71,72}

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. Podařilo se jim odvodit výpůjční profily pro dané oddělení a danou kategorii čtenáře. Důsledky zjištění byly validní jak pro akviziční, tak pro katalogizační oddělení.⁷³

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 a díky definovaným rozhraním by jedné skupině nebyly doporučovány publikace od skupiny druhé.⁷⁴

Podobným způsobem, jako se Whitmire⁷⁵ a Soria⁷⁶ snažily spojit využívání knihoven a studijní výsledky, se Sue Samson ve své práci pokusila spojit známky studentů s využitím elektronických informačních zdrojů přístupných v knihovně. V tomto případě byla potvrzena hypotéza, že lepší známky (respektive lepší průměrné hodnocení studenta) jsou asociovány s používáním knihovních e-zdrojů.⁷⁷ 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 knihovny více, a které méně. V průběhu se vynořily další otázky, které by bylo dobré dalším průzkumem prověřit. Například vyučující dle výsledků využívají knihovních e-zdrojů nejvíce ze všech – zajímavé by bylo zjistit, jaký vliv to má na jejich vědecké publikování.⁷⁸

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

⁷¹ ZWEIBEL, Stephen and 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.

⁷² Článek bohužel nedostatečně vystihuje, co do těchto interakcí zahrnuje.

⁷³ LEONARD, Michelle F., Stephanie C. HAAS and Vernon N. KISLING. Metrics and Science Mohograph Collections at the Marston Science Library, University of Florida. *Issues in Science and Technology Librarianship*. 2010, č. 62, s. 1. ISSN 1092-1206. Dostupný také z: <http://dx.doi.org/10.5062/F4PC308T>

⁷⁴ 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.

⁷⁵ 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.

⁷⁶ SORIA, Krista M., Jan FRANSEN and 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.

⁷⁷ Znovu vyvstává otázka, zda úspěšnější studenti využívají elektronických informačních zdrojů 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.

⁷⁸ 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.

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.⁷⁹ Ačkoli je obecná definice známá, není jasné, 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.⁸⁰

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.⁸¹

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 by 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.⁸²

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. Autoři provedli kategorizaci pětkrát, pro každou knihovnu zvlášť,

⁷⁹ AHMAD, Pervaiz, Mark BROGAN and 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, s. 36.

⁸⁰ AHMAD, Pervaiz, Mark BROGAN and 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

⁸¹ AHMAD, Pervaiz, Mark BROGAN and 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.

⁸² BOTER, Jaap and Michel WEDEL. *User categorization of public library collections. Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. ISSN 0740-8188. Dostupný také z: <http://dx.doi.org/10.1016/j.lisr.2005.01.004>

a zdůrazňují, že ve všech by měl být aplikován jim příslušný model, který bude odpovídat konkrétnímu složení uživatelů. Nedoporučovali příliš výsledky generalizovat.⁸³

Oproti tomu otázku řazení knih v akademické knihovně řešili výzkumníci z novozélandské Univerzity Waikato Sally Jo Cunningham a Eibe Frank. Podobně jako Hsiao-Tieh Pu a Chyan Yang⁸⁴ aplikovali metodu nákupního košíku na data o výpůjčkách. Snažili se nalézt mezi nimi asociace a odhadnout vzdálenost, kterou musí uživatel urazit, aby získal jím vybrané dokumenty. Šlo jim především o ověření funkčnosti rozmístění dokumentů v knihovně.⁸⁵

Zpracování fondu

Do této kategorie spadají analýzy, týkající se zpracování bibliografických záznamů. Tyto záznamy, ačkoliv se jeví globálně ne příliš významné, mohou hrát pro knihovnu důležitou roli, jak naznačil článek Lynn Silipigni Connaway a Timothy Dickeyho.⁸⁶

Protože se jedná o výzkumníky spojené s činností OCLC, mohli využít datových zdrojů, kterými OCLC disponuje. Jejich hlavním záměrem bylo vytvořit funkční databázi nejdůležitějších nakladatelství z celého světa. Jejich předchůdci do jisté míry potvrdili, že není produktivní tyto informace zpracovávat ručně a tak již předem počítali se zapojením shlukování⁸⁷ pomocí prefixu ISBN. Tento postup poměrně zjednodušil první krok; avšak při vytěžování dat z různých databází se objevil problém, související se zadáním či opravou údajů, kdy bylo nutné využít pomoci člověka. Pro ulehčení jeho úlohy využili programu, který prováděl porovnání tzv. trigramů (shluků tří písmen) z názvů s využitím fuzzy logiky. Finální návrh byl pak poskytnut ke kontrole lidskému oku a výsledkem byl vznik Publisher Name Authority File Database.⁸⁸

České a slovenské studie

Bohužel, data mining je v českých a slovenských knihovnách poměrně opomíjeným tématem. Existují pouze dvě práce, které přináší výsledky aplikace DM na knihovní data.⁸⁹ Jednou je případová studie Pavlína Omastové, Radky Římanové a Ctibora Škuty, kteří v termínech z vyhledávání Národní technické knihovny hledali nové vhodné termíny pro Polytematický strukturovaný heslář.⁹⁰ 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.

⁸³ BOTER, Jaap and Michel WEDEL. *User categorization of public library collections. Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. ISSN 0740-8188. Dostupný také z:

<http://dx.doi.org/10.1016/j.lisr.2005.01.004>

⁸⁴ PU, Hsiao-Tieh and Chyan YANG. Enriching user-oriented class associations for library classification schemes. *The Electronic library*. 2003, vol. 21, no. 2. ISSN 0264-0473. DOI 10.1108/02640470310470507.

⁸⁵ CUNNINGHAM, Sally Jo and Eibe FRANK. Market basket analysis of library circulation data. In: *Proceedings of the 6th international conference on neural information processing*. Perth, 1999. s. 825–830. Dostupný z: <http://www.cs.waikato.ac.nz/~eibe/pubs/SJ-EF-Market-Basket.pdf>

⁸⁷ SILIPIGNI CONNAWAY, Lynn, Timothy J. DICKEY. An experimental authority file and prototype application author. *Library Resources & Technical Services*. 2011, vol. 55, iss. 4. s. 182–194. ISSN 0024-2527.

⁸⁷ Jako DM metody.

⁸⁸ SILIPIGNI CONNAWAY, Lynn, Timothy J. DICKEY. An experimental authority file and prototype application author. *Library Resources & Technical Services*. 2011, vol. 55, iss. 4. s. 182–194. ISSN 0024-2527.

⁸⁹ Jednou z nich je bakalářská práce. Ačkoliv se při zpracování přehledu primárně zaměřujeme na odborné studie, kvůli nedostatku zdrojů jsme výjimečně sáhli také po závěrečných pracích. A z nich uvádíme jediného zástupce.

⁹⁰ OMASTOVÁ, Pavlína, Radka ŘÍMANOVÁ a Ctibor ŠKUTA. Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety. In: *Sborník z 19. konference, konané ve dnech 13.–15. září 2011 v Českých Budějovicích*. Ostrava: Sdružení knihoven ČR, září 2011, s. 101–109. ISBN 978-80-86249-62-9.

Druhou aplikaci DM na data knihovny představuje práce Michaly Soškové, která zpracovávala pomocí nástrojů SQL databáze a programu Excel údaje ze souborného online katalogu Masarykovy univerzity. 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.⁹¹

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.⁹² Její článek poukazuje na možnosti využití DM při komplexním řízení kvality, které jsme jako praktické využití zmínili také v části Systémy pro podporu rozhodování.⁹³

O něco kratší práci s vysvětlením pojmu bibliomining a několika příklady publikovala na Slovensku také Katarína Lalahová.⁹⁴

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.

Závěr

Ačkoliv se data miningové zpracování dat v českých a slovenských knihovnách příliš nerozšířilo, díky zahraničním autorům máme k dispozici celou řadu studií, kterými se lze při provádění DM v knihovnách inspirovat. Jedná se o studie v oblastech akvizice a měření výkonu knihovny, při zavádění služeb nových i evaluaci služeb stávajících, dále při jednání o financích, zlepšování OPACu knihovny nebo jejích webových stránek, případně při obhajobě jejích výdajů, a v dalších oblastech.

Nejvíce studií využívajících DM se věnovalo problematice uživatelského chování, a to analýze toho, jakým způsobem čtenáři využívají fond, nebo u akademických knihoven, jak ovlivňuje aktivita uživatelů v knihovně retenci studentů. Jak bylo zmíněno výše, možnosti využití dat se různí podle toho, která data jsou k dispozici, což také znamená, dle používaného automatizovaného knihovního softwaru a případně dalších softwarových produktů.

Pokud se knihovna rozhodne svá data zpracovávat pomocí DM, bude se muset kromě problémů jako vyčištění, propojení a uložení dat, potýkat také se zajištěním podmínek pro opakování analýz a se správnou interpretací výsledků. Zvláště v posledním bodě bude spoléhat na správnost uložení dat, korektnost provedení analýzy a schopnosti člověka, který je prováděl.

⁹¹ 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

⁹² SEDLÁČKOVÁ, Beáta. Data mining a jeho uplatnění při poznávání knihovnických jevů a zákonitostí. *Knihovna*, 2007, roč. 8, č. 11–12, s. 21–22. ISSN 1335-7026, ISSN 1336-0965. Dostupné z:

http://www.snk.sk/images/snk/casopis_knihovna/2007/november-december/21.pdf

⁹³ PAPAVALASOPOULOS, Sozon and Marios POULOS. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, vol. 33, iss. 4/5, s. 261–271. ISSN 0143-5124.

⁹⁴ LALUHOVÁ, Katarína. Bibliomining – data mining pre knižnice. *ITlib: Informačné technológie a knižnice*. 2007, č. 4. ISSN 1336-0779. Dostupné z: <http://www.cvtisr.sk/itlib/itlib074/lalahova.htm>

Zmíněné studie popisují především, ale ne výlučně, aplikaci DM (či vytváření teoretických modelů) v akademickém prostředí. Dle zjištění přehledové studie můžeme soudit, že neakademické knihovny mají s DM zpracováním dat ze svých procesů větší problémy. Důvodů může být několik, nedostatek financí, nedostatečná informovanost či nedostatek kvalifikovaných pracovníků.⁹⁵ Ovšem, také se může jednat o motivaci akademických knihoven co nejvíce o svých výsledcích a nových postupech spravovat ostatní.

Určitě ale doufáme, že tato přehledová studie pomůže zvýšit povědomí o DM zpracování dat v tuzemských knihovnách; v dalších pracích se hodláme věnovat konkrétnějším aplikacím DM s příkladem konkrétní akademické knihovny.

⁹⁵ Pro získání podrobnějších informací o tomto problému bychom museli provést specializovaný průzkum.

Literatura

- AHMAD, Pervaiz, Mark BROGAN and 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, s. 36.
- BANERJEE, Kyle. Is Data Mining Right for Your Library?. *Computers in Libraries*. November/December 1998, s. 28–31.
- 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.
- BOTER, Jaap and Michel WEDEL. *User categorization of public library collections. Library & information science research*. 2005, roč. 27, č. 2, s. 190–202. ISSN 0740-8188. Dostupný také z: <http://dx.doi.org/10.1016/j.lisr.2005.01.004>
- BRACKE, Paul J. Web usage mining at an academic health sciences library: an exploratory study. *Journal of the Medical Library Association*. 2004, roč. 92, č. 4, s. 421–428. Dostupný také z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521513/>
- 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
- 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
- CLEYLE, Su et al. Approaching Librarianship from the Data: Using Bibliomining for Evidencebased Librarianship. *Library hi tech*. 2006, roč. 24, č. 3, s. 369–375. ISSN 0737-8831.
- 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.
- CUNNINGHAM, Sally Jo and Eibe FRANK. Market basket analysis of library circulation data. In: *Proceedings of the 6th international conference on neural information processing*. Perth, 1999. s. 825–830. Dostupný z: <http://www.cs.waikato.ac.nz/~eibe/pubs/SJ-EF-Market-Basket.pdf>
- DENTON, William. On Dentographs, A New Method of Visualizing Library Collections. *Code4Lib Journal*. 2012, roč. 16, s. 171–179. ISSN 1940-5758. Dostupný také z: <http://journal.code4lib.org/articles/6300>
- DICKEY, Timothy J. Books as Expressions of Global Cultural Diversity. *Library Resources and Technical Services*. 2011, roč. 55, č. 3, s. 148–162.
- EMMONS, Mark and 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.

FINNELL, Joshua and 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 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.

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. 9. 2. 2016]. Dostupné z: https://assessment.trinity.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf

JOHNSON, Doug. A Data Mining Primer and Implications for School Library Media Specialists. *Knowledge Quest*. 2004, roč. 32, č. 5, s. 32–35.

KAO, S.-C., H.-C. CHANG and C.-H. LIN. Decision support for the academic library acquisition budget allocation via circulation database mining. In: *Information Processing and Management*. 2003, roč. 39, č. 1, s. 133–147. ISSN 0306-4573.

KITAJIMA, Mitsuro, Kensuke BABA and 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.

LACROIX, Eve-Marie and 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, č. 4. ISSN 1336-0779. Dostupné z: <http://www.cvtisr.sk/itlib/itlib074/laluhova.htm>

LEONARD, Michelle F., Stephanie C. HAAS and 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. ISSN 1092-1206. Dostupný také z: <http://dx.doi.org/10.5062/F4PC308T>

MELETIOU, Aristeidis and Anthi KATSIRIKOU. Methodology of analysis and interrelation of data about quality indexes of library services by using data and knowledge mining techniques. *Library Management*. 2009, vol. 30, no. 3. s. 138-147. ISSN 0143-5124. DOI 10.1108/01435120910937311

Ministerstvo kultury ČR. *Koncepce rozvoje knihoven ČR na léta 2011–2015 včetně internetizace knihoven* [PDF]. [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/CKbtmRCRemci>

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. The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries*. ALA, 2003, no. 4. ISSN 2163-5226.

NICHOLSON, Scott and Jeffrey M. STANTON. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In: H. NEMATI and C. BARKO eds. *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey (PA): Idea Group Publishing, 2003, s. 247–262.

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

OMASTOVÁ, Pavlína, Radka ŘÍMANOVÁ a Ctibor ŠKUTA. Elektronická stopa čtenáře v knihovnickém systému aneb co ví Aleph NTK i bez ankety. In: *Sborník z 19. konference, konané ve dnech 13.–15. září 2011 v Českých Budějovicích*. Ostrava: Sdružení knihoven ČR, září 2011, s. 101–109. ISBN 978-80-86249-62-9.

PAPAVLASOPOULOS, Sozon and Marios POULOS. Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*. 2012, vol. 33, iss. 4/5, s. 261–271. ISSN 0143-5124.

PU, Hsiao-Tieh and Chyan YANG. Enriching user-oriented class associations for library classification schemes. *The Electronic library*. 2003, vol. 21, no. 2. ISSN 0264-0473. DOI 10.1108/02640470310470507.

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>

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. ISSN 1335-7026, ISSN 1336-0965. Dostupné z: http://www.snk.sk/images/snk/casopis_kniznica/2007/november-december/21.pdf

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, Timothy J. DICKEY. An experimental authority file and prototype application author. *Library Resources & Technical Services*. 2011, vol. 55, iss. 4, s. 182–194. ISSN 0024-2527.

SILIPIGNI CONNAWAY, Lynn, Edward O'NEIL and Chandra PRABHA. Last Copies: What's at Risk?. *College and research libraries*. 2006, roč. 67, č. 4, s. 370–379. ISSN 0010-0870. Dostupný také z: <http://www.oclc.org/research/publications/archive/2006/connaway-crlo7.pdf>

SORIA, Krista M., Jan FRANSEN and 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

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

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.

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.

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 and 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. In: *International Journal of Electronic Business Management*. 2012, roč. 10, č. 2, s. 163–173. ISSN 1741-5063.

ZWEIBEL, Stephen and 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.