EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Adatbányászat a könyvtárakban

SZAKDOLGOZAT

Témavezető:
Dr. Kiszl Péter
egyetemi adjunktus

Készítette:
Mátyás Melinda
informatikus könyvtáros szak

Budapest, 2009

Tartalomjegyzék

1. BEVEZETŐ - MIÉRT ÉRDEKES AZ ADATBÁNYÁSZAT A KÖNYVTÁRAKNAK?

Az adatbányászat eljárását ma főként az üzleti, gazdasági, információtechnológiai életben alkalmazzák. Az eljárást használó üzleti elemzők a nagy mennyiségű adatot tároló adattárházaikban keresnek az adatbányászat segítségével olyan mintákat, amelyeket az ember gépi segítség nélkül nem tudna felfedezni.

A marketinghez használva elkülöníthetnek vásárlói csoportokat, vagy meghatározhatják azoknak a körét, akik korábban már válaszoltak a leveleikre vagy jó esélyt látnak rá, hogy válaszolnak, és csak nekik postáznak leveleket, ezzel is költséget takarítva meg.[1]

Ennek tükrében a könyvtárakban is egyre fontosabbá válik, hogy megfeleljenek a felhasználók differenciált igényeinek, jobban megismerjék a használóikat és a szervezetük munkafolyamatait a hatékonyabb és gazdaságosabb működés érdekében.

Az is fontos, hogy a könyvtárak reagáljanak az információs társadalom által hozott változásokra; például az információk mennyiségi növekedésének korában számadó információkkal szolgálhat és egyre több technikai lehetőség is kínálkozik rá, hogy nagy adathalmazokat elemezzenek, az intézmény irányítása céljából hasznosnak vélt adatokból kimutatásokat készítsenek, vizualizálják az adatokat. A könyvtárak esetében nagy adathalmazokat főként a könyvtári integrált rendszerekben tárolnak és kezelnek, ezen adatok új megközelítése és felhasználása szolgáltat alapot az adatbányászat alkalmazásához a könyvtárakban.

Az új felfogás az, hogy nemcsak a szolgáltatások és könyvtári működés támogatására használják fel ezeket az adatokat, hanem ezen információk egyfajta stratégiai erőforrássá, kiaknázható lehetőséggé váljanak a könyvtári szolgáltatások fejlesztésére, egyes szolgáltatások indoklására, tendenciák megfigyelésére.

Az adatoknak ilyen fajta kezelésére kínál lehetőséget az adatbányászat, amelyet a könyvtárakban is sikerrel használnak, tudatos alkalmazással eddig nem ismert tudásra tehetnek szert a szolgáltatások használatáról és a követendő fejlődési irány előrejelzéséről.

Sőt, egyes hazai könyvtárakban már látszik is a lehetőség az adatbányászat egy-egy változatának használatára. Például ilyennek tekinthetjük, hogy a Fővárosi Szabó Ervin Könyvtárban elkészítettek egy kimutatást a leggyakrabban kölcsönzött könyveiről, és

[1] ADRIAANS-DOLF ZANTINGE, Pieter: Adatbányászat. Panem. Budapest. 2002. pp. 19-20.

ezt a honlapjukon nyilvánossá is tették[2]. Ha ezt elemzik adatbányászati eszközök segítségével, szert tehetnek stratégiailag fontos, az adatok mélyén fekvő tudásra. Vagy például volt olyan könyvtár, ahol feltették azt a kérdést, hogy egy adott kiadó mely folyóirataira fizetnek elő, és milyen mértékben változott az áruk az elmúlt években. Vagy pedig, hogy milyen fajta e-könyveket használnak a felhasználók, és ezek a használatok milyen felhasználói csoportokhoz köthetők.[3] Ezen kérdések megválaszolásához a könyvtári integrált rendszer vagy az általában meglévő könyvtári gépi eszközök nem elegendőek, pedig a válaszok érdekesek és hasznosak lennének a könyvtári döntéstámogatáshoz vagy akár az állományfejlesztéshez.

Külföldön több könyvtár felismerte a precízebb, számokkal és bizonyítékokkal jobban alátámasztott döntéshozás eszközeinek szükségességét. Sam Clay, a Fairfax County Public Library[4] igazgatója szerint a könyvtáraknak érdemes üzleti modelleket alkalmazni, szerinte ezeket a modelleket használva nem kell spekulálni a döntéshozáskor vagy a megérzésekre hallgatni, hanem ki kell használni, hogy az adatbányászat által eredményül kapott tudás alapján döntsenek és cselekedjenek.[5] A könyvtára körülbelül húsz évre visszamenő tendenciákról szóló adatokat gyűjtött össze, amelyeket vállalkozó szellemű módon fel is használnak.[6] Az egyik adatbányászat-felhasználási példájukat nézve: van egy 2007 júniusa óta üzemelő blogjuk[7], amelyen hétről hétre feltüntetik az aktuális hét legtöbbször kölcsönzött könyveit, valamint az átlagosan legtöbbször kölcsönzött könyvek listáját. Továbbá az olvasók a blogról közvetlenül ismertetőt vagy megjegyzést írhatnak (*write a review* menüpont) valamint előjegyeztethetik a kívánt könyveket *(reserve the book)*.

Ez a fajta blog jó lehetőséget ad arra, hogy a könyvtár keretein (oldalain) belül összefogja a felhasználókat és a népszerű könyvek ismertetésével, megjegyzések hozzáfűzésével aktív szereplőjévé tegyék őket a könyvtári, olvasói életnek, a könyvek azonnali előjegyzésének lehetőségével pedig megkönnyítsék az előjegyzés majd a kölcsönzés igénybevételét.

---

[2] A szolgáltatás elérhető az alábbi címen: http://www.fszek.hu/?article_hid=23981 [2009-02-16].
[3] CULLEN, Kevin: Delving into data. [elektr. dok.] = Library Journal. 130. köt., 13. sz. 2005. EBSCO Azonosító: 03630277.
[4] A könyvtár hivatalos weboldala: http://www.fairfaxcounty.gov/Library/ [2009-01-24].
[5] CULLEN, Kevin: Delving into data. [elektr. dok.]. „[We] have to apply business models to what we do,… Clay says: „Don't think, feel or intuit. Do it, because what you know."
[6] CULLEN, Kevin: Delving into data. [elektr. dok.].
[7] A blog elérhetősége: http://allfairfaxreads.blogspot.com/ A hivatalos oldalukról a „Good reading" nevű menüpontra kattintva érhető el. [2009-01-24].

2. ADATBÁNYÁSZAT ÉS A KÖNYVTÁRAK

Láthatjuk, hogy vannak olyan könyvtárak, amelyekben már észrevehetően alkalmazzák az adatbányászat eljárását, de általánosságban nem rég óta kezd bekerülni a könyvtári gyakorlatba. Az adatbányászat és a könyvtárak volt a fő kutatási témája Scott Nicholsonnak, valószínűleg ő foglalkozott legbehatóbban a területtel, számos tanulmányt írva róla és kutatócsoportot szervezve. Nicholson jelenleg az amerikai szirakúzai egyetemen dolgozik adjunktusként, valamint a Master of Science in Library and Information Science program irányítója szintén a szirakúzai egyetemen.[8] Jelenleg a fő kutatási területe a játék a könyvtárakban *(Gaming in libraries)*, amelynek vizsgálatához, a korábbi kutatási területéhez kapcsolódva felhasznál könyvtári adatbányász-eszközöket és elképzeléseket.[9] Ő és szerzőtársa, Jeffrey Santon kívánta bevezetni a *bibliomining* szakkifejezést az adatbányászat és könyvtárak elnevezésű témára, amely alatt a s Nicholson az adatbányászat, az adattárházak és a bibliometria fogalmát érti[10]. A szerző tanulmányaiban így is hivatkozik rá, dolgozatunkban Mikulás Gábort követve könyvtári adatbányászatot[11] írunk vagy az adatbányászat a könyvtárakban kifejezést használjuk.

A keresők tanúsága szerint a *bibliomining* kifejezés főként Nicholson írásaival kapcsolatban jelenik meg, de egyes példák szerint mások is használják ezt az új terminust az adatbányászat és a könyvtárak kontextusában. Például ezt a szót használja egy indiai szerzőpáros egy, a témához kapcsolódó tanulmány címében: „*Bibliomining Processes for Integrated Library System*"[12], valamint érdekes, hogy a szakkifejezést spanyol nyelvterületen átvették „*bibliominería*" néven, ahogy a madridi szerző tanulmányának címéből[13] látszik.

Jelen dolgozat nagymértékben támaszkodik Scott Nicholson írásaira[14], összefoglalva az eredményeit, és megnézve, hogy hogyan lehet azokat hazai területen felhasználni. Kitérünk a hazai felhasználás meglevő példáira is, de a tapasztalatok

---

[8] [NICHOLSON, Scott]: Career. http://scottnicholson.com/career/index.html , [2009-01-24].
[9] [A bibliomining honlap főoldala], http://www.bibliomining.com/ [2009-01-24].
[10] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship. [elektr. dok.] http://bibliomining.com/nicholson/approach.htm [2009-09-07].
[11] MIKULÁS Gábor: Adatbányászat a könyvtárakban. Referátum. [elektr. dok.] http://www.gmconsulting.hu/inf/cikkek/207/index.php [2009-01-24].
[12] [A British Library Direct tanulmány-adatai]: http://direct.bl.uk/bld/PlaceOrder.do?UIN=153304283&ETOC=RN&from=searchengine [2009-01-24].
[13] ROMERO, Jorje Candás: Minería de datos en bibliotecas: bibliominería [elektr. dok.]. http://www2.ub.edu/bid/consulta_articulos.php?fichero=17canda2.htm [2009-01-24].
[14] A szerző elismerésre méltó módon a weboldalán nyilvánossá tette a témáról megjelent tanulmányait: http://bibliomining.com/nicholson/ [2009-01-24].

alapján hazánkban az adatbányászat könyvtárakkal kapcsolatban még nem igazán ismert és alkalmazott terület. Így inkább átfogó képet adunk az adatbányászat fogalmáról és a könyvtárakban való alkalmazás lehetőségeiről, amelyeknek a köre igen széles, mivel a könyvtárak nagy mennyiségű adatot kezelnek, és az adatbányászat lényegében minden olyan területen hasznosítható, ahol nagy mennyiségű, használhatónak ítélt adatok állnak rendelkezésre, és szükség van a döntéshozói vagy a szervezeti munka támogatására, megfigyelésére, fejlesztésére. Scott Nicholson művein kívül felhasználunk a témánkhoz kapcsolódó, más szerzőktől származó írásokat, a téma kurrens volta miatt főként folyóiratokban megjelent tanulmányok a források. Leginkább azok a források bizonyultak hasznosnak, amelyekben konkrét könyvtári adatbányászati példákat is leírnak vagy említenek.

Sok közülük elérhető teljes szöveges formátumban internetes tartalomszolgáltatók adatbázisaiban (például EBSCOhost[15]), az ezekre való hivatkozáskor feltüntetjük az EBSCOhost által adott azonosítószámot. Szintén a téma kurrens volta és természete miatt sok internetes forrást is felhasználtunk, ezeknél megadjuk link formátumban a pontos elérhetőséget és az oldal legutóbbi letöltésének dátumát. Az internetes tartalmak gyors változása miatt a legfontosabb elektronikus formátumú írások másolatát a dolgozathoz mellékelt CD tartalmazza.

A könyvtári adatbányászat alkalmazásának főbb és behatárolhatóbb területei a példák alapján a döntéstámogatás, a menedzsment, a tájékoztatás és az állományfejlesztés. Ezekről a későbbi fejezetekben lesz szó.

Először ismerkedjünk meg az adatbányászat fogalmával, a hozzá szükséges feltételekkel, eszközökkel. Majd nézzük meg, hogy egy könyvtárban hogyan valósulhat meg az eljárás felhasználása, melyek a speciálisan könyvtárra jellemző feltételek, források az adatbányászathoz. Az alkalmazás előző bekezdésben felsorolt főbb területei előtt pedig áttekintjük, hogy mely általános, átfogóbb környezetben, milyen szemléletmód mellett alkalmazható az adatbányászat a könyvtárakban, erre lesz példa az adatbányászat a bizonyítékokon alapuló könyvtári munkában.

2.1. Alapfogalmak

2.1.1. A tudásfeltárás és az adatbányászat

Az adatbányászat (angol nyelvű terminológiával *data mining*) egy nagyobb folyamatba, a tudásfeltárás folyamatába (*KDD – Knowledge Discovery in Databases*) illeszkedik. A

[15] Az EBSCOhost weboldala: http://www.ebscohost.com/ [2009. 01. 24.]

tudásfeltárás az adatbázisokban azt jelenti, hogy előre nem ismert, közvetett, hasznosítható információt, tudást szeretnénk kinyerni az adatokból nem triviális, nem nyilvánvaló módon.[16] A tudás itt az adatelemek közötti kapcsolatokat, mintákat jelenti. Az első tudásfeltárásról szóló, 1995-ös montreali konferencián úgy fogalmazták meg, hogy a tudásfeltárás az a teljes folyamat, „amelynek során az adatokból kinyerjük az információt".[17]

Tisztázandó, hogy a tudásfeltárás és az adatbányászat nem egymás szinonimái, hanem az adatbányászat a tudásfeltárásnak egyik kulcsfontosságú eleme, a tudásfeltárás folyamatának egyik állapota, amelynek során történik a minták felismerése, rejtett összefüggések felfedezése az adatokban.

A tudásfeltárás összességében nem új technika, hanem több tudományágat átfogó terület, tartalmazhatja a gépi tanulást, a statisztikát, az adatbázis-technológiát, a szakértői rendszereket és az adatok láthatóvá tételének egyes területeit is. A tudásfeltárás életre hívása ilyen formában kapcsolódik annak felismeréséhez, hogy az információt termelési tényezőnek tekintsük. [18]

2.1.2. Az adattárházak

A tudásfeltáráshoz szorosan kapcsolódnak az adattárházak *(datawarehouses)*, amelyek egy központi adattár szerepét töltik be. Az adattárházak használata is üzleti mintán alapszik, a Vállalati Erőforrás-Gazdálkodás Tervező Rendszerrel (*ERP – Enterprise Resource Planning System*) rendelkező cégek az információikat automatikusan adattárházakba mozgatják.[19]

Az adattárház alkalmazásánál az elemezni kívánt adatokat először ki kell nyerni (exportálni) a forrásul használt adatbázisból vagy egyéb információ-tároló rendszerből, és ezeket a kinyert adatokat át kell tölteni az adattárházba. Az a különbség a forrás-rendszerben és az adattárházakban található adatok szervezése között, hogy az adattárházakban az információk az adatbányászat alkalmazásához megfelelő formátumban találhatók, valamilyen meghatározott témához kapcsolódva szerveződnek.

---

[16] ADRIAANS-DOLF ZANTINGE, Pieter: Adatbányászat, p. 144.
[17] *Uo.*, 17.
[18] *Uo.*, 17. p.
[19] CULLEN, Kevin: Delving into data. [elektr. dok.].

Az adattárházak fő jellemzője, hogy általában különösen sok adat, nagy adathalmazok találhatók bennük, mert az adattárház adatait és az adatok történeti változását állandóan tárolni szeretnék bennük.[20]

Az adattárházak így lehetővé teszik, hogy a tudásfeltárás folyamatához szükséges adatokat ezekben tárolják, ne az adott szervezet, így a könyvtár mindennapi használatra szánt rendszerét terheljék, ne csak annak erőforrásait használják. Továbbá az adattárház elemzésre van optimalizálva, így megkönnyíti a mintakeresést az adatbányászat folyamata során.[21]

A könyvtárak esetében az adattárházakban lévő adatok fő forrásául a könyvtári integrált rendszerek szolgálhatnak, ahogy ezt Mikulás Gábor[22] és Kevin Cullen[23], a Colorado Állambeli Egyetem könyvtárának digitális projektekkel foglalkozó könyvtárosa, is kiemeli. A könyvtári integrált rendszerek ilyetén használatáról a jelentőségük miatt lentebb, külön alfejezetben szólunk. Az adattárházba mozgatott adatok a könyvtárak esetében származhatnak még a web szerver naplózásából, a könyvtárközi kölcsönzések adataiból, a könyvtárhoz tartozó terület demográfiai adataiból[24] vagy a könyvtár pénzügyi adatbázisából is[25]. Ezeken kívül bármely más forrásból, könyvtári használathoz kapcsolódó adatbázisból, amellyel az adott könyvtár rendelkezik, és úgy gondolja, hogy a bennük tárolt információk elemzése hasznos lehet. A lényeg a források használhatóságánál, hogy az adatok szervezett formában legyenek tárolva, egyértelműen azonosíthatók legyenek az adatelemeik, és az adattárházba való integráláshoz exportálható formátumúak legyenek, technikailag meg tudják oldani a beépíteni kívánt adatok konverzióját.

Felmerülhet a kérdés, hogy mindenképp szükséges-e adattárház használata az adatbányászathoz. A válasz az, hogy nem mindenképpen, de hasznos, ha rendelkeznek vele, mert több lehetőséget biztosít az adatbányászat sokrétű, összetettebb használatához, valamint jobban eloszthatják a használt erőforrásokat. A könyvtári integrált rendszerből általában közvetlenül is készíthetnek statisztikát például a kölcsönzésekről, vagy azokról a használati adatokról, amelyekről a könyvtárban használt integrált rendszer képes összegzést adni, ezekhez a jelentésekhez nem

[20] ADRIAANS-DOLF ZANTINGE, Pieter: Adatbányászat, p. 9.
[21] CULLEN, Kevin: Delving into data. [elektr. dok.].
[22] MIKULÁS Gábor: Adatbányászat a könyvtárakban. Referátum. [elektr. dok.].
[23] CULLEN, Kevin: Delving into data. [elektr. dok.].
[24] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship [elektr.dok.].
[25] TÓTH Erzsébet: Adatbányászatra irányuló törekvések könyvtári területen. [elektr. dok.] = Könyvtári Figyelő 48. köt. 3. sz. 2002. http://www.ki.oszk.hu/kf/kfarchiv/2002/3/toth.html [2009-09-07].

szükséges az adattárház. De azt is figyelembe kell venni, hogy a beépített jelentések nem mindig válaszolják meg egy dinamikus szervezetben felmerülő kérdéseket.[26] Például, ha a történeti adatokat is szeretnének vizsgálni, tendenciákat kívánnak kutatni, több mező vagy adatcsoport összekapcsolásával szeretnének vizsgálódni, ezekre a feladatokra jó szolgálatot tehet egy adattárház. Az integrált rendszerekbe épített jelentéskészítő alkalmazásokat és az adattárház nyújtotta lehetőségeket legcélszerűbb egymás kiegészítéseiként kezelni, és nem egymást kizárva.

Mivel az adattárházak optimálisan működnek nagyobb mennyiségű adatcsoporttal is, a kisebb könyvtárak költségvetési szempontból megfontolhatnák adatmegosztó konzorciumok létrehozását.[27] Így egy nagy adatbázisban külön lennének tárolva az egyes könyvtárak adatai, mind hozzáférhetne a sajátjához, sőt, ha további együttműködésre is nyitottak a konzorcium tagjai, akár egymás elemzéseit, adatait, szolgáltatásait is összehasonlíthatnák, például egy-egy új szolgáltatás bevezetése előtt, vagy annak felméréséhez, hogy milyen szolgáltatásokat nyújt egy-egy könyvtár, és a használati adatok szerint mennyire népszerű, eredményes az adott szolgáltatás.

A könyvtárak esetében az adattárházakban található adatok fő forrása tehát a szervezet mindennapos használatára szánt vagy operatív rendszere, az integrált könyvtári rendszer, amelyet az előbb hivatkozott Kevin Cullen az ERP-rendszerekkel állít párhuzamba, és az adattárházak alkalmazásához kiindulópontnak tekinti azokat.[28]

Teljes mértékben egyet is értünk vele, mivel a könyvtárakban általában az integrált rendszerekben tárolják a legtöbb használati adatot, amelyek elemezhetők, hasznosíthatók lehetnek. Alapesetben ezek a rendszerek adatokat tárolnak a könyvtár gyűjteményéről, a kölcsönzésekről, a felhasználókról, a kiadásokról, az ezeken kívül megtalálható információk integrált rendszerenként különbözőek, illetve az sem elhanyagolható, hogy az előbb felsorolt típusok tárolási formája is más és más a különböző rendszerekben.

Az integrált rendszerekbe épített jelentéskészítő, statisztikai eszközök adatbányászati alkalmazása érdekes terület, erre az alábbiakban bővebben kitérünk.

---

[26] CULLEN, Kevin: Delving into data. [elektr. dok.].
[27] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. [elektr. dok.] http://bibliomining.com/nicholson/odmcom.html [2009-09-07].
[28] CULLEN, Kevin: Delving into data. [elektr. dok.].

2.2. Az adatbányászat és az integrált könyvtári rendszerek

A könyvtári integrált rendszerekkel kapcsolatban Cullen kiemeli, hogy a könyvtárak sajnos gyakran azt tapasztalják, hogy az integrált rendszer adatai egy „fekete dobozba" vannak zárva, és csak a rendszer gyártója által biztosított jelentések segítségével nyerhetők ki.[29] Így ennek elkerülése végett a könyvtáraknak fontos figyelembe venniük, hogy egy relációs adatbázis alapú integrált rendszerrel bírva nem szükségszerűen könnyű adattárházat létrehozni, mivel ehhez a rendszernek nyitottnak kell lennie, külső rendszerek számára elérhetőnek vagy a bennük található információknak exportálhatóknak és jól felépített szerkezetűeknek kell lenniük.[30]

A szerző arra is rámutat, hogy bár a gyártók nem mindig teszik egyszerűvé az információk kinyerését általuk forgalmazott rendszerekből, vannak pozitív példák, amelyek azt mutatják, hogy egyesek felismerték a jelentőségét a rendszereikben tárolt adatok mélyebb szintű elemzési eszközeinek vagyis az adatbányászatnak, és szélesebb körű eszköztárat biztosítanak magában az integrált rendszerhez kapcsolt alkalmazások segítségével ennek kihasználására.

Például a Horizont (az ELTE–BTK legtöbb könyvtárában ennek az integrált rendszernek egy korábbi változatát használják) gyártó Dynix, ma SirsiDynix[31] üzleti intelligencia-programok gyártóival társulva (a MicroStrategy-vel és a SwiftKnowledge-dzsal, amely olyan adat-megjelenítő technológiákkal támogatja, mint az OLAP – *Online Analytical Processing* – Online Analitikus Feldolgozás és a *data cube* – dinamikus többdimenziós adatkocka[32]) olyan egyedülálló alkalmazásokat épített az integrált rendszerbe, mint a Web Reporter, amely támogatja az azonnali automatikus jelentéskészítést, a végfelhasználónak web alapú kezelőfelületet biztosítva[33] vagy a Director's Station, amely szintén web alapon működő elemző eszköz, a menedzserek egyéni igényei szerint kiválasztott használati adatokból képes jelentéseket készíteni[34]. Továbbá adatokat tud költöztetni az integrált rendszerből, valamint külső információkat

[29] Uo.
[30] Uo.
[31] A SirsiDynix hivatalos oldala az általa forgalmazott integrált rendszerekkel: http://www.sirsidynix.com/Solutions/Products/integratedsystems.php [2009. 01. 25.]
[32] CULLEN, Kevin: Delving into data. [elektr. dok.].
[33] Uo.
[34] A gyártó ismertetőjéből: http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/Symphony_Features_Benefits.pdf [2009-01-25].

is képes fogadni, amelyeket, ha az adattárház adatbázisa fel tud használni, rögtön használhatnak a lekérdezéshez.[35]

A SirsiDynix másik figyelemre méltó elemzési eszköze a Normative Data Project for Libraries (NDP)[36], amely az előfizetők számára lehetővé teszi, hogy a döntéshozás támogatásához lássák más könyvtárak használati adatait (észak-amerikai könyvtárak adatait szolgáltatják), összehasonlítsák a sajátjukéival, valamint a könyvtári használati adatokhoz demográfiai és földrajzi adatokat illesszenek.[37]

Erre fog hasonlítani Scott Nicholson ötlete az adatbányászat alkalmazása a bizonyítékokon alapuló könyvtárosság témakör kapcsán, aki szerint hasznos lenne, hogy a könyvtárak megosszák egymással az adattárházaik információit[38]. Ebben az esetben az adattárházak adatait közvetlenül az adattárházat használó könyvtárak osztanák meg, megoldva a különböző adatformátumok egységesítését egy megosztott alkalmazásba konvertáláshoz, és nem egy szolgáltatónak fizetnének elő az információért, ahogyan a SirsiDynix-nél bemutatott szolgáltatásban tehetik.

Bár ezen alkalmazások egy része hazánkban nem hasznosítható (a helyhez kötöttség miatt), érdekes látni, hogy olyan gyártók, akik felismerték az adatbányászatban rejlő lehetőségeket, milyen széles körű eszköztárat tudnak kínálni az alkalmazáshoz. Ezt kifejezik azzal is, hogy gyártó mottója is a meglévő adatok dinamikus felhasználására utal: „*Bringing knowledge to life.*”[39] (A tudás életre hívása.). Ezzel egybevág, hogy Cullen szerint az integrált rendszereket gyártók vannak a legjobb helyzetben ahhoz, hogy integrálják szolgáltatásukba az adatbányászó, analitikus eszközöket.[40]

Egy részről valóban praktikus az integrált rendszerekben tudni ilyen jellegű szolgáltatásokat, de sok múlik a gyártók nyitottságán, a könyvtár igényeihez való alkalmazkodóképességén, hogy belevágnak-e ilyen jellegű szolgáltatások fejlesztésébe és karbantartásába, biztosításába a könyvtáraknak.

Más részről, ha nincsenek ilyen nagy hatékonyságú elemző eszközök beépítve az integrált rendszerbe, a könyvtáraknak lehetőségük van használni a jelenlegi

---

[35] CULLEN, Kevin: Delving into data. [elektr. dok.].
[36] Áttekintés az alkalmazásról a gyártó honlapjáról (képekkel): http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/NormativeDataProject.pdf [2009-01-25].
[37] Tömör leírás a szolgáltatásról a SirsiDynix oldalán: http://www.sirsidynix.com/Solutions/Products/analytical.php [2009-01-25].
[38] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship [elektr.dok.].
[39] A gyártó logójáról: http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/DirectorsStation.pdf [2009-01-25].
[40] CULLEN, Kevin: Delving into data. [elektr. dok.].

rendszerükben lévő statisztikai eszközöket, vagy ha vállalkozó szelleműek és megvan rá a lehetőségük, a mélyebb szintű, részletesebb elemezések, adatbányászó eljárások alkalmazásához adattárházat is hozhatnak létre. Ekkor nagyobb részben vagy teljes egészében az adattárház fogja biztosítani az adatbányászó eljárások használatának felületét, az integrált rendszeren belüli jelentéskészítő eszközök használata valószínűleg visszaszorul, míg adattárház nélkül csak azokra támaszkodhatnak.

Az adattárházak könyvtári használatára példa a Pennsylvaniai Egyetem Data Farm vagy Penn Library Data Farm[41] elnevezésű adattárháza, amelyben még feldolgozatlan adatokat tárolnak a szervezetről, annak tevékenységeiről, tranzakcióiról és felhasználóiról, annak érdekében, hogy a váratlanul felmerülő kérdésekre az adatok alapján válaszra vagy megerősítésre találjanak. Jelenlegi céljuk az adattárház használatával a felhasználók jobb szolgálata. A jövőbeli terveik pedig, hogy jobban megfigyeljék a szervezetük munkafolyamatait. [42]

Nézzünk meg egy szemléletes használati példát, amelyet Cullen ír le[43] a Penn Library Data Farm hasznáról: az egyetem diákjai panaszkodtak a kevés használható ruhatári szekrény miatt, a felsőbb vezetés pedig hezitált, hogy bármit is tegyen az ügyben. A Data Farm adatbányászó csapata hozzáfogott kivizsgálni az esetet, elkülönítette a vizsgálathoz szükséges adatokat, és öt évre visszamenő elemzést készített a szekrények használati adataiból. Azt találták, hogy valóban növekszik a szekrények használatának mértéke, de csak a szekrények harmadát használják, és a szám folyamatosan csökken. Ezután rájöttek, hogy nem használt szekrények kulcsai egyszerűen eltűntek. Ha a kutatás időtartamát nézzük: a kiinduláshoz szükséges elemzés elkészítése nem tartott tovább harminc percnél.

[41] Data Farm. University of Pensylvania Library. A Data Farm tematikus oldaláról: http://datafarm.library.upenn.edu/ [2009-01-26].
[42] CULLEN, Kevin: Delving into data. [elektr. dok.].
[43] Uo.

3. AZ ADATBÁNYÁSZAT ALKALMAZÁSI TERÜLETEI A KÖNYVTÁRAKBAN

3.1. Az adatbányászat a bizonyítékokon alapuló könyvtári munkában

Az adatbányászat eljárása hasznos és gyümölcsöző lehet a bizonyítékokon alapuló könyvtári munka segítéséhez. A bizonyítékokon alapuló könyvtári munka (EBL – *Evidence Based Librarianship*) az Evidence Based Health Care-ből, vagyis a bizonyítékokon alapuló orvoslásból alakult ki és vált önálló módszerré a könyvtári munkában az 1990-es évek végén.[44] Lényege, hogy a könyvtári munka gyakorlatát közelítsék a bizonyítékok alapján megtervezett munkához, és kiküszöböljék a keresések gyakori felületességét, valamint jobban megszűrjék a nagy mennyiségű információt.[45] Tágabb értelemben lehetővé teszi, hogy az olvasóközönséget jobban a valós igények és adatok alapján segítsék. Ezért a módszer nagymértékben épít a felmérések, statisztikák információira, azokban kutatja fel az elemző a hatékony működéshez szükséges információkat.[46]

Itt kapcsolódik az adatbányászat eljárásához, mivel a könyvtári adatbányászat éppen ezeknek a bizonyítékul szolgáló adatoknak az elemzésére, mintázatainak felismerésére koncentrál, ezért a hagyományos bizonyítékokra épülő könyvtári munkához praktikus, egyedi bizonyíték-központú és a könyvtár sajátosságait figyelembe vevő eszközöket kínál.

A bizonyítékokon alapuló könyvtári munka hagyományos módszere valóban az, amit fent említettünk, hogy elemzi a felméréseket és a statisztikákat, akár úgy, hogy egy probléma felmerülésénél a könyvtáros először azonosítja a problémát, majd rákeres a meglévő kutatási eredményekre.[47] Azonban bizonytalan, hogy ezek a kutatási eredmények mennyire relevánsak az adott könyvtár számára (hacsak nem a saját eredményeikről van szó), mert ezek az eredmények függnek az adott könyvtár sajátosságaitól, a kutatásban alapul vett mérések pontosságától és a kutató gondosságától is.[48]

---

[44] Evidence Based Librarianship – Bizonyítékokon (precedenseken) alapuló könyvtári munka. [elektr. dok.] = KIT Hírlevél 5. sz. 2003. http://www.kithirlevel.hu/index.php?oldal=cikk&c=746 [2009-02-17].
[45] Uo.
[46] Uo.
[47] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship [elektr.dok.].
[48] Uo.

Az adatbányászat felhasználásának az adott könyvtárban ezzel szemben az a nagy előnye, hogy a könyvtár saját adatai kerülnek elemzésre, biztosak lehetnek benne, hogy a bizonyítékok valóban a könyvtárnak megfelelő adatokból származnak.

Az alkalmazás menete alapjában úgy zajlik, mint a többi adatbányászati eljárásnál. Először összegyűjtik a meglévő és hasznosítható adatokat az adattárházba, főleg az integrált könyvtári rendszerből, ehhez kapcsolhatnak demográfiai adatokat[49] és a web-szerverek naplózásait. Ezután következik az adatok tisztítása, az elemzéshez szükségtelen adatok törlése és az összetartozó adatcsoportok összekapcsolása egy egyedi azonosító segítségével (erre szolgál az SQL-adatbázisokban pl. az elsődleges kulcs). Ekkor (az adatok összekapcsolása után) kerülhet sor a személyes adatok törlésére is, a felhasználók személyiségi jogainak védelmében, a törölt személyes adatok helyére kerülhetnek például egyedi azonosítószámok. Az összegyűjtés előnye, hogy különböző formátumú adatok egy helyre kerülnek, a keresés szempontjából azonos formátumúak lesznek, megkönnyítve az elemzésüket.[50]

Az ily módon létrejött adattárház felhasználásának menete úgy indul, hogy a könyvtárosok felhasználókról és/vagy a szolgáltatásokról tesznek fel kérdéseket. Azokra az adatokra alapozva a kérdéseket, amelyeket összegyűjtöttek az adattárházban, majd meghatározzák, hogy a kérdés vizsgálatához milyen további adatok hiányoznak vagy szükségesek még, a kutató ezeket is összegyűjti, és integrálja az adattárházba a kérdés vizsgálatához.

A kutatás során az elemzéshez bekapcsolhatnak statisztikai, adatbányászati, bibliometriai eszközöket is, azért, hogy felfedezzék az adatokban található mintázatokat. Ezek a mintázatok nyújthatnak bizonyítékokat a döntéshozáshoz, és egyben gyakran új kutatási területeket is kínálnak.[51]

A könyvtári integrált rendszer adatait alapul véve a statisztikai elemzéshez egy példa az állományvédelem és állománygyarapítás területéhez: annak vizsgálata, hogy a felhasználók mely dokumentumokat kölcsönzik a leggyakrabban, ezen dokumentumok csoportjának az állapotára például fokozottabb figyelmet lehet fordítani, ha szükséges, könyv esetében köttetésre lehet küldeni. Ezzel mintegy előre jelezhetik a később felmerülő szükségleteket, és még a felmerülésük előtt vagy a felmerülésükkor tudnak rá reagálni a könyvtárosok vagy a könyvtárvezetők.

---

[49] Uo.
[50] Uo.
[51] Uo.

Vagy másik példaként a gyakran kölcsönzött könyvek kimutatását összekapcsolhatják az előjegyzett könyvek adataival, és megnézhetik, hogy melyikre van a legnagyobb kereslet, ezután megfontolhatják az új példányok vásárlását az adott könyv(ek)ből. Más hasonló alkalmazási példa, ha elemzik a könyvtárközi kölcsönzéssel kért könyvek adatait, és ha úgy látják, hogy bizonyos könyveket gyakran kérnek, a szerzeményezési döntéshozók megfontolhatják az adott könyv beszerzését.

Az adatbányászati eszközzel történő kutatásra példa a tájékoztatás területéhez: ha a web-szerver vagy az online tájékoztatáshoz használt program vagy webes alkalmazás naplózza az online felületen történő beszélgetéseket, kérdésfeltevéseket és válaszokat a tájékoztató könyvtáros és a felhasználó között, bizonyos időközönként elemezhetik ezeket az adatokat a tájékoztató könyvtáros bevonásával (azért előnyös a bevonása, mert jobban tudhatja, hogy melyek a tipikus kérdések, segíthet a szelekcióban a sok kérdés-válasz között). Az elemzés után pedig készülhet egy lista a gyakran ismételt kérdésekről, és ezt GYIK-ként (Gyakran Ismételt Kérdésekként) feltölthetik a könyvtár honlapjára vagy a más típusú információkból születhet egy új menüpont, amely a felhasználók igényeinek megfelelően praktikusan gyűjti össze az egybe tartozó információkat, vagy olyan információkat, amelyekre a felhasználók igénye általában együtt szokott felmerülni. Így a könyvtári kommunikációban látható módon biztosíthatják a felhasználók kérdéseinek, felmerülő igényeinek figyelembevételét. Az eszközök a példák alapján megvannak hozzá, utóbbihoz akár adattárház igénybevétele nélkül is. Mivel ma sok könyvtár használja online tájékoztatási felületként a különböző online-böngészőn át használható vagy kliens-alapú ún. csevegő programokat, mint az Egyetemi Könyvtár[52] a meboo[53] elnevezésűt vagy az Országos Széchenyi Könyvtár[54] a skype[55] nevűt, ezek naplózási funkciójának bekapcsolásával meglenne egy forrás az adatok gyűjtéséhez és elemzéséhez.

Ezen könyvtári alkalmazások példáival át is tértünk az adatbányászat könyvtárakban való alkalmazásának területeire, amelyeket a következőkben kategóriákra, könyvtári munkaterületekre bontva vizsgálunk.

---

[52] A meboo az Egyetemi Könyvtár blogján („klogján") a jobb sávban: http://egyetemi.klog.hu/ [2009-02-18].
[53] A meboo szolgáltatás főoldala: http://www.meebo.com/ [2009-02-18].
[54] A skype-elérhetőség jelzése az OSZK főoldalán a bal sávban: http://www.oszk.hu/ [2009-02-18].
[55] A skype kliens-program magyar nyelvű változatának főoldala: http://skype.hu/ [2009-02-18].

3.2. A könyvtári adatbányászat alkalmazása a menedzsment és a döntéstámogatás területén

A könyvtári menedzsment és döntéstámogatás összekapcsolódó fogalmak, ténylegesen a menedzsment takar könyvtári munkafolyamatot, de a döntéstámogatást is kiemeltem, mivel napjainkban egyre nagyobb teret nyernek a döntéstámogató szoftverek, döntéstámogatást segítő eljárások a könyvtárakban is. Ilyen eljárás az adatbányászat is, a döntéstámogatáshoz általában eszközként használják, lényegében egy láncszem a döntéstámogató eljárások alkalmazásakor. A következőkben adatbányászati szempontból nézzük meg, hogy a könyvtári adatbányászat hogyan segítheti a döntéstámogatást, milyen szerepet tölthet be a döntéstámogatásban.

A könyvtári adatbányászatot a könyvtári menedzsment területén (a felhasználás néhány területét kiemelve) alkalmazhatják a költségvetési, a személyi és a munkaidő-beosztási döntésekhez.

Költségvetési kérdésekben támpontokat, elemzendő számszerű adatokat és tendenciákat adhat ahhoz, hogy a vezetők ésszerűbben döntsenek az intézmény költségvetéséről, több részlegből álló könyvtár esetén elemezzék, hogy melyik részlegnek mennyi tőkét célszerű juttatni.

A személyi részt nézve vizsgálhatják az alkalmazottak elvégzett vagy folyamatban levő munkafolyamatait, beosztását, például, hogy kit mikor célszerű beosztani (például a tájékoztatópultnál feltett kérdések napi megoszlása szerint), valamint az elvégzett munka összesítése, elemzése alapján rálátást kaphatnak a dolgozó teljesítményére.

Döntéstámogatás területén pedig abból a célból használhatják fel, hogy a menedzserek, a könyvtárvezetők vagy a döntéshozók igazolják egy-egy döntés szükségességét, döntéshozatal előtt a könyvtári adatbányászattal összegyűjtött információk alapján elemezzék a lehetőségeket, és kiválasszák az intézményüknek leginkább megfelelőt. Hasznos lehet például ilyen jellegű elemzések elkészítése egy új szolgáltatás bevezetése vagy meglévő szolgáltatás változtatása, átalakítása előtt.

3.2.1. A könyvtári adatbányászat költségvetési alkalmazásai

A következőkben tekintsük át részletesebben, könyvári példákkal ezeket az alkalmazásokat. Először nézzük a költségvetési alkalmazását. A DMBA *(data mining based decision support model)* elnevezésű, adatbányászatot felhasználó modellt használva a taiwani Kun Shan műszaki egyetem könyvtára (Library of Kun Shan University of Technology – LKSUT) egy komplex rendszert épített ki a könyvtár

költségvetésének célszerűbb, pontosabb adatokra épülő elosztásához.[56] A DMBA segítségével elemzik az akadémiai könyvtár kölcsönzési adatait, azért hogy támogassák a költségvetési döntéshozást az egyes részlegek számára juttatandó összegről, a szerint számolva, hogy az egyes részlegek mennyire dolgoznak a saját szakterületükön, hány olyan könyvet forgatnak az adott részlegen, amelyek a részleg saját szakterületéhez tartoznak. Ezen számítási eredmények alapján (amelyeket különböző formulák és algoritmusok felhasználásával kapnak meg) osztják el a költségvetést a részlegek között.[57] A módszert felhasználó egyetemi könyvtár vezetői szerint a költségvetés elosztásához szükséges számolni, értékelni, minősíteni, mivel ezek nélkül nem lehet elosztani a rendelkezésre álló pénzt.[58]

Ennek vizsgálatára különböző technikákat használnak, ilyenek a döntéshozatali technikákhoz tartozó[59] célprogramozás „*goal programming*" és a statisztikai módszerek. A célprogramozásban matematikai modellek fejlesztésével képesek közel optimális megoldást nyújtani összetett, egymással versengő elemeket, célokat tartalmazó problémákra, azzal, hogy kiválasztják a konkrét célokat, kritériumokat és felállítanak egy fontossági sorrendet közöttük, vagyis súlyozzák ezeket, előre meghatározva hozzájuk tartozó kritériumokat.[60] A statisztikai módszerek elsősorban arra koncentrálnak, hogy az előre meghatározott kritériumoknak (melyeket hierarchikus döntési fák tartalmaznak) mennyi a megegyező, megosztott aránya.[61]

A folyamat egészében igen összetett feladat mélyen feltárni az értékes információkat a tárolt kölcsönzési adatbázisban, mivel nem elég az egyes részlegeknél használt dokumentumok százalékos aránya, hanem a célravezetőbb eredményért az egyetem hozzáteszi a kölcsönzési adatok előre meghatározott mutatóját. Ez az erősség definiciója (*strenght*), amely megmutatja, hogy egy-egy dokumentum milyen szorosan tartozik az adott részleg szakterületéhez. Ezen kívül alkalmazzák az SQL-adatbáziskezelő nyelvet annak érdekében, hogy képesek legyenek igen nagy mennyiségű adatot tartalmazó adatbázisokban (pl. a kölcsönzési adatoké) keresni.[62]

---

[56] WU, C.-H.: Data mining applied to material acquisition budget allocation for libraries: design and development = Expert Systems with Applications, 25. sz. 2003. p. 403.
[57] Uo. p. 402.
[58] Uo. p. 407.
[59] KUN-PÁL Gábor: 57. fejezet: Döntéshozatal több kritérium felhasználásával, in *NCGIA CC* [elektr. dok.] http://gisfigyelo.geocentrum.hu/ncgia/ncgia_57.html [2009-07-31].
[60] WU, C.-H.: Data mining applied to material acquisition budget allocation for libraries, p. 402.
[61] Uo.
[62] Uo.

3.2.2. Személyi döntésekhez való alkalmazása

Scott Nicholson írása szerint a döntéshozás segítéséhez az adatbányászat használatával eredményes lehetne monitorozni a könyvtári alkalmazottak szokásait, viselkedését, bármennyire is kényelmetlenül hangozhat ez elsőre a könyvtárosok számára, a szerző szerint a kevesebb kiadás és a bizonyítékok iránti (felsőbb) igény és a tudatos és gondos teljesítmény nyomon követést igényel.[63]

További személyi döntésekhez való alkalmazás a kölcsönzési pult könyvtárosai munkájának elemzése, amely segítségével optimalizálni lehetne az ott dolgozók számát.[64] Például, ha megállapítanak egy tendenciát arról, hogy mikor vannak általában nagy sorok a pultoknál, arra az időszakra ideiglenesen több kölcsönző könyvtárost hasznosíthatnának az olvasók hosszabb ideig tartó várakoztatásának elkerülésére. Természetesen ez függ az adott könyvtár adottságaitól is, miszerint hány kölcsönzési pulttal rendelkezik, a Fővárosi Szabó Ervin Könyvtár Központi Könyvtárában a raktárból kért könyvek kiadási pultjával együtt három kölcsönzési pult van, és általában mindnél ül egy-egy kölcsönző könyvtáros, így ott ezt nem tudnák alkalmazni, hacsak nem bővítik a könyvkiadási pultok számát.

3.2.3. A könyvtári adatbányászat alkalmazása a döntéshozáshoz

Nézzük meg, mely könyvtári munkafolyamatokkal kapcsolatos döntéshozáshoz lehet hasznos az adatbányászat elemzéseinek használata.

A könyvtárközi kölcsönzést tekintve Nicholson szerint használhatnák a könyvtárközi kérések elemzését új előírások szükségletének alkalmazására.[65]

A szerzeményezést tekintve a költségvetés általában szűkös kerete miatt hasznos lehet az eladó, kiadó és az árak kiválasztásánál. Valamint oda lehetne figyelni arra, hogy mennyi idő telik el egy könyv rendelése és a polcra kerülése között, megvizsgálva, hogy mennyi idő kell a különböző típusú dokumentumok polcra kerüléséhez, ezeket dokumentum-típusonként csoportosítani, majd az átlagos időt hozzárendelni és azt számadónak tekinteni a további rendeléseknél, elkerülve az ok nélkül hosszan elhúzódó bevételezési munkát.[66]

---

[63] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[64] Uo.
[65] Uo.
[66] Uo.

3.2.4. Alkalmazások a döntéstámogatásban

A könyvtári döntéstámogatásban történő alkalmazását részben már érintettem a bizonyítékokon alapuló könyvtári munka (angol rövidítéssel: EBL) kapcsán. Az EBL eljárásban a meglevő vagy körvonalazott döntéshez való bizonyítékokon van a hangsúly, és ezek összegyűjtéséhez szolgál egy hasznos eszközzel az adatbányászat, a döntéstámogatásnál pedig inkább azon van a hangsúly, hogy a még nem feltétlenül kialakult döntéshez gyűjtsenek adatokat, kialakítsák a lehetséges opciókat. Ezt a könyvtári adatbányászat segítségével statisztikai, bibliometriai és adatbányászati eszközökkel érhetik el, azért hogy felfedezzék az adatokban található (sok esetben rejtett) mintázatokat.[67]

A döntéstámogatásban nem utolsó sorban előnye az adatbányászatnak (elsősorban az EBL kapcsán), hogy gyors döntések meghozásához is praktikus, mivel a közös felületre gyűjtött adatok lekérdezésével rövid időn belül kigyűjthetik a bizonyítékul szolgáló, a döntéshozás szempontjából releváns információkat. Az összegyűjtött adatok elemzéséhez a menedzserek használhatnak OLAP (*Online Analytical Processing* – Online Analitikus Feldolgozás) rendszereket, amelyek lehetővé teszik a gyűjtött adatok elemzését anélkül, hogy a menedzserek vagy könyvtári döntéshozók értenének az adatbázis lekérdező nyelvéhez. A DREW (The Digital Reference Electronic Warehouse – Digitális Tájékoztatási Elektronikus Tárház) elnevezésű adattárház-projekt például OLAP-eszközökre építi a szolgáltatásait.[68]

A DREW projekt célja volt (a projektet 2005-ben hívta életre Scott Nicholson és David Lankes, ma a honlapjai alapján az eredeti tervek szerinti formában nem látszik működni) egy multidiszciplináris tudásbázis építése a digitális tájékoztatási folyamat mélyebb megértése elősegítőjeként.[69] OLAP eszközöket az adattárházban archivált tranzakciók (kérdés-válaszok) adataiból szokásos vagy egyéni jelentés készítésére[70] használhatók a projektben.

Külföldi alkalmazási példaként kiemelhetjük Sam Clay, a Fairfax County Public Library igazgatója meglátását: szerinte szükség van a könyvtárakban a minőségi

---

[67] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship [elektr.dok.].
[68] Uo.
[69] NICHOLSON, Scott: Bibliomining Applications in Digital Reference: Using Data Warehousing and Data Mining to Improve Management and Decision-Making. [elektr. dok.]. http://www.webjunction.org/white-papers/-/articles/content/439065?_OCLC_ARTICLES_getContentFromWJ=true [2009-08-07].
[70] LANKES, R. David – NICHOLSON, Scott: Archiving Human Intermediation: The Digital Reference Electronic Warehouse (DREW) Project. [elektr. dok.] http://quartz.syr.edu/rdlankes/Presentations/2004/drewasist.pdf , 5 [2009-08-07].

döntéshozásra, így a döntéstámogató rendszerekre, akár üzleti modellek alkalmazásával is, a könyvtára körülbelül húsz évre visszamenőleg gyűjt tendencia-adatokat, amelyeket vállalkozó szelleműen használnak fel a döntéstámogatáshoz.[71]

A könyvtári döntéstámogatáshoz haszonnal alkalmaznak üzleti intelligencia (mint a Cognos[72]) vagy döntéstámogató programokat is a külföldi példák szerint. A waterloo-i egyetem könyvtárosa különösen hasznosnak találta döntéshozási információ-forrásként a Cognos program külső jelentéskészítő eszközeit, így például éves kiértékelést készíthettek minden épület után fizetett biztosítási díjról.[73]

A Fairfax County Public Library igazgatója előszeretettel használja a Director's Station nevű alkalmazást, amely adatbányászó és analitikai eszköz, amelyet a SirsiDynix (a Horizon nevű integrált könyvtári rendszer gyártója) és a SwiftKnowledge közösen fejlesztett ki. Az igazgató vezetői döntések meghozásához, támogatási kérelmekhez és marketingdöntésekhez használja a Director's Station alkalmazást.[74]

A fentebbi döntéstámogatási eszközök használata mindenképp megkívánja az adattárház meglétét, mivel az ebben gyűjtött adatokat elemzik, és teszik láthatóvá, értelmezhetővé a felhasználó számára a különböző analitikai eszközökkel. Az adattárház és az elemző eszközök beszerzése, fenntartása viszonylag nagy beruházást kíván a könyvtár részéről, de ha valóban szüksége van rá a könyvtárnak, és sikeresen alkalmazzák (képzett elemzőkkel, innovatív hozzáállással), akkor nagy eséllyel megtérülne a rá fordított összeg. Adattárház hiányában általában az integrált könyvtári rendszerek statisztikái is szolgálhatnak bizonyos információkkal a döntéstámogatáshoz (például a gyakran használt dokumentumok kimutatása), főleg, ha ismerik a meglévő eszközöket és kreatívan fel is használják azokat.

3. 3. A könyvtári adatbányászat alkalmazása a gyűjteményszervezésben

A gyűjteményszervezés számos területén alkalmazhatják sikerrel a könyvtárak az adatbányászatot, így az állománygyarapítás és -apasztás, az állománykarbantartás és nem utolsó sorban az állományalakítás, állományelemzés területén.

Azért alkalmazhatják sikerrel, mert az adatbányászat eljárásában felhasznált legfontosabb forrásban, az integrált könyvtári rendszerben tárolt adatok nagyon sok felhasználható információt tartalmaznak a meglévő könyvtári állományról (optimális

[71] CULLEN, Kevin: Delving into Data. [elektr. dok.].
[72] A program leírása a wikipédiában: http://en.wikipedia.org/wiki/Cognos [2009-08-08].
[73] CULLEN, Kevin: Delving into Data. [elektr. dok.].
[74] CULLEN, Kevin: Delving into Data. [elektr. dok.].

esetben a könyvtár egész állományáról), a dokumentumok bibliográfiai adatairól, a példányszámairól és a kölcsönzési adatokról. Ha ezeket az adatokat összegyűjtik egy adattárházban, és elemzik, sok értékes információhoz juthatnak, addig fel nem tárt mintázatokra, összefüggésekre jöhetnek rá az elemzők, amelyeket a könyvtár vezetőinek, gyűjteményszervezésért felelősöknek továbbítva felhasználhatnak a költséghatékonyabb, ésszerűbb és valós adatokon alapuló, átgondolt munkához.

3.3.1. Az állományelemzéshez való alkalmazás

A bibliográfiai adatokból nyerhető információk (cím, szerzőség, kiadás, fizikai leírás, elérés módja, azonosítószám, tárgyszó, esetlegesen plusz adatok a Dublin Core Metaadat formátum alkalmazása esetén a web alapú dokumentumoknál) főként az állományelemzés, a gyűjtemény megértéséhez hasznosak, így például alapszintű gyakorisági információk nyerhetők belőlük. Az összegyűjtött adatokat ezután össze lehet hasonlítani más könyvtárak adataival, kiértékelhetik a saját állományuk méretét, és az állomány egy-egy szakterületen való részesedését.[75] Ez alapján azt is megtudhatják pontos adatok alapján, hogy mely szakterület az erőssége a könyvtáruknak, melyikre érdemes nagyobb hangsúlyt fektetni emiatt, vagy melyik lényegesebb szakterületen kell felzárkózniuk.

A bibliográfiai adatok elemzése a gyűjtemény minőségi mutatóinak felállításához is hasznos lehet, azáltal, hogy egy-egy dokumentumot összevetnek a különböző listákon található helyükkel, amely listák megmutatják, hogy egy adott korcsoport számára vagy szakterületnek melyek az elismert dokumentumai.[76]

Az ilyen jellegű bibliográfiai adat-elemzést elektronikus könyvtárakban is szinte ugyanúgy lehet alkalmazni, mint a hagyományos könyvtárakban. Annyi különbséggel, hogy a katalogizálás és a MARC-formátumban leírt adatok helyett a bibliográfiai adatok általában metaadatként szerepelnek.[77] Sőt, elektronikus könyvtárak esetében még könnyebb is lehet az elemzendő adatok rendszerből való kinyerése, összeállítása, mivel alapesetben nem feltétlenül a könyvtári integrált rendszer mintáját követve, zárt adatbázisokban tárolják ezeket az adatokat, hanem az elemzésekhez szabadabban kinyerhető formátumokban. A Magyar Elektronikus Könyvtár[78] (MEK) oldalán található statisztikák bizonyítják ezt, sok szempont szerint lehet tájékozódni a MEK

[75] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[76] Uo.
[77] Uo.
[78] A MEK honlapjának címe: http://mek.oszk.hu [2009-08-10].

összetételéről, a dokumentumok számáról, tárgyáról, és így tovább. Nem utolsó sorban fel lehet használni ezeket a statisztikai adatokat az adatbányászathoz, az adatbányászati elemzésekhez. Mivel élő hazai példa, megnézzük részletesebben, milyen információkkal szolgálnak ezek a kimutatások az elektronikus könyvtárról.

3.3.1.1. Adatbányászati példa-elemzés a MEK Statisztika oldalának felhasználásával

A fent említett statisztikák a MEK főoldaláról érhetők el, az oldal alján a jobb alsó sarokban található Statisztika nevű hivatkozásra kattintva.[79] Ahogy az oldalhoz tartozó bevezetőben írják, a kimutatások és forgalmi adatok a mek.oszk.hu szerverén elérhető gyűjteményre vonatkoznak (nem tartalmazzák a régi cím – mek.iif.hu gyűjteményének lekérdezési adatait).[80] Az adatbányászati szempontból értékes és a gyűjteményszervezéshez kötődő statisztikák a következők:

- A gyűjtemény megoszlása (napi frissítéssel)[81]
- Dokumentum-letöltési adatok (havi frissítéssel)[82] és
- Összesített dokumentum-letöltési adatok (a havi letöltési adatok összesítése maximum az utolsó 12 hónap)[83]

3.3.1.1.1. A gyűjtemény megoszlása

A gyűjtemény megoszlása nevű statisztika a MEK állományának aktuális összetételét mutatja meg, adatbányászati szempontból az állományelemzéshez szolgálhat hasznos információkkal. Az adatbányászati elemző számára részletes bontásban mutatja, hogy mely szakterületen, hány dokumentum található a gyűjteményben. A statisztikában látható a dokumentumok száma összesen (jelenleg 7211 db), ebből kiemelve a magyar nyelvűek számát (jelenleg 6599 db). Továbbá tartalmazza a műfaj, témakör – kétféle bontásban: egy átfogóbb (humán területek, kultúra, irodalom; kézikönyvek és egyéb műfajok; műszaki tudományok, gazdasági ágazatok; társadalomtudományok; természettudományok és matematika) és kisebb szakterületekre bontva (anyagtudományok, kohászat; bányászat; etc.), valamint formátum és státusz szerinti bontást. A műfaj szerintiből kiderül, hogy a legtöbb dokumentum a tanulmányokhoz tartozik (1729 db), ezt követi a regény (942 db) és a versek (763 db). Témakör szerint

---

[79] A pontos hivatkozás, A MEK számokban: http://mek.oszk.hu/html/statisztika.html [2009-08-10].
[80] A MEK számokban, http://mek.oszk.hu/html/statisztika.html [2009-08-10].
[81] Megoszlás, http://mek.oszk.hu/html/megoszlas.html [2009-08-10].
[82] Letöltési statisztika, http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=stat [2009-08-10].
[83] Összesített dokumentum-letöltési adatok, http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=stat_rcum , [2009-08-10].

legtöbb dokumentum a humán területek és kultúra, irodalom alatt található (4365 db), ezt követi a társadalomtudományok témakör (3064 db). A formátum-statisztikából kiderül, hogy a legtöbb dokumentumot PDF formátumban tárolják (5813 db), ezt követi a HTML-formátumban tároltak csoportja (3939 db). Valószínűleg azért alakult így, mert ezek a legegységesebben, legprecízebben megjeleníthető formátumok, bár a HTML esetében vitatkozhatnánk ezen, de általában nem túl összetett ezeknek a HTML-fájloknak a kódja, így általában minden böngésző egységesen jeleníti meg ezeket a dokumentum-oldalakat. Az is megfigyelhető, hogy az új dokumentum-formátumok is megjelentek, egyelőre még nem nagy számban, de a későbbiekben valószínűleg növekedni fognak ezek a számok, ilyen a LIT (616 db) és a DjVu (566 db). Kisebb hátránya a formátumok statisztikájának, hogy a fájlformátumok rövidítése szerepel a listán, és nem adják meg a feloldását, de mivel egységesített formátumok és elterjedtek, viszonylag könnyen visszakereshetők egy keresőmotor használatával (pl. yahoo vagy google segítségével). Végül mit olvashatunk ki a státusz statisztikájából? Első látásra nem sokat, hacsak nem ismerjük fejből a listán szereplő rövidítéseket, sajnos itt sem szerepel a rövidítések feloldása, itt ez nagyobb hátrány, mivel eléggé specifikus rövidítések, de ha valaki tájékozott a digitalizált dokumentumok forrásainak világában, akkor néhánynak rájöhet a feloldásra (például a „BHI” = Bibliotheca Hungarica Internetiana). Az sem egyértelmű, hogy mit takar ez a státusz-információ, úgy látszik, hogy sokszor azt, hogy honnan származik az adott dokumentum (például a „BHI”), sokszor pedig, hogy a MEK-ben milyen „státuszban” szerepel, így erre utal az „ÚJ” és a „TÖRÖLT”. A MEK számokban nevű statisztikát gondozó Drótos László MEK-könyvtáros szíves és gyors tájékoztatása alapján kiderült, hogy mit rejtenek ezek a rövidítések, valamint arról is adott felvilágosítást, hogy mire használják ezeket a statisztikai adatokat a MEK-ben. A táblázat e szerint főként belső használatra szánt, azt mutatja meg egyrészt, hogy hány könyv származik a megjelölt forrásokból (például a „BHI” takarja a megszűnt Bibliotheca Hungarica Internetiana projektből átvett könyveket, az „ISZT” az Internet Szolgáltatók Tanácsa támogatásával készült vagy megvásárolt könyveket). Másrészt pedig, hogy a MEK-en belül hány könyvnek mi a státusza (például a „HIÁNY” a valamilyen szempontból hiányos könyvet jelöli), és hogy mik a további teendők vele (például a „KORR” a korrektúrázásra váró könyvek számát mutatja). Adatbányászati szempontból a feladatok megtervezéséhez, és a megfelelő alkalmazottak különböző feladathoz való rendeléséhez lehet hasznos ez a táblázat, annak felfedésével, hogy a számadatok mögött milyen dokumentumok

rejlenek. Például, ha korrektúrára van pénze és embere a könyvtárnak, e lista segítségével és a dokumentumok adataival könnyen kiválaszthatják, melyek a korrektúrázásra váró tételek.

A felhasználás módjairól kiderült, hogy a témakör-megoszlási jelentés a jövőbeli gyarapítás szempontjából érdekes az elektronikus könyvtárnak, ha módjuk van rá, akkor lehetőleg olyan témájú könyveket vásárolnak vagy digitalizálnak, amely témához tartozóból még kevés van. A főoldalról elérhető Sikerlista[84] pedig abban orientálja a könyvtárosokat, hogy melyek a keresett írók, témák és műfajok. Drótos László azt is hozzátette, hogy a MEK-nek kevés befolyása van arra, hogy mi kerül a gyűjteménybe, mert nagyrészt kapják a dokumentumokat, és nem ők végzik a digitalizálást. Azt is kiemeli, hogy leginkább a saját web szerverük statisztikáját figyelik (Webalizer[85]), főként az éves trendeket, amelyek megmutatják, hogy hogyan növekszik a MEK ismertsége és forgalma. Felhívja a figyelmet arra a belső információra is, hogy 2008 decembere óta két szerverük van (a nagy terhelés miatt), a mek.oszk.hu címen lévő szerver és a mek.niif.hu tükörszerver, így a Webalizer statisztikájának számadatait meg kell duplázni, hogy reális képet kapjunk. Drótos László beszámolója alapján egy pozitív példát láthatunk a meglévő statisztikai adatok felhasználására, ha nem is tudatosan, adatbányászathoz hasonló módszereket használhatnak, konstruktívan állnak hozzá a számadatokból kinyerhető hasznos információkhoz.

Mint láthattuk, alapszintű elemzések levonhatók első ránézésre is az ilyen jellegű statisztikai adatokból, aki hivatásos elemző és ismeri az adott könyvtár (itt a MEK) gyűjteményszervezési céljait, terveit és a befolyásoló tényezőit valószínűleg ezeknél értékesebb összefüggéseket tud kiolvasni az adatokból.

3.3.1.1.2. Dokumentum-letöltési adatok és összesített dokumentum-letöltési adatok

Mivel a két statisztikai táblázat felépítése ugyanolyan, csak az időintervallumban különböznek, együtt tárgyalom ezeket.

A letöltési statisztikák a mindenkori előző hónap vagy előző év dokumentum letöltéseit mutatják szerző szerinti ábécé sorrendben. Pozitív, hogy a statisztikák bevezető leírásában megadják, hogy hogyan kell értelmezni a táblázatban szereplő számokat, e szerint:

„A felső szám az egyetlen fájlból álló (HTML, RTF, PDF, ZIP stb.) művek letöltéseinek számát jelzi. Az alatta levő halványabb érték pedig azt mutatja, hogy a

[84] A sikerlista elérhetősége: http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=siker [2009-09-08].
[85] Webstatisztika mek.oszk.hu: http://mek.oszk.hu/webalizer/ [2009-09-08].

többdarabos verzió - ha van ilyen - saját kezdőlapját hányan nézték meg. Ennél ugyanis az egyes részek külön-külön való mérése nagyon félrevezető lenne, így csak az index.htm nyitóoldalt számoljuk. A jobbszélső oszlop e két szám összege."[86]

A szerzői ábécé sorrend miatt az adatbányászati elemző akkor kap használható információt belőle, ha arra kíváncsi, hogy egy szerző műveit vagy egy adott művét hányszor töltötték le az előző hónapban vagy az előző évben. Megtudhatjuk például, hogy Rejtő Jenő: Az elveszett cirkáló című művét 145-ször töltötték le az elmúlt hónapban, a rovásírással írt változatát pedig lényegesen kevesebbszer, 45-ször, de A boszorkánymester című művéhez képest ez a rovásírásos letöltés több mint kétszeresen nagyobb (ott 174 és 17 a letöltések száma).[87] A szerző szerinti betűrendes szerkezetű lista mellett a fentebb említett „Sikerlista" hivatkozásról[88] elérhető oldalról tudhatjuk meg, hogy melyek a legnépszerűbb könyvek a letöltésük gyakorisága szempontjából.

Összességében nagyon pozitív, hogy a Magyar Elektronikus Könyvtár készít, rendszeresen frissít és közzétesz ilyen jellegű használati statisztikákat, lehetőséget adva vagy felmutatva a lehetőségét az elemzések vagy konkrétabban az adatbányászati elemzések készítésének.

3.3.2. Állománygyarapítás és -apasztás

A továbbiakban nézzünk állománygyarapítási, szerzeményezési alkalmazási példákat. Ha az adott könyvtár az integrált könyvtári rendszerében nyomon követi a dokumentumok rendelését a kiválasztástól a beérkezésig és a könyvtárban való feldolgozásig, olyan hasznos, elemzendő információkra tehet szert, mint a rendelés forrása (melyik kiadótól, beszerzőtől), a szerzeményező könyvtáros, a mű ára, a rendelés és a beérkezés dátuma és az esetlegesen fellépő problémák a rendeléssel. Valamint néhány esetben akár arról is, aki kiválasztotta, kérte az adott dokumentumot.[89] Nicholson szerint a könyvtárak azon kívül, hogy feltárják a beszerzőkkel történt problémákat, ritkán elemzik ezeket az adatokat a kompetitív előny eléréséhez.[90] Rámutat továbbá, hogy ezek az adatok főleg logisztikaiak, amelyek segítségével rábukkanhatnak például a késő rendelés okára, valamint arra, hogy ezen kívül még sok

---

[86] Letöltési statisztika: http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=stat [2009-08-10].
[87] Letöltési statisztika – R betű: http://mek.oszk.hu/html/vgi/kereses/keresesuj.phtml?indextomb=&mod=keres&offset=0&tip=stat&szerzo_abc=r [2009-08-10].
[88] [MEK–Sikerlista]: http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=siker [2009-09-14].
[89] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[90] Uo.

feltáratlan kutatási területet tartalmaz az információforrások árainak megértéséhez, az irányításukhoz és az árak előrejelzéséhez.[91]

A másik gyűjteményszervezéshez sok hasznos információval szolgáló forrás a kölcsönzési információk, amely adatcsoportok segíthetnek a vásárlásban és a selejtezésben is, például úgy, hogy a gyakran kölcsönzött dokumentumok esetében megfontolhatja az adott könyvtár az új példányok beszerzését, a ritkán kölcsönzöttekből esetében pedig ötleteket kaphat a selejtezendő példányokhoz.[92] A gyakran kölcsönzött dokumentumok mintázatát elemezve a könyvtár előre jelezheti a jövőbeli igényeket, hogy valószínűleg milyen jellegű dokumentumokra lesz szükség és mennyit célszerű ezekből rendelniük.[93] (Például, ha egy népszerű, több kötetből álló regény második részét gyakran kölcsönzik, és tegyük fel még nem jelent meg a harmadik része, már előre számolhatnak a következő rész megvételével, és azt is megbecsülhetik a második rész kölcsönzései alapján, hogy minimum hány példányra lesz szükség.)

A másik oldalt nézve, a hiány, veszteségek megelőzése érdekében is végezhetnek mintázatkereséseket az elveszett vagy ellopott könyvek adatait és ezzel párhuzamosan a magas olvasói büntetéseket elemezve. Ezek az információk segíthetnek a megfelelő előírások meghozatalához, a veszteségek megelőzése érdekében. A hiányosságok jobb megértéséhez elemezhetik a sikertelen kéréseket vagy kereséseket, ezeknek két forrása lehet: a tájékoztatópultnál lezajló sikertelen kérés és az OPAC-ban történő, a dokumentum hiánya miatt eredménytelen keresés.[94] Mindkét forrás azt igényli, hogy dokumentálják a sikertelen tranzakciókat, így a tájékoztató könyvtáros részéről a hiányzó könyvet, az OPAC esetében pedig azt, hogy kinyerjék a rendszerből a hiányzó dokumentum miatti eredménytelen (vagy 0 eredményt hozó) lekérdezéseket. Az előbbi esetben sok könyvtárban megvan ez a dokumentálás dezideráta-lista formájában, ennek feldolgozása, elemzése, minták keresése a kért dokumentumokban az adatbányászati munka.

3.4. A könyvtári adatbányászat alkalmazása a tájékoztatásban és a felhasználói kapcsolattartásban

Továbblépve a könyvtári adatbányászat tájékoztatásban és az olvasói kapcsolattartásban betöltött funkcióira, először értelmezzük, hogy az adatbányászattal kapcsolatban mit értek a tájékoztatásban és az olvasói kapcsolattartásban való alkalmazáson. A

[91] Uo.
[92] Uo.
[93] Uo.
[94] Uo.

tájékoztatásban olyan területeken lehet hasznos az adatbányászat, amelyeken lehetőségünk van az eljárás alkalmazásával megkönnyíteni, praktikusabbá, felhasználóközpontúbbá tenni a tájékoztatás elemeit. Azért teheti felhasználóközpontúbbá, mert az adatbányászati eljárás alapján tett elemzések felhasználása felhasználóktól származó használati adatokat, vagy megjegyzéseket, könyvtárosi megfigyeléseket párosítja a könyvtáros szakértelmével, amellyel képes az adatokat a felhasználó számára praktikusan értelmezhető formába rendszerezni. Erre a későbbiekben látunk példát. Az olvasói kapcsolattartáshoz való felhasználásba beleértendő az olyan jellegű eszközök vagy módszerek használata (ezeket lehet kifejleszteni, felismerni az adatbányászati elemzések felhasználásával), amelyek megkönnyítik, praktikusabbá teszik az olvasókkal való kommunikációt, az olvasók elérését. Ide sorolom például a felhasználói magatartás és olvasói szokások elemzését és megértését.

Elmondható, hogy az alkalmazásnak ez az egyik leggyümölcsözőbb területe. A leggyümölcsözőbb, és egyben a személyi jogsértés területén a legnagyobb veszélyeket is hordozó, így ha az adatelemzőnek, a felhasználókról szóló személyesebb információk elemzése közben, különös körültekintéssel kell elvégeznie az adatok anonimizálását, például kódokkal való cseréjét vagy egyéb módszert alkalmazását, amellyel törli a személyes információkat, amelyekkel egyértelműen azonosítható lenne az adott felhasználó. Színesítésképp röviden leírom, az egyik amerikai krimi[95] példáján át, hogy hogyan lehet visszaélni a személyes adatokkal. A krimi történetében valószínűleg az adatbányászat eljárását alkalmazták személyiségi jogsértő módon, még ha kitalált történet is, figyelemreméltó jelenséget szemléltet. A nyomozás során a detektívek úgy jutottak el a sorozatgyilkoshoz, hogy a gyilkosságok alapján meg tudták mondani, hogy milyen könyveket olvas a keresett személy, az egyik detektív egy FBI-os kollégájához fordulva segítségért kapott egy listát a környékbeli könyvtár kölcsönzéseiről, és az egyik illetőre pont ráillett a detektívek előzetes listája, történetesen pont a sorozatgyilkosra. Az FBI részéről úgy állították be ezt az adatgyűjtést, hogy a szélsőséges dokumentumokat olvasók neveit gyűjtik egy külön adatbázisba, ezeket nyilvánosan nem használhatják fel, viszont sokszor segítheti a nyomozás egy-egy lépését. A személyiségi jogok és adatbányászat kapcsolatára még külön kitérünk a 4., Az adatbányászat előnyei, hátrányai és egyéb kérdései című fejezetben.

[95] A film címe: A hetedik, ismertetése az IMDB film-adatbázisában: http://www.imdb.com/title/tt0114369/ [2009-08-11].

A 3.2.4. pontban, a döntéstámogatáshoz való alkalmazás leírásakor már érintettünk egy tájékoztatásban való alkalmazási példát, a DREW (The Digital Reference Electronic Warehouse – Digitális Tájékoztatási Elektronikus Tárház) elnevezésű adattárház-projektet az OLAP-eszközök használata kapcsán. A DREW projektet 2005-ben hívta életre Scott Nicholson és David Lankes[96], célja (volt – mivel ma a honlapja alapján[97] nem látszik működni) egy multidiszciplináris tudásbázis építése a digitális tájékoztatási folyamat mélyebb megértése elősegítőjeként.[98] A rendszert a tervezetében egy XML-sémára építették, amely egy helyen tárolja a különböző forrásokból, kutatóktól érkező tájékoztatási tranzakciókat egy megosztott archívumot építve.[99] A projekt első megvalósulni készülő alkalmazása a Reference Extract (REX), egy keresőmotor, amely az eredeti tervek szerint a digitális tájékoztatási tranzakciók nyilvános archívumában keres.[100] A Reference Extract projekt aktuális weboldalán láthatunk a készülő keresőről három videót, amelyekben David Lankes mutatja be a tervezetet.[101] Megtudhatjuk ezekből, hogy egy olyan keresőmotor indítását tervezik, amely a tájékoztató könyvtárosok által megbízható forrásként megjelölt adatforrásokban keres. A DREW projektet nem említi egyik videóban sem, vagy a háttérben dolgoznak vele vagy Lankes Nicholson és a DREW projekt nélkül folytatta a Reference Extract projektet, erre utal az is, hogy a Lankes „Details” című videóban a tájékoztatási adatbázissal, megbízható gyűjteményekkel rendelkezőket együttműködésre, a projekt támogatására bíztatja. Ha ez valóban így van, akkor valamelyest eltávolodtak az adatbányászat-alapú DREW-projekttől, bár kétségtelenül így is reményt keltő az internetes tájékoztatás területén az elképzelés, de (valószínűleg mivel még tervezet) nincs elég nyilvános információ a hátteréről, hogy milyen eszközökre épít, használ-e adatbányászati eljárásokat, ha igen, hogyan, így mélyebben nem lehet vizsgálni adatbányászati szempontból.

---

[96] NICHOLSON, Scott: The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services [elektr. dok.] http://www.bibliomining.com/nicholson/nicholsonbibliointro.html [2009-08-11].

[97] A honlapon csak egy „iis” szöveg olvasható: http://drew.syr.edu/ [2009-08-11].

[98] NICHOLSON, Scott: Bibliomining Applications in Digital Reference: Using Data Warehousing and Data Mining to Improve Management and Decision-Making. [elektr. dok.].

[99] A DREW projekt 2005-ös weboldala az Internet Archive Way Back Machine web-archiváló keresővel előhívva: http://web.archive.org/web/20050723011710/http://drew.syr.edu/ [2009-08-11].

[100] LANKES, R. David – NICHOLSON, Scott: Reference Extract: Extending the Reach of Digital Reference through Collaborative Data Warehousing [elektr. dok.]. http://www.ieee-tcdl.org/Bulletin/v2n1/nicholson/nicholson.html [2009-08-11].

[101] Planning Reference Extract, http://referencextract.org/ [2009-08-11].

3.4.1. A tájékoztatási szolgáltatások fejlesztése

Egyszerűbben, akár adattárház nélkül is alkalmazható eljárás, ha a tájékoztatási szolgálat a gyakran előforduló korábbi kérdésekből és válaszokból készít egy külön gyűjteményt, megelőzve az újra és újra felmerülő kérdések és válaszok ismételgetését. Egy egyedi felépítésű Gyakran Ismételt Kérdések (GYIK vagy angol rövidítéssel FAQ – *Frequented Asked Questions*) oldallal rendelkezik a las vegas-i Clark County Library. Ebben a leggyakoribb öt féle kérdéstípust csoportosították, úgy jelenítve meg, hogy a kérdés eleje mindig ugyan az (pl. „*Can I...*" – „Tudok-e..."), és a csoportok melletti legördülő menüből választhatják ki a felhasználók az őket érdeklő kérdéseket.[102] Nem tudható biztosan, hogy adatbányászati eszközöket használtak az összeállításhoz, de feltételezhető. Mindenesetre jól szemlélteti, hogy ilyen jellegű, ötletesebb, praktikusabb formában is létre lehet hozni az adatbányászati eszközök segítségével elkészített új kérdés-válasz gyűjteményt.

3.4.2. A felhasználói szokások és a felhasználói magatartás megértése

A felhasználói viselkedés megértéséhez, elemzéséhez fontos információforrások lehetnek a web szerverek naplózásai, abban az esetben, ha a könyvtár technikusai megtalálják a megfelelő módját az adatbányászathoz hasznos naplózás-részek kiszűréséhez. Ekkor információkat kaphat a könyvtár arról, hogy mit keresnek a felhasználók, megtalálták-e azt, megkapták-e a választ a kérdésükre (ha elektronikus úton tették fel), ezen információk birtokában a könyvtárosok jobban hozzájárulhatnak a felhasználói kapcsolatok támogatásához és a tudásmenedzsment segítéséhez. Az is kiderülhet ezekből a naplózásokból, hogy melyek bizonyultak a leghasznosabb információforrásoknak, ezeket tudatosítva a könyvtárak stratégiai haszonra tehetnek szert[103], pontosabb képet kapva arról, hogy mely információforrások az erősségei a könyvtáruknak.

A web szerver és külső látogatottság- és felhasználói aktivitás-mérő szolgáltatások naplózásainak statisztikájára és az elemzés lehetőségeire példaként nézzük meg a MEK Statisztika oldaláról[104] az ide vonatkozó kimutatásokat. Ezek a következők:

- Medián webaudit (napi frissítéssel)[105]

---

[102] A FAQ a könyvtár odalán: http://www.lvccld.org/faqs/index.cfm [2009-08-11].
[103] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[104] A MEK számokban, http://mek.oszk.hu/html/statisztika.html [2009-08-11].
[105] Webaudit, http://webaudit.hu/object.8D7ED62D-1CED-4E84-BAD3-4944E7412406.ivy?session.folder=10136 [2009-08-11].

- A honlap forgalma (Extreme tracking – folyamatos frissítéssel)[106]
- A web-szerver forgalmi adatai (Webalizer – napi frissítéssel)[107]

Sajnálatos módon az első két kimutatás jelenleg nem érhető el, a Medián webaudit-ra mutató link nem a hivatkozott oldalra ugrik, hanem a szolgáltató főoldalát hozza be, az Extreme tracking látogatottság-mérő pedig a statisztikai adatokra téve a felhívás szövegdobozát, azt jelzi, hogy a használt verzió elavult és kéri, hogy frissítsenek az újabbra.

A harmadik, a web-szerver forgalmi adatai, amelyik az OSZK szerverén található, viszont működik és a kért oldalt tölti le, a vizsgálódás szempontjából végeredményben ez a legfontosabb a forrás három közül (bár a korábban működő Medián webaudit is érdekes használati adatokkal szolgált).

A web-szerver forgalmi adataiból igen sok szempontból tájékozódhatunk a mek.oszk.hu oldal látogatottságáról, oldalainak letöltéséről és gyakorisági adatokról. A Webalizer főoldalán a mindenkori utolsó 12 hónap statisztikáját kérhetjük le az aktuális hónaptól kezdve. A hónapok felsorolása mellett láthatunk egy összesített statisztikát, ez az elmúlt 12 hónapra vonatkozik. Ebből egy adatelemző kiolvashatja, hogy mikor volt legnagyobb a lapok, fájlok és a keresési találatok letöltésének száma. Megfigyelhető, hogy az őszi és téli hónapokban mind magasabb, míg áprilistól kezdve erőteljesen csökken. A hónap statisztikáinak oldalán egy-egy külön táblázatban jelenítik meg a különböző kimutatásokat. Példaként nézzük meg, hogy mit tudhatunk meg adatbányászati szempontból egy már lezárult hónap, 2009 júliusának tábláiból. A navigációt megkönnyítik az oldal felső részén elhelyezett ún. horgonyok, amelyek egyben a táblák tematikai felosztását is mutatják. Megtudhatunk napi és óránkénti adatokat, valamint láthatjuk a legsűrűbben nézett URL-ek és keresési szavak listáját, ki- és belépési információkat, adatokat a bejelentkezett felhasználókról és a felhasználók által használt böngészőkről.

Adatbányászati szempontból a felhasználói szokásokat és felhasználói magatartást vizsgálva leghasznosabb információkat a legsűrűbben nézett URL-ekből, a Be- és kilépési lapokból és a Keresési szavak listájából tudhatunk meg. A legsűrűbben nézett URL-címeknél a MEK kereső-oldala áll az első helyen, úgy látszik, a legtöbb felhasználó az egyszerű keresést használja. Az Összetett keresés oldala is szerepel a listán, méghozzá a 31. helyen, ezt is gyakran használják, úgy látszik, ez általános

[106] [A honlap forgalma], http://extremetracking.com/open?login=mek2adm [2009-08-11].
[107] Webstatisztika mek.oszk.hu, http://mek.oszk.hu/webalizer [2009-08-11].

tendenciának mondható, az ezt megelőző hónapban, és a legnagyobb látogatottságot mutató, 2008 novemberében is a 27-28. helyen szerepelt. Az Összetett keresés oldalát megelőzi a MEK akadálymentes változatának keresője, az RSS hírcsatornája, a MEK könyvkereső oldala, a Pallas Nagylexikon keresőoldala és a Humán területek, kultúra, irodalom nevű témakör weblapja. Érdekes, hogy Humán területek, kultúra irodalom témakörének weblapjának népszerűsége egybeesik azzal az adattal, hogy ebben a témakörben tárolja a MEK a legtöbb dokumentumot, ahogy a gyűjteményszervezéssel kapcsolatos statisztikák elemzéséből kiderült. Ezt tekinthetjük nem triviális úton felfedezett összefüggésnek, az ilyen összefüggések feltárása a tudásfeltárás lényege, és a feltárás eszköze az adatbányászat. Ezen statisztikai adatok lehetséges elemzésére, kiválogatására mutattunk ezzel egy példát, a Magyar Elektronikus Könyvtár munkatársai, elemzői – ha vannak erre szakosodottak valószínűleg sokkal több hasznosítható információt tudnának kiszűrni ezen adatok elemzésével.

Elkanyarodva MEK Statisztikája kapcsán felhozott példától a felhasználói szokásoknak egy másik információforrása lehet a helyben használat elemzése, amellyel elemzendő mintákat kaphatnak a könyvtárhasználatra.[108] Ehhez arra van szükség, hogy nyomon kövessék a helyben használatot, rögzítsék (legcélszerűbb számítógépen) a helyben használt dokumentumok adatait, mielőtt azok visszakerülnének a helyükre. Ehhez hasonlóan azt is nyomon követhetik a rejtett összefüggések feltárásához, hogy a könyvtárközi kölcsönzések során milyen dokumentumokat kérnek, hogyan használják azokat (helyben olvasásra és/vagy fénymásolásra) és hogy melyek a leggyakoribb kölcsönadó könyvtárak.[109]

A felhasználói szokásokat a könyvtárak gyakran különböző felmérésekkel igyekszenek vizsgálni. Az adatbányászat alkalmazása részben megkérdőjelezi a felmérések szükségességét vagy kizárólagosságát, mivel gyakran olyan adatokra is rákérdeznek ezekben a felmérésekben, amelyeket maguk is ki tudnának deríteni az integrált könyvtári rendszerükből.[110] Az OSZK Könyvtári Intézet Könyvtörténeti és Könyvtártudományi Szakkönyvtár 2007-es évi elégedettségi kérdőívét[111] átnézve azokra a kérdésekre, hogy „Mióta beiratkozott olvasója a Könyvtártörténeti és

[108] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[109] Uo.
[110] ESTABROOK, Leigh S.: Sacred trust or competitive opportunity: Using patron records. [elekr. dok.] = Library Journal, 2. sz. 1996. EBSCO azonosító: 03630277 [2009-08-11].
[111] Könyvtári Intézet Könyvtártudományi Szakkönyvtár: Elégedettségi kérdőív az olvasói igények felmérésére 2007. 04. 16. – 2007 06. 30.

Könyvtártudományi Szakkönyvtárnak (KSZK)?" vagy az életkort érintő kérdésre egy megfelelő statisztikát készítő vagy adatelemzésre képes integrált könyvtári rendszerrel bírva annak használatával is tudtak volna választ kapni, abban az esetben, ha csak erre az adatra kíváncsiak. Sőt, azzal az előnnyel, hogy a vizsgálatban a könyvtár összes beiratkozott olvasója szerepel, nem csak azok köre, akik kitöltötték az adott kérdőívet.

3. 5. Egy komplex külföldi példa bemutatása az adatbányászatra: Penn Library Data Farm

A Penn Library Data Farm a pennsylvaniai egyetem könyvtárának „adatfarmja", ahol olyan információkat tárolnak a könyvtár használatáról, amelyek még feldolgozatlanok, arra az esetre, ha felmerülne egy kérdés, amelyet ezen adatok elemzésével könnyen meg tudnak válaszolni. Az adatfarmon adatokat találunk a szervezetről, a tevékenységeiről, a könyvtári tranzakciókról és a felhasználókról. Az adatfarm létének célja a felhasználók jobb szolgálatának segítése.[112] Az idézett tanulmány írásakor (2005-ben) még csak tervbe volt véve a jövőbeli használata: a könyvtári munkafolyamatok megfigyelése, mára a honlap alapján úgy látszik, hogy megvalósult, de csak a bejelentkezett felhasználók érhetik el az adatok személyes természete miatt.[113] Az adatfarmon gyűjtött adatok felhasználhatók mind a menedzsment és döntéstámogatás, a gyűjteményszervezés, mind a tájékoztatás és olvasói kapcsolattartás céljára, ezért tárgyaljuk ezt a sokrétű példát egy külön alfejezetben, de az alfejezeten belül megtartva a könyvtári munkafolyamatok szerinti besorolást adatbányászat felhasználási lehetőségeinek bemutatásakor.

Először tehát nézzük, hogy az adatfarm honlapján[114] található lekérdezések közül melyek segíthetik a menedzsmentet és a döntéstámogatást. Azért beszélek lekérdezésekről, mert az adatfarm honlapján tulajdonképpen csak erre van lehetőségünk, az elemzést az egyetemi könyvtár könyvtárosai végzik, így arra már nincs rálátásunk, de a lekérdezések alapján így is képet kaphatunk arról, hogy a könyvtár milyen célokból hasznosíthatja ezeket az elérhetővé tett használati adatokat.

Az elektronikus folyóirat-adatbázis előfizetési döntésekhez használható az „E-Resource Tracking"[115] nevű oldal, amelyen folyóiratcím, forrás-típus (pl. e-könyv, e-folyóirat, kép) és tárgykör szerint kérdezhetjük le a könyvtár ERED[116] nevű oldalán

[112] CULLEN, Kevin: Delving into Data. [elektr. dok.].
[113] About the Penn Library Data farm: http://datafarm.library.upenn.edu/census/index.html [2009-08-26].
[114] Penn Library Data Farm: http://datafarm.library.upenn.edu/ [2009-08-27].
[115] USR Report Builder: http://datafarm.library.upenn.edu/cgi-bin/ursform/ursform.pl [2009-08-27].
[116] E-resources search: http://www.library.upenn.edu/cgi-bin/res/sr.cgi [2009-08-30].

összegyűjtött elektronikus források a használati adatait. A folyóirat címeknél szabad szavas keresésre, a másik kettőnél pedig kötött szavas keresésre van lehetőségünk. A kezdeti találati listán belül jobban specifikálhatjuk, hogy pontosan mely folyóirat(ok)ra vagy más forrás(ok)ra, milyen időszakra és milyen használati módokra (egyetemen belüli vagy kívüli) keresünk. A végeredményül kapott oldalon pedig megtekinthetjük az adott források használatának számát. A könyvtári alkalmazottakkal kapcsolatos döntéshozáshoz hasznos lehet a Gate Counts[117] hivatkozásról elérhető belépésszámláló, amely két könyvtárába való felhasználói belépéseket számolja. Kérhetünk hét napjai és órái szerinti bontást, amelyből megállapítható, hogy mely órákban a legnagyobb a könyvtárak látogatottsága, tehát mikor van szükség több alkalmazottra. Egy másik szempontú lekérdezésből pedig megtudhatjuk, hogy mely egyetemi részlegekből mennyien látogatják a könyvtárat, ez abban lehet a döntéshozók segítségére, hogy milyen szaktudású tájékozató könyvtárosokra lehet főleg szükség. Ezeken kívül még van egy-két jelentés, amely hasznos lehet menedzsmenthez és a döntéstámogatáshoz, de ezeket csak könyvtári azonosítóval és jelszóval vagy könyvtári e-mail címmel lehet elérni. Ilyen a könyvtár integrált könyvtári rendszerével, a Voyager-rel készített Voyager Fund[118] kimutatás, amely a könyvtár kiadásait mutatja meg. Egy másik Voyager-rel készített lekérdezéshez viszont hozzáférhetünk, amely a kölcsönzési adatokról, a szekrénykulcs – és laptop használatokról ad jelentéseket. A szekrénykulcs használati jelentésének sikeres adatbányászati elemzésére és felhasználására olvashattunk egy gyakorlati példát a 2.2. Adatbányászat és integrált könyvtári rendszerek című alfejezet végén.

A gyűjteményszervezés segítésére a Penn Library Data Farm oldaláról használható a Borrow Direct Data Repository[119] és az EZBorrow Data Repository[120]. Mindkét lekérdezőfelület a könyvtárközi kölcsönzésekről készít jelentést, az első a BorrowDirect nevű konzorcium tagjai közötti könyvtárközi kölcsönzésekről, a második pedig az EZBorrow nevű konzorcium tagjai közöttiekről. Ezen lekérdezések egy része csak a konzorciumban szereplő könyvtár azonosítójával tekinthető meg. Két lekérdezéstípust tekinthetünk meg szabadon: egy összesített jelentést a Library of Congress osztályozási rendszerével listázottat és egy osztályozási rendszer feltüntetése nélkülit, valamint egy egyes könyvtárakra lebontott jelentést. A lekérdezésből megtudhatjuk, hogy melyik

[117] Data Farm – Visitor Reports: http://datafarm.library.upenn.edu/visit_gatecounts.html [2009-08-30].
[118] Voyager Fund: http://proxy.library.upenn.edu:4500/loggedin/df-vfunds.html [2009-08-30].
[119] Borrow Direct Data Repository: http://datafarm.library.upenn.edu/bdrpt.html [2009-08-30].
[120] EZBorrow Data Repository: http://datafarm.library.upenn.edu/bdrpt_ez.html [2009-08-30].

tagkönyvtárból kérték a legtöbb könyvet, hogy melyik könyvtárnak milyen arányban sikerült teljesíteni a kéréseket, valamint azt is, hogy mennyi idő telt el a kérés és a teljesítés között. A döntéshozók vagy könyvtárközi kölcsönzésért felelősek segítséget kaphatnak annak eldöntéséhez, hogy mely könyvtárakkal számolhatnak bizonyosan eredményes könyvtárközi kölcsönzésekkel. A könyvtárközi kölcsönzések másik lekérdezőfelületén arról kaphatunk adatokat, hogy mely témakörök a legnépszerűbbek, melyek dokumentumait kérték a legtöbbször. A gyűjtemény összetételének megértéséhez szolgálhat hasznos adatokkal a Collection Management Report Builder[121] nevű lekérdezőfelület, amely a Voyager könyvtári rendszerrel készült lekérdezések közé tartozik és csak könyvtári azonosítóval kérdezhető le. Hasonló célokat szolgálhat a Collection Inventory[122] felület, amelynek lekérdezései szintén csak azonosítóval érhetők el.

A felhasználói szokások tanulmányozásához, megértéséhez hasznosak lehetnek a fentebb említett Gate Counts nevű hivatkozásról elérhető könyvtári belépés-számláló, valamint a szintén fentebb hivatkozott Charge Location Circulation linkről elérhető lekérdezések, különösen a szekrénykulcs- és laptop-használati adatok valamint a felhasználói csoportok szerint készített kölcsönzési számláló[123].

Összességében a Penn Library Data Farm a nagyon sok szempontú lekérdezőfelületeivel, az interneten elérhetővé tett lekérdezési oldalaival hasznos segítsége lehet az egyetemi könyvtárnak főleg a döntéshozás és a gyűjteményszervezés területén, legkevésbé talán a felhasználói szokásokat elemezhetik a segítségével, de ha az első kettőt eredményesen, a felhasználók érdekeit szem előtt tartva használják fel, minden bizonnyal a felhasználói elégedettséghez is hozzájárulnak, még ha indirekt módon is a szolgáltatások megfelelő szervezésével. A Penn Library Data Farm egy jó példa arra, hogy hogyan hasznosítható az adatbányászat a könyvtári területen, az alkalmazás sikerességét mutatják a pozitív alkalmazási példák is.

[121] Collection Management Report Builder: http://datafarm.library.upenn.edu/collrpt.html [2009-08-30].
[122] DataFarm – Inventory Reports: http://datafarm.library.upenn.edu/inventory2.html [2009-08-30].
[123] Voyager Reports – Page Data Farm: http://datafarm.library.upenn.edu/circptrdbrd.html [2009-08-30].

4. A KÖNYVTÁRI ADATBÁNYÁSZAT ELŐNYEI, HÁTRÁNYAI ÉS EGYÉB KÉRDÉSEI

4.1. A könyvtári adatbányászat előnyei

Az adatbányászat könyvtári területen történő felhasználásának fő előnye, hogy az adott könyvtár a saját adataival dolgozhat, állapíthat meg addig nem ismert összefüggéseket, mintázatokat, amelyek segíthetik a könyvtárban a döntéshozást, a szolgáltatások fejlesztését, a várható szükségletek előrejelzését és a felhasználók igényeinek és tevékenységeinek mélyebb szintű megértését és a felmerülő igények kielégítését.

A döntéstámogatást tekintve számít leginkább pozitívnak az adatbányászat alkalmazásakor, hogy a saját adataival dolgozhat a könyvtár vagy ha esetlegesen megosztott adattárházat használ, akkor akár más, hasonló státuszú, használói körű könyvtárak adataival is összehasonlíthatja saját működési adatait, állapíthat meg összefüggéseket. A szolgáltatások fejlesztéséhez az adatbányászat alkalmazásának hangsúlya azon van, hogy elemezze a meglévő szolgáltatásokat, azok sikerességét, kihasználtságát, megosztott adattárház használata esetén pedig akár összehasonítsa a sajátjait más könyvtárak szolgáltatásaival. A várható szükségletek vagy éppen a szükség hiányának felismerésében is a mintázatok elemzésének a lehetősége a legfőbb pozitívum. A felhasználók szokásainak, tevékenységeinek vizsgálatához szükséges adatok talán a legkevésbé dokumentáltak, legkevésbé vannak úgy tárolva, hogy kinyerhetők legyenek egy adatbázisból, ezért itt sok múlik azon, hogy a könyvtárban mennyire dokumentálják majd elemzik a felhasználókkal folytatott interakciókat, például a tájékoztatási pultnál elhangzott kérdéseket vagy esetlegesen az online felületre érkezett kérdéseket. A többi felsorolt területet tekintve megfelelő adatbányászati eljárással kinyerhetők az elemzéshez a működési adatok, egy jó elemző képességű könyvtáros pedig értékes információkat, összefüggéseket tud feltárni bennük.

Egy 2002-ben megjelent, könyvtári adatbányászatról szóló tanulmányban[124] Tóth Erzsébet is számba veszi annak előnyeit és hátrányait, de abban az adatbányászat könyvtári területen való alkalmazását szűkebb keresztmetszetben vizsgálja, mint jelen dolgozat. Az említett tanulmányban az adatbányászat alkalmazását leginkább a könyvtár által szolgáltatott adatbázisokban történő keresések megkönnyítésére érti valamint úgy nevezi meg az eljárást, mint „korszerű döntéstámogató rendszert”.[125] Az

[124] TÓTH Erzsébet: Adatbányászatra irányuló törekvések könyvtári területen. [elektr. dok.].
[125] Uo.

adatbányászat előnyeit két pontban emeli ki, az első: „Gyorsabb és alaposabb dokumentum hozzáférés megvalósítása a hagyományos katalógushoz képest."[126] A szerző dokumentum hozzáférést itt valószínűleg a teljes szövegű adatbázisokban található dokumentumok elérésére érti, ahogy az általa fentebb írtakból következik („A teljes szövegű adatbázisok az on-line katalógushoz képest jobban megfelelnek az adatbányászati technológiák követelményeinek, hiszen ez utóbbi frissítése meglehetősen nehézkes és költséges."[127]) Az első előnyt arra értheti, hogy adatbányászati technológiák segítségével csoportosítják a szöveges adatbázisban található információkat, dokumentumokat, és ezen csoportosításon belüli keresés segíti a felhasználót a megfelelő dokumentum megtalálásában. A Tóth Erzsébet által említett másik előny pedig: „A keresett információ könnyen megtalálható a gyűjteményben anélkül, hogy a felhasználók külön segítséget kérnének a könyvtárostól."[128] Ezt a tanulmányából következően úgy kell érteni, hogy a könyvtárosok az adatbányászati elemzések segítségével a gyűjteményt (valószínűleg elektronikus gyűjteményt) előre úgy strukturálták, hogy megkönnyítse a keresést a felhasználók számára. Sajnos a szerző nem fejti ki bővebben, hogy hogyan érti ezen előnyöket, ezért szorítkoztam arra, hogy saját következtetéseket vonjak le a tanulmány többi része alapján, amelyek szintén elég tömören vannak megfogalmazva. Az adatbányászat hátrányainál már könnyebb lesz a helyzet ezt a tanulmányt nézve, mert részletesebben kifejti a felsorolás pontjait.

4.2. A könyvtári adatbányászat hátrányai

Az adatbányászat hátrányai főként technikai feltételekből, szükségletekből adódnak, nem vagy csak kis részben alkalmazási hátrányból. A technikai feltételeket nézve az egyik legnagyobb hátránya vagy negatívuma, hogy a valódi alkalmazásához költséges befektetéssel egy adattárházat kell létrehozniuk a könyvtáraknak, és oda áttölteni a könyvtár működési adatait, főleg az integrált könyvtári rendszerből. Itt felmerülhet az adatok konverziójának problémája, amelynek leküzdésében az integrált könyvtári szoftver forgalmazói tudnának segíteni (mivel ezek az alkalmazások általában zárt forráskódúak, és csak a forgalmazó technikusai férnek hozzá közvetlenül a bennük levő adattáblákhoz), valószínűleg költséges megoldással. Ezt a hátrányt enyhítendő, ahogy fentebb elhangzott, alkothatnának konzorciumot a könyvtárak, amelyek adattárházat

[126] Uo.
[127] Uo. a szerző által itt hivatkozott tanulmány: BANERJEE, Kyle: Is data mining right for your library? = Computers in Libraries, 18. köt. 10. sz. 1998. pp. 28-31.
[128] TÓTH Erzsébet: Adatbányászatra irányuló törekvések könyvtári területen. [elektr. dok.].

kívánnak használni, és osztott adattárházként használhatnák ugyanazt, amelynek előnyei is lennének az adatok összehasonlítása szempontjából. Ehhez kapcsolódik, hogy a magas szintű elemzéshez és az elemzések viszonylag könnyű és átlátható elkészítéséhez szintén költséges szoftverekre lenne szükség, például OLAP-szoftverekre, de egy ügyes programozóval valószínűleg a könyvtárban is tudnának írni elemző algoritmusokat, az elemzéseket vizuálisan megjelenítő programokat. Ehhez arra van szükség, hogy a könyvtár vezetősége jól behatárolja, hogy mire szeretné használni az adatbányászatot, használható elemzési és adatfigyelési, adatgyűjtési szempontokat adjon az elemzőknek és a programozóknak. Szükség van továbbá az eredményes alkalmazáshoz a könyvtárat, a kutatott szakterületet jól ismerő elemző könyvtárosra, aki képes „olvasni" az adatbányászat során kapott mintázatokban, átlátja az összefüggéseket, és felhasználható formában prezentálja a vezetőségnek vagy a döntésért felelős kollégáinak. Így viszonylag nagy felelősség hárul az elemzőre, hogy milyen munkát végez, mit tár fel az elemzett adatokban, lehet hátrányt, nem megfelelő eredményt hozó szereplő is, illetve ha jól végzi munkáját, lehet előnyt hozó résztvevő a tudásfeltárás folyamatában. Így az adatbányászat alkalmazásának lehetnek technikai és személyi hátrányai, de első megközelítésben mindenképp a technikaiak az erősebbek.

Tóth Erzsébet a fentebb említett tanulmányában is első helyen említi a hátrányok között az adattárolási és lekérdezési szabványok hiányát és az adatkonverzió sokszor nehézkes megoldását.[129] Hátrányként írja ezen kívül, hogy könyvtári területen „jelenleg még nem tesztelték az adatbányászati technikák sikeres alkalmazását".[130] Ezt a jelenlegi helyzet szerint a nemzetközi viszonylatokat nézve meg lehet cáfolni a sokoldalú elemzést nyújtó Penn Data Farm és a fentebb példaként említett a taiwani Kun Shan műszaki egyetem könyvtárának DMBA elnevezésű adatbányászati eszköze (lásd a 3.1.1.1. A könyvtári adatbányászat költségvetési alkalmazásai című alfejezetben) használatának ismeretében. Hazai viszonylatban elképzelhető, hogy nem tesztelték az alkalmazását ilyen komplex rendszerekkel vagy tudatosan nem alkalmazzák még az eljárást, a jövőbeli kutatás célja lehet feltárni, hogy mennyire ismerik a hazai könyvtárakban ezt az eljárást, illetve ha nem is ismerik, mennyire hasznosítják az adatbányászat elemeit a mindennapi munkában valamint, hogy mennyire nyitottak az adatbányászat alkalmazására. Tóth Erzsébet szintén említi hátrányként a technikai akadályokat, csak más értelemben, mint itt használtam, ő azt emeli ki, hogy az okozhat

[129] Uo.
[130] Uo.

gondot, „ha az adatok jelentése nincs pontosan meghatározva az adatbányászati eszközök számára, akkor azok nem képesek az információk közötti relációk felismerésére."[131] Ez valóban probléma lehet, de meglátásom szerint fejlett adattárházzal, adatbányászati eszközökkel, megfelelő programozókkal ez a probléma kiküszöbölhető.

Végeredményben összegezve az előnyöket és a hátrányokat az adatbányászat felhasználásakor szinte minden azon múlik, hogy mennyire szakértő, célirányos módon használják fel, mennyire vannak tisztában azzal, hogy mire szeretnék használni és ezzel párhuzamosan technikailag mire képesek felhasználni a meglévő adatbányászati eszközeiket. Ha ez a két tényező összhangban van, akkor szinte biztos az adatbányászati eszközök sikeres alkalmazása akár egyszerűbb, akár bonyolultabb szoftvert alkalmazó eljárásokról legyen szó.

4.3. A könyvtári adatbányászat alkalmazásának egyéb kérdései

Nicholson az alábbiakban foglalta össze a könyvtári adatbányászathoz kapcsolódó kérdéseket:

- a személyiségi jogok védelme,
- az adatbányászathoz használt adattárházak összekapcsolásához hiányzó szabványok,
- mi a helyzet azokkal a könyvtárakkal, ahol a használati adatokat nem számítógépen tárolják (vagyis nem használnak integrált könyvtári rendszert – a dolgozat írójának megjegyzése), mivel ilyenkor szinte lehetetlen az adatbányászat használata,
- problémaként felmerülhet, hogy egy könyvtárban sok használati adat felmérés vagy interjú eredménye és ezek az adatok nincsenek benne az adattárházban, erre Nicholson azt a megoldást látja, hogy ezek az eredmények kiegészítő adatok lehetnek, amelyek erősítik az adatbányászat során kapott bizonyítékokat.[132]

A következőkben vegyük sorra ezeket a kapcsolódó kérdéseket, mivel valóban létező és fontos problémákat vetnek fel az adatbányászat könyvtári alkalmazásáról.

[131] Uo.
[132] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship. [elektr.dok.].

4.3.1. A személyiségi jogok védelme

A személyiségi jogok védelméhez hozzá lehet venni az etikai kérdéseket, főleg, ha a könyvtár például olyan adatbányászatra alapuló szolgáltatást készül bevezetni, amelyben a felhasználóknak külön profilt hoznának létre az olvasói kártya számával való belépés után. Ez azzal járna, ahogy Nicholson is helyesen emeli ki, hogy a könyvtáraknak tájékoztatniuk kellene a felhasználókat, hogy mire használják fel az adataikat, sőt, kívánatosabb lenne, ha engedélyt is kérnének tőlük e-mailben, telefonon vagy személyesen, ezzel biztosítva őket, hogy csak azok adatait elemzik, akik hozzájárultak.[133] Az ilyen szintű elővigyázatosság szükségességét meg lehet fontolni azokban az esetekben, amikor nem dolgoznak a felhasználók személyes adataival, hanem még az adatbányászati eljárás elején törlik azokat, ekkor nem feltétlenül kell engedélyt kérni.

4.3.2. Hiányzó szabványok az adatok megosztásához és konvertálásához

Az adattárházak összekapcsolásához hiányzó szabványok kérdéséhez kapcsolódik, és fontosabb probléma az adattárházba kerülő adatok formátumának egységessége, konvertálhatósága, amelynek megoldása főleg az integrált könyvtári szoftver gyártójának együttműködésén, a szükséges feltételek biztosításán múlik. Itt az az alapprobléma, hogy az integrált rendszer adatbázisában nincs szabványosan meghatározva minden adatelem jelentése, egy részük meg van, amelyek a MARC-formátumhoz kapcsolódó adatelemek, de a használati adatokhoz, Nicholson kifejezésével a „*transaction-level*”[134] adatokhoz nagy valószínűséggel nincs kialakult szabvány és gyártónként változik, hogy az adatbázisokban ezeket hogy írják le. Az elektronikus források használatának méréséhez már létezik egy COUNTER nevű projekt, amely szabványokat határoz meg a méréshez, és ezen szabványok szerint hoz létre online statisztikákat, amelyeket fel lehet használni az adatbányászathoz. A projekt tagjai nagy részben kiadók, közvetítő szereplők, könyvtárak, könyvtári konzorciumok és fiókkönyvtárak.[135] A projekt léte egy előremutató jel az adatbányászat alkalmazásának megkönnyítéséhez, az elektronikus források használatának leírása szabványokkal könnyebb feladat, mint egy integrált könyvtári rendszerben szabványokat létrehozni, így lehet, hogy ez utóbbi még várat magára.

---

[133] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[134] NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship. [elektr.dok.].
[135] COUNTER – Online usage of electronic resources, http://www.projectcounter.org/ [2009-08-26].

4.3.3. Az integrált könyvtári rendszerek hiánya a könyvtárakban

Az utolsóként említett kérdés hazánkban nagyon is élő probléma, mivel sok kisebb település könyvtárában még nem használnak integrált könyvtári rendszert, így azokon a helyeken valóban szinte lehetetlen a használati adatok elemzése, mivel csak manuálisan vannak dokumentálva. Az ilyen könyvtárakban, ha az integrált könyvtári rendszer használati adatainak vizsgálatára nincs is lehetőségük, arra például van, hogy szabadon hozzáférhető internetes szolgáltatásokat vegyenek igénybe a tájékoztatáshoz, ahol lehet naplózni a beszélgetéseket, kérdéseket, vagy akár a tájékoztató könyvtáros levelezésének kérdéseit és válaszait is naplózhatják későbbi elemzés céljára. Vagy azt a módszert is alkalmazhatják, hogy a helyben használatot dokumentálják és elemzik. Ilyen helyzetben sok múlik a könyvtárosok találékonyságán és problémamegoldó képességén, hogy a megfelelő technikai eszközök hiányában milyen más ügyes módszereket használnak fel.

4.3.4. A digitális könyvtárak és az adatbányászat

Kapcsolódó kérdésként ezeken kívül kiemelhetjük a digitális könyvtárak és az adatbányászat valamint a hagyományos könyvtárak és az adatbányászat kapcsolatát, amely két típust az alkalmazás leírásakor nem igazán különböztettem meg, határoztam meg a különbségeket az adatbányászat alkalmazásával összefüggésben.

A bemutatott adatbányászati eljárások főként a hagyományos könyvtárakra vonatkoznak, de azt is számításba kell venni, hogy a könyvtárak egyre nagyobb arányban jelenítik meg a dokumentumaikat digitális formátumban is[136], mint az OSZK DK — az OSZK Digitális Könyvtár[137] vagy eleve digitális könyvtárként működnek, mint hazánkban a Magyar Elektronikus Könyvtár. Ahogy Nicholson is kiemeli, míg a két típusú könyvtárban az alkalmazottaknak mások a feladatai, a tevékenységek ugyanazok maradnak, így az eszközök azonosítása, amelyek a gyűjteményhez szükségesek, a szolgáltatott dokumentumok beszerzése, és azok szolgáltatása a felhasználóknak, valamint segíteni a felhasználókat abban, hogy elérje ezeket a dokumentumokat, elektronikus vagy személyes tájékoztatás segítségével.[138] A digitális könyvtárakat nézve a legtöbbnek nincs kölcsönzési funkciója, a felhasználó egyszerűen letöltheti vagy kinyomtathatja a kívánt anyagot, de a hagyományos könyvtárral szemben

[136] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].
[137] OSZK DK: http://oszkdk.oszk.hu/ [2009-08-26].
[138] NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining. [elektr. dok.].

a digitális könyvtár képes nyomon követni a felhasználó egész látogatását a digitális könyvtárba[139], pontosabban a digitális könyvtár honlapjára. A MEK számokban elnevezésű statisztika vizsgálatakor (a harmadik fejezetben) láthattuk, hogy milyen hasznosnak bizonyulhatnak ezek a nyomon követett adatok adatbányászati szempontból. Digitális könyvtárat vizsgálva az adatbányászattal többnyire szorosabban a gyűjtemény használatáról kaphatunk információkat, a hagyományos könyvtárhoz képest kevésbé ismerhetjük meg a felhasználók egyéb interakcióit a könyvtárral, hacsaknem tartanak fenn egy jól működő online tájékoztató szolgáltatást is, amelynek a naplózásait/ tranzakcióit elemezhetik.

[139] Uo.

5. AZ ADATBÁNYÁSZAT HAZAI ALKALMAZÁSA – LEHETŐSÉGEK ÉS IGÉNYEK

Ahogyan Tóth Erzsébet is kiemeli a tanulmányában[140], már 2002-ben, teljes mértékben rendelkezésünkre álltak a technológiai feltételek az adatbányászathoz, de az alkalmazása „még gyerekcipőben jár". Ahogyan láttuk, ma külföldön már kevésbé járnak gyerekcipőben, sok helyen felismerték az adatbányászat könyvtári alkalmazásának lehetőségét, előnyeit, és ki is használják azokat (Lásd a Penn Library Data Farm 3.5 és a Fairfax County Library 3.2.4 példáját). Az alkalmazás előrelendítő tényezője, ha az adatbányászat felhasználásának lehetőségét, szükségességét a könyvtár vezetője vagy menedzsmenttel foglalkozó szakembere ismeri fel, olyan ember, aki döntéshelyzetben van a könyvtár irányításában, ekkor neki megvan a lehetősége, hogy véghez vigye a felhasználás ötletét. Ha nem a vezetőségből születik az alkalmazás ötlete, akkor leginkább úgy lehet meggyőzni az előnyeiről a döntéshozókat, hogy amennyire az ingyenes és viszonylag kevés időráfordítást igénylő lehetőségek engedik, a vállalkozó szellemű könyvtáros próbaképp alkalmazza munkájában az adatbányászatot, és az eredményeit bizonyítékképpen ismerteti a vezetőségnek vagy ha nagyon sikeres volt az alkalmazás, akkor a vezetőség magától észreveszi, hogy valamilyen változás történt. Ilyen alkalmazásra kínálnak lehetőséget az ún. web 2.0-ás alkalmazások, például az online tájékoztatás naplózása (lásd 3.1. pont utolsó előtti bekezdése); vagy a gyakran kölcsönzött könyvek listájának figyelemmel követése (lásd FSZEK példája a bevezetőben).

Hazánkban könyvtári területen kísérletképpen már alkalmazták az adatbányászatot. (Jezerniczky Ani, aki az adatbányászatból írta a szakdolgozatát[141] a Weka ingyenes adatbányász program bemutatásáról – mivel egy könyvtárban dolgozott, az utolsó fejezetben példaképp, a TextLib integrált könyvtári rendszer adatait felhasználva néhány szempontból bemutatja a lehetséges elemzéseket.) Levélváltásunk alapján az ő tapasztalatait összegezve ismertetem a hazai alkalmazást tesztelő 'adatbányász' szemével.

A hazai alkalmazásban felmerült fő kérdések:

- a felhasználás szükségessége,

[140] TÓTH Erzsébet: Adatbányászatra irányulót törekvések könyvtári területen. [elektr. dok.].
[141] Jezerniczky Istvánné: Az adatbányászat egy könyvtári alkalmazása. [Szakdolgozat]. Dunaújvárosi Főiskola Mérnök Informatikus BSc Hálózati szakirány. 2009.

- az adatbányászat lehetséges felhasználási módjai,
- könyvtárosi igény az alkalmazásra,
- a könyvtárosok hozzáállása,
- technikai kérdések, a személyiségi jogok kérdése.

5.1. A felhasználás szükségessége

A felhasználás szükségességét támasztja alá, hogy gyakran a könyvtárosok máshogy emlékeznek a felhasználók szokásaira, a könyvtárba járókra, például arra, hogy melyik korosztály látogatja a könyvtárat, vagy hogy melyik könyveket kölcsönzik gyakrabban, Ani bevallása szerint a TextLib rendszerből kinyert nyers adatok is meglepőek voltak. Az adatbányászattal a valós adatokra tudnának építeni, ahogyan a bizonyítékokon alapuló könyvtári munka és adatbányászat alfejezetben (3.1) bemutatásra került.

5.2. Az adatbányászat lehetséges felhasználási módjai

A könyvtári adatbányászat kipróbálója szerint mindenképp hasznos lenne, ha legalább egy-két évente vizsgálnák a kölcsönzői kört, a lemorzsolódást és az állománygyarapítást, mivel ezen vizsgálat eredményeinek tudatában jobban meg lehetne célozni a lemorzsolódás-gyanús olvasókat, jobban fel lehetne mérni az olvasói igényeket és valós adatokra építve igazodni lehetne hozzájuk. Szerinte ilyen jellegű elemzés elvégzésével és kiértékelésével hatékonyabb munkát érhetnének el, valamint új olvasókat is lehetne hozni a könyvtárba. Akár azt is működőképesnek látná, hogy az éves statisztikát a papíradatlapra írt helyett számítógépen készítsék el, egy adatbázis felhasználásával, és adatbányászati eszközökkel megvizsgálnák az éves munka hatékonyságát.

Összefoglalóan a következő területeken tartaná hasznosnak az adatbányászat alkalmazását, ezek sok helyen összecsengnek a fentebb bemutatott alkalmazási lehetőségekkel, más helyeken pedig a tapasztalatok alapján újabb szempontokkal bővülnek:

1. A beiratkozott olvasók kölcsönzési szokásainak elemzése — milyen gyakran látogatják a könyvtárat, hány könyvet visznek ki egyszerre, az év melyik időszakában aktívabbak; az élethosszig tartó tanulás *(lifelong learning)* hogyan befolyásolja a könyvtár használatát; a diákok meddig maradnak felhasználók: az általános iskolától a felsőfokú intézményig vagy előbb lemorzsolódnak-e, mi a lemorzsolódás oka.

2. Kik azok, akik megszűnnek használni a könyvtárat, melyik korcsoport a legveszélyeztetettebb, hogyan lehetne megcélozni őket a megtartáshoz.

3. A kölcsönzési adatok alapján milyen könyvekre van igény, a könyvtárközi kölcsönzéseket is figyelembe véve.

A felhasználói szokások vizsgálatához láttunk példákat a 3.4.2. alfejezetben, itt Ani új szempontokkal bővíti a megfigyelési pontokat, mint az élethosszig tartó tanulás befolyásoló ereje, és annak megfigyelése, hogy egyes korcsoportok meddig maradnak beiratkozott olvasók. Látszik, hogy nagy hangsúlyt fektet az alkalmazás lehetőségeinél a lemorzsolódásra, az erre való figyelés ötlete Mikulás Gábortól származik, valószínűleg tapasztalati példákra, üzleti alkalmazási mintákra építve. A kölcsönzési adatok vizsgálatára szintén láttunk alkalmazási ötleteket és példákat a 3.3. alfejezetben.

5.3. A könyvtárosok hozzáállása és igény az adatbányászatra

A könyvtárosok hozzáállására – a tapasztalatok szerint – jellemző, hogy lassan, de fejlődik a számítógépes eszközökre való nyitottságuk (főként a gépi katalogizálás, az elektronikus adatbázisok egyre nagyobb szükségessége és felhasználása miatt), de kevésbé tartják hasznosnak az önmonitorozásról, szervezet értékeléséről szóló módszereket (mint a könyvtári menedzsment és a minőségbiztosítás), amelyek első látásra lehet, hogy plusz munkának látszanak, de felelősségteljesen végezve a szervezet hasznára is válhatnak. Ani azt is hozzátette, hogy főként az idősebb generáció ragaszkodik a régi könyvtárosi magatartáshoz; ezen valószínűleg a képzés, új ismeretek megtanítása, tisztázása valamelyest segíthet. A megkérdezett könyvtári adatbányászat-tesztelő ennek fényében úgy látja, hogy előrevivő lenne, ha lenne igény az adatbányászatra, de tapasztalatai szerint sajnos nincs.

5.4. Technikai és személyiségi jogi kérdések

A technikai feltételekről sajnos beigazolódott, hogy valóban nem könnyű az integrált könyvtári rendszerekből adatokat kinyerni egy külső program számára (amelyről a 2.2. pontban is szó volt), de pozitívum, hogy a példa szerint nem lehetetlen feladat. Aninak nagy nehézségek árán és a TextLib fő programozója (Thék György) segítségével több mint egy hónap alatt sikerült kinyernie az adatokat a Weka adatbányász program számára. Elmondása szerint a legnagyobb nehézséget az adattisztítás jelentette, mert sokszor találkozott hiányos adatmezőkkel és mivel bonyolult adatbázisokat kellett létrehoznia. Összefoglalva szerinte megvalósítható az adatok exportálása, csak a programozóknak nagyon sok mindennek kell megfelelniük, alkalmazkodniuk kell a könyvtárosok igényéhez. A személyiségi jogok kérdése valóban kényes terület, ahogyan a 4.3.1. pontban ki is tértünk rá, könyvtárak részéről ez az ellenállás egyik legnagyobb

forrása, miszerint félnek a személyes adatok illetéktelen kezekbe kerülésétől. Anit emiatt utasította el az egyik megyei könyvtár, amikor az adatbázisukat szerette volna használni mintaként, hogy nagyobb felhasználószámú adatbázison tudja tesztelni az adatbányászatot. Egyrészt érthető a félelmük a személyes adatok védelméről szóló törvény(ek) (pl. 1992. évi LXIII. törvény) miatt, másrészt alaptalan, ha az adatbányászati eljárás elején a személyes adatok valóban törlésre kerülnek vagy kódszámmal helyettesítik azokat az adatok tisztításakor (az eljárás menetét lásd a 3.1. pontban).

Összefoglalóan Ani mindenképp hasznosnak tartja az adatbányászat alkalmazását könyvtári területen, mert ezzel sokkal jobban meg lehetne célozni az olvasói igényeket és növelni lehetne a könyvtárak népszerűségét. Pozitív példa, hogy hazánkban is megvan a lehetőség az adatbányászat alkalmazására, könyvtár részéről megfelelő szakemberek és kellő nyitottság birtokában.

6. BEFEJEZÉS

Összességében úgy tűnik, hogy az adatbányászat valóban eredményesen alkalmazható, létjogosultsággal rendelkezik a könyvtári területen is. A feldolgozott szakirodalom és a külföldi alkalmazások alapján legjobban a könyvtári menedzsment és döntéshozás majd a tájékoztatás és a gyűjteményszervezés területén lehet felhasználni. A felhasználói kapcsolattartással, felhasználói szokások és magatartás elemzésével kapcsolatos alkalmazások a személyiségi jogok védelme miatt a legkényesebbek, a legnagyobb körültekintést igénylők, így emiatt kezdetben nehézségekbe ütközhet az eljárás alkalmazása ezeken a területeken. Az adatbányászat alaklmazására kétféle út áll a könyvtárak előtt: vagy olyan integrált rendszert választanak (vagy eleve olyannal rendelkeznek), amelyik képes adatbányászati elemzések elkészítésére (mint a SirsiDynix által forgalmazott új Horizon) vagy megtalálva a módját, az integrált rendszerükből kinyerik a használni kívánt adatokat áttöltik egy adattárházba, és külön adatbányászó programokkal (ingyenesen használható például a WEKA nevezetű) készítik el az elemzéseket. Mindkét megoldás bizonyos költség-ráfordítást igényel a könyvtárak részéről, ha ingyenes adatbányász-programot használ az adott könyvtár, akkor körülbelül ugyanakkora ráfordítást, ha fizetős programot, akkor a második opció nagyobbat igényel. A magyarországi alkalmazást tekintve a könyvtári integrált rendszerek fejlődését feltételezve az első lehetőségre látunk nagyobb esélyt. Az adattárházzal és adatbányász-programmal kombinált második lehetőség viszont szabadabban használható annyiban, hogy nincs az integrált könyvtári rendszerhez kötve, ha az adott könyvtár rendelkezik tehetséges és innovatív könyvtárosokkal és programozókkal (vagy ideális esetben könyvtáros programozókkal), akkor ennek az utóbbinak a használata a kézenfekvőbb a könyvtárnak. Így a második lehetőség több külső, anyagi, személyzeti feltételhez kötött, de szélesebb is a felhasználási lehetőség területe benne. A könyvtári adatbányászat alkalmazásában nem elhanyagolható tényező, hogy szükség van hozzá a könyvtári vezetők, könyvtárosok szemléletváltására, arra, hogy felismerjék, hogy a meglévő könyvtárhasználati adataikból stratégiailag előnyt hozó, gazdaságosabb működést lehetővé tevő hasznos információk nyerhetők ki. Ezután következhet, hogy azon gondolkozzanak, milyen megoldással tudnák használni az adatbányászat eljárását. A hazai könyvtárak egyre növekvő nyitottságát látva a számítógépes technológiai megoldásokra, nem tartom elképzelhetetlennek, hogy hamarosan nagyobb érdeklődés mutatkozik a könyvtári adatbányászat iránt.

BIBLIOGRÁFIA

Felhasznált irodalom:

[NICHOLSON, Scott]: Career. http://scottnicholson.com/career/index.html , [2009. 01. 24.].

ADRIAANS-DOLF ZANTINGE, Pieter: Adatbányászat. Budapest, Panem, 2002.

BANERJEE, Kyle: Is data mining right for your library? = Computers in Libraries, 18. köt. 10. sz. 1998. pp. 28-31.

CULLEN, Kevin: Delving into data [elektr. dok.] = Library Journal. 130. köt, 13. sz. 2005. EBSCO azonosító: 03630277 [2009-08-07].

ESTABROOK, Leigh S.: Sacred trust or competitive opportunity: Using patron records. [elektr. dok.] = Library Journal, 2. sz. 1996. 48-49, EBSCO azonosító: 03630277.

Evidence Based Librarianship – Bizonyítékokon (precedenseken) alapuló könyvtári munka. [elektr. dok.] = KIT Hírlevél 5. sz. 2003. http://www.kithirlevel.hu/index.php?oldal=cikk&c=746 [2009-02-17].

JEZERNICZKY Istvánné: Az adatbányászat egy könyvtári alkalmazása. [Szakdolgozat] Dunaújvárosi Főiskola Mérnök Informatikus BSc Hálózati szakirány. 2009.

KUN-PÁL Gábor: 57. fejezet: Döntéshozatal több kritérium felhasználásával, in NCGIA CC [elektr. dok.] http://gisfigyelo.geocentrum.hu/ncgia/ncgia_57.html [2009-07-31].

LANKES, R. David – NICHOLSON, Scott: Archiving Human Intermediation: The Digital Reference Electronic Warehouse (DREW) Project. http://quartz.syr.edu/rdlankes/Presentations/2004/drewasist.pdf [2009-08-07].

LANKES, R. David – NICHOLSON, Scott: Reference Extract: Extending the Reach of Digital Reference through Collaborative Data Warehousing. http://www.ieee-tcdl.org/Bulletin/v2n1/nicholson/nicholson.html [2009-08-11].

MIKULÁS Gábor: Adatbányászat a könyvtárakban. referátum.[Elektr. dok.]. http://www.gmconsulting.hu/inf/cikkek/207/index.php [2009-01-24].

NICHOLSON, Scott: Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship [elektr. dok.]. http://bibliomining.com/nicholson/approach.htm [2009. 09. 07.].

NICHOLSON, Scott: The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services [elektr. dok.] http://www.bibliomining.com/nicholson/nicholsonbibliointro.html [2009-08-11].

NICHOLSON, Scott: Bibliomining Applications in Digital Reference: Using Data Warehousing and Data Mining to Improve Management and Decision-Making. [Elektr. dok.]. http://www.webjunction.org/white-papers/-/articles/content/439065?_OCLC_ARTICLES_getContentFromWJ=true [2009-08-07].

NICHOLSON, Scott–STANTON, Jeffrey: Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. [elektr. dok.]. http://bibliomining.com/nicholson/odmcom.html [2009-09-07].

NICHOLSON, Scott: The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services, http://www.bibliomining.com/nicholson/nicholsonbibliointro.html [2009-08-11].

ROMERO, Jorje Candás: Minería de datos en bibliotecas: bibliominería [elektr. dok.] http://www2.ub.edu/bid/consulta_articulos.php?fichero=17canda2.htm [2009-01-24].

TÓTH Erzsébet: Adatbányászatra irányuló törekvések könyvtári területen. [elektr. dok.] = Könyvtári Figyelő 48. köt. 3. sz. 2002. http://www.ki.oszk.hu/kf/kfarchiv/2002/3/toth.html [2009. 09. 07.].

WU, C.-H.: Data mining applied to material acquisition budget allocation for libraries: design and development = Expert Systems with Applications, 25. sz. 2003. pp. 401-411.

Internetes források

About the Penn Library Data farm: http://datafarm.library.upenn.edu/census/index.html [2009-08-26].

[A Cognos program leírása a Wikipédiában], http://en.wikipedia.org/wiki/Cognos [2009-08-08].

[A DREW projekt 2005-ös weboldala az Internet Archive Way Back Machine web-archiváló keresővel előhívva], http://web.archive.org/web/20050723011710/http://drew.syr.edu/ [2009-08-11].

COUNTER – Online usage of electronic resources, http://www.projectcounter.org/ [2009-08-26].

ELTE Egyetemi Könyvtár, http://egyetemi.klog.hu/ [2009-02-18].

[A Fairfax County Public Library blogja], http://allfairfaxreads.blogspot.com/ [2009-01-24].

[FAQ a Las Vegas-Clark County Library oldalán], http://www.lvccld.org/faqs/index.cfm [2009-08-11].

Gyakran kölcsönzött dokumentumok a Fővárosi Szabó Ervin Könyvtárban, http://www.fszek.hu/?article_hid=23981 [2009-02-16].

Könyvtári Intézet Könyvtártudományi Szakkönyvtár: Elégedettségi kérdőív az olvasói igények felmérésére 2007. 04. 16. – 2007 06. 30.

[A SirsiDynix hivatalos oldala], http://www.sirsidynix.com/Solutions/Products/integratedsystems.php [2009. 01. 25.].

[A SirsiDynix termékeinek ismertetője], http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/Symphony_Features_Benefits.pdf [2009-01-25].

[A SirsiDynix rendszereinek alkalmazása], http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/NormativeDataProject.pdf [2009-01-25].

[SirsiDynix szolgáltatásainak leírása], http://www.sirsidynix.com/Solutions/Products/analytical.php [2009-01-25].

A MEK számokban, http://mek.oszk.hu/html/statisztika.html [2009-08-10].

[MEK]: Megoszlás, http://mek.oszk.hu/html/megoszlas.html [2009-08-10].

[MEK]: Letöltési statisztika, http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=stat [2009-08-10].

[MEK]: Összesített dokumentum-letöltési adatok, http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=stat_rcum , [2009-08-10].

[MEK]: [Sikerlista], http://mek.oszk.hu/html/vgi/kereses/kereses.phtml?tip=siker [2009-09-08].

[MEK]: Webstatisztika mek.oszk.hu, http://mek.oszk.hu/webalizer/ [2009-09-08].

Webaudit, http://webaudit.hu/object.8D7ED62D-1CED-4E84-BAD3-4944E7412406.ivy?session.folder=10136 [2009-08-11].

[A MEK honlapjának forgalma], http://extremetracking.com/open?login=mek2adm [2009-08-11].

Országos Széchényi Könyvtár, http://www.oszk.hu/ [2009-02-18].

OSZK DK: http://oszkdk.oszk.hu/ [2009-08-26].

Penn Library Data Farm, http://datafarm.library.upenn.edu/ [2009-01-26].

[Penn Library Data Farm]: USR Report Builder: http://datafarm.library.upenn.edu/cgi-bin/ursform/ursform.pl [2009-08-27].

[Penn Library Data Farm]: E-resources search: http://www.library.upenn.edu/cgi-bin/res/sr.cgi [2009-08-30].

[Penn Library Data Farm]: Data Farm – Visitor Reports: http://datafarm.library.upenn.edu/visit_gatecounts.html [2009-08-30].

[Penn Library Data Farm]: Voyager Fund: http://proxy.library.upenn.edu:4500/loggedin/df-vfunds.html [2009-08-30].

[Penn Library Data Farm]: Borrow Direct Data Repository: http://datafarm.library.upenn.edu/bdrpt.html [2009-08-30].

[Penn Library Data Farm]: EZBorrow Data Repository: http://datafarm.library.upenn.edu/bdrpt_ez.html [2009-08-30].

[Penn Library Data Farm]: Collection Management Report Builder: http://datafarm.library.upenn.edu/collrpt.html [2009-08-30].

[Penn Library Data Farm]: DataFarm – Inventory Reports: http://datafarm.library.upenn.edu/inventory2.html [2009-08-30].

[Penn Library Data Farm]: Voyager Reports – Page Data Farm: http://datafarm.library.upenn.edu/circptrdbrd.html [2009-08-30].

Planning Reference Extract, http://referencextract.org/ [2009-08-11].