

súlyt nyer az elektronikus archiválás megoldatlansága.

Az *Electronic-Archiving Initiative* – amelyet a JSTOR 2003-ban indított útjára – felmérést kezdeményezett 11 kis, közepes és nagy könyvtárban azzal a céllal, hogy az elektronikus forrásokra történő váltás milyen hatással van a könyvtári költségekre, különösen az előfizetési díjak mellett jelentkező kiadásokra (nevezzük ezeket kezelési költségeknek). A kutatás eredményeiről részletesebben a *D-LIB Magazine* 2003. 1. száma ad hírt, a teljes jelentés 2004-ben jelenik meg *Schonfeld, R. C.–King, D. W.–Okerson, A.–Fenton, E. G.* tollából *The nonsubscription side of periodicals: changes in library operations and costs between print and electronic formats* címen.

A vizsgált 11 felsőoktatási könyvtárban 100-tól 23 000-ig terjed az előfizetett nyomtatott folyóiratok száma; az elektronikusoké 5000-tól 20 000-ig. Az előfizetési költségek alsó értéke 250 000 dollár, felső határa 2,5 millió. Ezen összegen belül elektronikus folyóiratokra 15–80%-ot költenek.

A vizsgálat során az ún. életciklusmódszert alkalmazták: a hosszú távra szóló pénzügyi kötelezettségvállalást számították ki egy adott folyóiratcímre vonatkozóan egy adott év beszerzése alapján. E műveletet külön végezték el a nyomtatott és az elektronikus folyóiratokra nézve, így képesek voltak összehasonlítani a különböző formátumok kezelési költségeit is.

Megállapították, hogy az elektronikus formátummal kapcsolatos hosszú távú pénzügyi kötelezettség kevesebb, különösen a kisebb intézményekben. Bár különféle helyi tényezőket is figyelembe kell

venni, számos könyvtár számára valószínűleg csökkennek a kezelési költségek az elektronikus formátumra való áttérés következtében. Mivel azonban – különösen a kis és közepes könyvtárakban – feltehetően növekedni fog a gyűjtemény nagysága, a kezelési költségekben mutatkozó nyereséget felemészti a folyóiratcímek megnövekedett száma.

Habár a kezelési költségek nem csökkennek annyira, amennyire egyes optimista előrejelzések jósolták, a trend mégis az, hogy a kezelésre fordított kiadások csökkennek, ahogy a könyvtárak elektronikus folyóiratokra váltanak át. Mindenesetre számolni kell azzal, hogy bizonyos költségek, amelyeket a könyvtárak, illetve a kiadók viseltek, helyet cserélhetnek a két fél között (pl. a szerveren való tárolás).

Az archiválás megoldatlansága határozottan hátráltatja az elektronikus folyóiratokra való áttérést. Remélhető, hogy az a megtakarítás, amely a tanulmány szerint a kezelési költségekben mutatkozik, hozzá fog járulni az archiválás megvalósításához. Válaszra várnak azonban a következő kérdések: Elegendő lesz-e a várható megtakarítás általánosan elfogadható archiválási eljárások kidolgozására, vagy más forrásokból kell fedezetet találni erre? Késlelteti-e az archiválás megoldatlansága az áttérést az elektronikus formátumra? Vagy felgyorsul az áttérés, és az archiválás problémájától eltekintenek?

/SCHONFELD, Roger C.–FENTON, Eileen Gifford: *Library periodicals expenses: comparison of nonsubscription costs of print and electronic formats on a life-cycle basis.* = CLIR issues, 39. sz. 2004. május–június, p. 1–4./

(Papp István)

## Adatbányászat a könyvtárban

A *bibliomining* vagy könyvtárak számára végzett adatbányászat a könyvtári szolgáltatásból adatokat előállító adatbányászati és bibliometriai alkalmazás. (A bibliomining kifejezést a szerzők, S. Nicholson és J. Santon egy korábbi publikációjukban kívánták bevezetni az adatbányászat könyvtári alkalmazására. A cikkben a továbbiakban *könyvtári adatbányászatot* fogunk említeni.) A könyvtári adatbányászat a könyvtári folyamatokkal kapcsolatos adathalmaz feltárásának statisztikai és mintázatfelismerési eszköze, melynek célja a dön-

tés-előkészítés, illetve a szolgáltatások igazolása. (Az adatbányászat az ipari és kereskedelmi szektorban az utóbbi tíz évben bevetté vált tevékenység; a szervezet jelentős folyamatait elektronikus eszközökkel lefedő rendszerekben tárolt adatokat, illetve a közöttük lévő összefüggéseket tárja fel. Ennek célja lehet a tervezés, döntés-előkészítés, értékelés, kontrolling, minőségirányítás stb. Az ilyen rendszerek lehetővé teszik, hogy a vállalkozások azonosítsák a szolgáltatásukat igénybe venni nem kívánók – a lemorzsolódók – csoportját.

Például telefonszolgáltató esetében visszakeresik azokat az ügyfeleket, akiknek forgalma az utóbbi három hónapban 50%-os csökkenést mutatott. De azt is megállapíthatják, hogy mely szegmensek vennék szívesen, ha értéknövelt szolgáltatáshoz jutnának stb. E módszer által marketingköltséget takarítanak meg, miközben növelik az ügyfelek elégedettségét. A könyvtárak esetében az integrált könyvtári rendszerek által gyűjtött működési adatok jelenthetik a könyvtári adatbányászat alapját. – A ref.)

### **A könyvtári adatbányászat folyamatának lépései**

**1. A kutatás fókuszának meghatározása.** Ez lehet általánosabb vagy konkrétabb probléma, döntés előkészítése. Például költségcsökkentést megelőzendő egyre fontosabb igazolni és megvédeni az egyes könyvtári szolgáltatásokat. Ha nincs jól azonosított cél, a kutatás során túl sok mintázatot találunk, ami miatt bizonytalanná válhat a rendszer. (Mintázat lehet például az a korreláció, amelyben az 50-70 év közötti életkor, az alsóvárosi lakhely, illetve az évi 25-50-es könyvkölcsönzési szám „áll együtt”).

**2. A külső és belső adatforrások meghatározása.** Az adat legyen működésre utaló, nem aggregált és csekély feldolgozottságú. (Ezek kinyerése érdekében nem szabad letörölni idő előtt az adatokat adatvédelmi okokra hivatkozva.) Külső, könyvtári rendszeren kívüli források lehetnek például a demográfiai adatok, belsők az ügyfelek jellemzőit mutató adatok, kölcsönzési adatok, a könyvtári portál látogatottságát mutató logfájlok.

**3. Az adattárházba szánt adatok gyűjtése, tisztítása és anonimizálása.** Az adattárház: az integrált rendszertől független adatbázis, amely tisztított és anonimizált, elemzésre előkészített működési adatokat tartalmaz. (Az adattárház tehát esetünkben nem azonos az integrált könyvtári rendszerben tárolt adatokkal; onnan ki kell menteni őket az adattárházba. – A ref.) A tárolt adatokat ugyanis el kell választani a konkrét személyi adatoktól, megelőzve az illetéktelen felhasználást. Ennek érdekében hasznos egyeztetni jogi szakértővel. Kutatóintézetben működő könyvtárnak hasonló okból engedélyt kell kérnie az illetékes kutatási bizottságtól. Az adattárház szoftverét meg kell írni, és az alrendszer némi fenntartást is kíván. A tanulási folyamat lépcsőzetessége érdekében érdemes ke-

vésbé összetett adatbányászati cél megfogalmazásával és megvalósításával kezdeni.

### **4. A megfelelő elemzési módszer kiválasztása.**

A vezetői információs rendszer (Management Information System = MIS) adatai a működő rendszerből származnak. Ezek az adatok nincsenek megtisztítva. Egyes MIS-rendszereknek van kritikusérték-figyelő és -jelentő rendszere, amely rámutathat például az alacsony használat vagy elhasználódás miatt kivonandó dokumentumokra, vagy a meg nem válaszolt referenzkérdések számának túl magas voltára.

### **5. Minták azonosítása adatbányászattal, jelentéskészítés hagyományos elemzési eszközök által.**

Az online elemző adatfeldolgozás (Online Analytical Processing = OLAP) az adatok interaktív bemutatását kínálja a döntéshozó számára, mely alapján diagramokat, jelentéseket készíthet. (Kijelölhatja a kölcsönzés időbeli változását, az olvasószám napi eloszlását, vagy a különféle napszakokban egy könyvtárosra jutó használók számát stb.)

### **6. Az eredmények elemzése és alkalmazása.**

Az adatbányászat a korábban megszokottaktól eltérő, inspiráló adatelemzési módszer. Célja, hogy nagy mennyiségű adatból statisztikai és mesterséges intelligencia segítségével érvényes, új és felhasználható mintákat azonosítson. A múlt és jelen adatainak megértése a cél. Eszközei például a predikció, az osztályozás vagy a becslés. A kinyert adatok – főként az adatbányászati tevékenység első idejében – tesztelendők. Hasznos, ha azokat könyvtárosok is minősítik, értékelik, hogy ki lehessen küszöbölni az esetleges félreskálázásokat, programozási bakikat. A folyamat utolsó része pedig a jelentéskészítés bevezetése, rendszeressé tétele.

Az adatbányászathoz számos cég kínál szoftvereket. Legismertebbek például a SAS és az SPSS alkalmazásai, ám ezek nagyon költségesek. Létezik nyílt forrású szoftver is, például a Weka, amely olcsóbb ugyan az előbbieknél, ám nem túlságosan felhasználóbarát.

/NICHOLSON, Scott: The bibliomining process: Data warehousing and data mining for library decision making. = Information Technology and Libraries, 2003. december, p. 146–151./

(Mikulás Gábor)