

delemi felügyelet, a Gazdasági Versenyhivatal, az ombudsmanok és a Hírközlési Fogyasztói Jogok Képviselője hivatala rendszerében lévő ürt tölti ki. Fő feladatai közé tartozik:

- szakmai tanácsadás, online tanácsadás;
- információnyújtás közérdekű témákról;
- az érdeksérelmet okozó jelenségben illetékes szolgáltatók felkutatása, és a problémák kezelése;
- fórumrendszerek működtetése a kialakult konfliktusok kezelésére az internetes nyilvánosság felhasználásával, tájékoztatásával.

Az Infomediátor szolgáltatásai iránti igényt az információs társadalom olyan minőségi változásai keltették, amelyeket a hagyományos fogyasztóvédelem nem képes kezelni. Egyre bonyolultabb termékek és szolgáltatások jelennek meg, nő a távolság a felhasználó és a szolgáltató tudásszintje és földrajzi helyzete között. Új termékként jelenik meg az információ, és az információs szolgáltatások világméretű növekedése új veszélyforrásokat jelent a felhasználóknak. Az Infomediátor ellátja azokat a feladatokat is, amelyek a Barátságos Internet Fórum kezdeményezésénél fogalmazódtak meg, így a felhasználói nagyközönség sem marad érdekképviselő nélkül.

Barátságos Internet Fórum

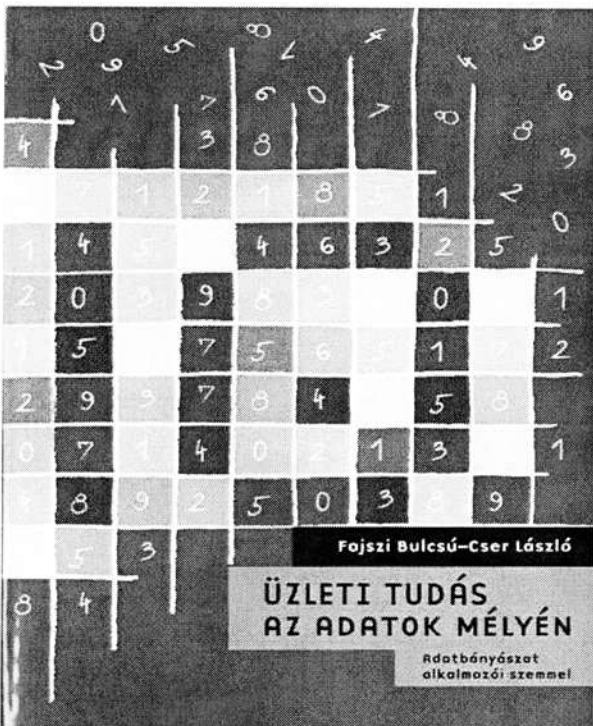
(BIF – www.baratsagosinternet.hu)

Az Inforum, a MATISZ és az MTE 2003 végén szándéknyilatkozatot írt alá a BIF létrehozásáról. A fórum célja az volt, hogy a biztonságosabb internetezés érdekében – a tagországok nemzeti programjaival és akcióival összehangoltan – az EU által követett fő cselekvési irányokat meghonosítsák Magyarországon. A tervek között szerepelt a tiltott tartalmak visszaszorítását támogató ún. forró drót (telefonos tanácsadás – hotline) szolgáltatás működtetése, az önszabályozás, a szakmai összefogás erősítése, a területtel összefüggő felvilágosító, ismeretterjesztő, kommunikációs tevékenység megszervezése, valamint az EU-ban már létező tartalomminősítő és -címkéző rendszerek meghonosítási lehetőségének vizsgálata. A BIF 2004 februárjában azon a napon rendezte konferenciáját, amelyen a *Biztonságosabb Internet* EU-projekt (Safer Internet Exchange project – www.saferinternet.org, <http://www.safer-internet.net>) keretében a világ 15 országában tartottak hasonló szakmai rendezvényeket.

/Inforum, Infomediátor, MTE, MATISZ, AJE honlapok, sajtóközlemények, hírlevelek/

Birkás Bence
(BME OMIKK)

Könyvtárosok is elkezdhetik...



Üzleti tudás az adatok mélyén : adatbányászat alkalmazói szemmel / Fajsi Bulcsú, Cser László. – Budapest : Budapesti Műszaki és Gazdaságtudományi Egyetem, Gazdaság- és Társadalomtudományi Kar Információs és Tudásmenedzsment Tanszék, 2004. – 260 p. : ill ; 21 cm
Bibliográfia a jegyzetekben
ISBN 963-421-558-0

Közhely, hogy a könyvtáraknak már régen időszerű lenne túllépniük a dokumentum- és adatszolgáltatáson, hogy egyre több és különböző irányú értéknövelt szolgáltatást kínáljanak. Egyik út az *adatbányászat*. Az adatbányászat az adatbázisokból olyan implicit és rejtett információkat, szabályszerűségeket, mintázatokat nyer ki, amelyek a gyakorlatban is jól használhatók. A matematikán kívül a statisztikát, a számítástudományon belül a mesterséges intelligenciát, valamint az információs technológia egyre szélesedő eszköztárát használja.

A Budapesti Műszaki és Gazdaságtudományi Egyetem Gazdaság- és Társadalomtudományi

*Karának információs és tudásmenedzsment tan-
széke* által kiadott könyv egyaránt szól szervezeti
vezetőknek és az adatbányászat alkalmazóinak. A
szerzők által is kifejezett cél arra is utal, hogy sike-
res adatbányászati projektekhez mindkét fél egy-
behangzó akaratára szükség van. A könyv alcíme
az alkalmazói szempontot hangsúlyozza, ami által
felhatalmazva érezhetjük magunkat az adatbányá-
szat saját (információs és könyvtári) szempontú
átgondolására.

A könyvtárak alapesetben két nagy adathalmazzal
is rendelkeznek: egyik az általuk kínált dokumen-
tumok adatai, másik a használattal összefüggő
(használói, illetve használati) adatbázis. Az integ-
rált könyvtári szoftverek a két oldalt egy rendszer-
be „integrálják”. Hogy az informatikai alkalmazás
lépcsői közül a könyvtárak melyik fokon állnak, a
következő korszakfelsorolás mutatja: 1. automati-
zálási korszak, 2. folyamatok átszervezése, 3. vál-
lalati (esetünkben: könyvtári) alkalmazások hasz-
nálata, 4. elektronikus kereskedelem, 5. informati-
kai alapú üzleti modellek, 6. csatlakozás elektroni-
kus szövetségi rendszerekhez. Nos, a könyvtárak
jelentős hányada rendelkezik integrált rendszerrel
(3. szint), és még sok fejlesztési lehetőség áll előt-
tük. Annál is inkább, mert a könyvtáraktól a fenn-
tartó elvárja, hogy gazdálkodjon állományával,
munkatársaival, épületével és berendezésével. A
meglévő adatvagyonnal való gazdálkodás azon-
ban még alig elvárás, annak ellenére, hogy igen
értékes forrást jelenthet. E forrást az információs
szektor szédületesen fejlődő vállalkozásai – Ama-
zon.com, Google, Yahoo! stb. – ki is használják
(egyelőre pl. testreszabott reklámok, a keresetthez
hasonló könyvek ajánlása, honlapok testre szabá-
sa, honlapon való navigáció optimalizálása). És
sejthető, hogy e folyamat még sok meglepetést
tartogat.

Könyvtárunkban már ma is gyakran kérdezzük:
mely elkülöníthető használói csoportok látogatják
könyvárunkat (mert honlapon, hírlevélben stb. meg
kell szólítani őket)? Állományfrissítés vagy helyhi-
ány esetén az állomány mely részeit érdemes
leginkább csökkenteni? Mely fiatalok fognak az
iskola végeztével eltűnni a könyvtár látóköréből?

Az integrált könyvtári rendszerek még alig rendel-
keznek – például a fenti kérdésekben is használ-
ható – döntéstámogató modulokkal. A döntéstá-
mogató rendszerek töltik ki a rést a korlátozott
emberi befogadóképesség és a hatalmas adat-
mennyiség között (gondoljunk pl. a kölcsönzések
adatainak tárára). E rendszer legrafináltabb mód-

szere a hatékonyságnövelésre pedig éppen az
adatbányászat (datamining). De joggal nevezhet-
nénk tudásbányászatnak is, amely az üzleti intelli-
gencia (az ember természetes értelmi képességét
mesterséges körülmények között utánozó rend-
szer) fontos része.

Az adatbányászat többféle megközelítéssel élhet.
A *felfedezés* esetében nincsenek előfeltevések és
keresett összefüggések, például mikor a kölcsön-
zési statisztika év közbeni vagy több évig tartó
alakulását vizsgáljuk. A *célzott adatbányászat* a
részletekben megbúvó összefüggéseket keresi. Ez
történhet akkor, amikor a felsőoktatási könyvtárban
azt próbáljuk kideríteni, hogy az idei évben miért
lett hirtelen kevesebb az elsőévesek beiratkozási
aránya. A *döntésautomatizálás* akkor praktikus, ha
nagyon gyors vagy nagyon sok ismert döntést kell
meghozni. Az előző évi tapasztalatok alapján kiad-
hatjuk-e és meddig az adott olvasótermi dokumen-
tumot az évnek ebben a szakában?

Az adatbányászathoz többféle ismeret szükséges:

- adatbányászati szakértelem és tapasztalat,
- üzleti tudás (könyvtárosul fogalmazva inkább: a
szervezeti célok, eszközök, ügymenet ismerete),
- adatismeret, amely a tárolt adatok és a közöttük
való tranzakciók informatikai jellemzőire vonat-
kozik.

Mind e hasznos tudás gyakran nem található meg
egészében a szervezetben, ezért mindenhol gya-
kori, hogy az ismeretek egy részét külső szakértő
adja.

„Eddig is tudtuk, hogy a testnevelés szakosok jön-
nek legkevesebbszer a könyvtárba” – hangozhat el
egy felmérési projekt eredményének hallatán, ám
abban már kevésbé lehetünk „fejből” biztosak,
hogy mennyivel kevesebbszer, vagy hogy az évek
során a látogatások száma milyen tendenciát mu-
tat, illetve részleteiben milyen motivációk állnak a
hallgatói aktivitás változása mögött. Az egyhar-
mad–kétharmad szabály szerint akkor ad jó ered-
ményt az adatbányászati munka, ha a feltárt ösz-
szefüggések kétharmada igazolja a terület munka-
társainak sejtését, egyharmada teljesen új ismeret-
et ad.

Az adatbányászat üzleti alkalmazásaira – ügyfél-
szegmentáció, ügyfélérték-számítás, adósminősí-
tés, a lemorzsolódások vizsgálata, csalásfelderíté-
s, keresztértékesítés, bolti együttvásárlások, sze-
mélyazonosítás (oroszországi hadifoglyok), inter-
netes viselkedési szokások elemzése – ezen is-

mentetés írója nem tud minden esetben könyvtári megfelelőt írni. Mindazonáltal a könyvtári rendszer (vagy meghatározott részének) kölcsönzési adatai területre és használói szegmensekre lebontva az egyes könyvtárak állománymenedzsment-gyakorlata, a könyvtár-politikai döntéshozók vagy a könyvkiadók számára egyaránt értékes forrást jelentenek. A módszerrel elég pontosan lemérhető Az olvasás éve vagy a Nagy Könyv mozgalom valós hatása is. Figyelemmel kísérhető, hogy a beszerzési keret 1%-os növelése vagy csökkentése hogyan érezte magát az állomány minőségében vagy a használati adatokban. Az eljárás tehát érveket adhat a könyvtár kezébe, illetve segíthet a saját erős és gyenge pontok felderítésében is.

Mindezen eljárások meghatározott adatbányászati módszertanon alapulnak. Ennek pontjai:

- Az *üzleti cél* meghatározása. (Miben szeretnénk döntést hozni könyvtárunkban? Mondjuk: csökkentjük az oktatási rendszerből való kikerülést utáni könyvtári lemorzsolódást.)
- Az *elemzési feladat* megfogalmazása: felügyelt és felügyelet nélküli tanulás. (A korábban lemorzsolódottak milyen előzetes könyvtárhasználati mintát mutattak?)
- A *modellek* felállítása. Lehetséges összefüggések körének megválasztása; felügyelt tanulóhoz prediktív modellek, pl. döntési fák, neuronhálók görbék által határolt tartományokhoz, regressziók egyenesekkel vagy síkokkal való felosztása; felügyelet nélküli tanulóhoz klaszterezési eljárások, önszervező rendszerek és vizuális technikákon alapuló modellek. (Például a tanulmányi célú dokumentumok használatának azonosítása az adott csoportban.)
- A modellek *megvalósítása*. (Benne visszacsatolás: megfelel-e a modell az üzleti céloknak? Algoritmusok futtatása a kurrens kölcsönzési adatokon, a lemorzsolódni hajlamos használók azonosítása.)

Felmerül a kérdés, hogy mivel több vagy más az adatbányászat, mint a statisztika. Nos, az adatbányászat célzottan üzleti (vagy közszolgálati) alkalmazásokat szolgál, míg a statisztika ennél általánosabb jellegű. Pl. nem elég az együttműködést kimutatni (mondjuk: aki gyerekkönyveket kölcsönöz, visz-e zene-CD-eket is), hasznos tudni, hogy melyik ösztönzi a másik iránti érdeklődést. Az adatbányászatban megengedhetők elhanyagolások és „nagyvonalúságok” (pl. egy változó normális eloszlásának feltételezése). Így a statisztika a *milyen*, az adatbányászat a *miért* kérdésre válaszol inkább.

Szintén magyarázatot érdemel az adatbányászat és a hagyományos adatelemzések viszonya. Ahogy a szerzők kifejtik: az utóbbi – jellegeből adódóan – csak előre meghatározott dimenziók mentén lehetséges, így nem lehet például az ismeretterjesztő dokumentumok használatának fiókkönyvtárak közötti területi megosztására keresni. Ezért célszerű az adatbányászat eszközeivel feltárni az üzletileg fontos dimenziókat (célzottan feltenni a kérdéseket), ezek alapján adattárházat építeni (adattárház: az integrált rendszerből külön, adatbányászatra szolgáló tárló), és hagyományos adatelemzéssel folyamatosan nyomon követni az üzleti eseményeket. Így az adatbányászat feltáró módszer, a nagyüzemi kiaknázást viszont hagyományos módon érdemes végezni.

A könyv előszava azt ígéri, hogy egyaránt szól üzletemberekhez (ideérthetjük a könyvtárvezetés szakembereit) és informatikusokhoz. Ha ez teljesen nem is valósul meg, *Az adatbányászatról egyszerűen* fejezet joggal tarthat igényt széles szakmai nyilvánosságra, és a könyvtárosképző intézmények kötelező olvasmányok listáján való szereplésre. A további fejezetek – *Amit az üzleti intelligenciáról tudni kell*, *Az üzleti intelligencia alkalmazásai*, *Adatbányászati projektek szervezése*, *Az adatminőség kérdése*, *Az adatbányászat technológiai háttere* – szintén izgalmas könyvtári gondolatokat ébresztenek, de nehezebb olvasmányt jelentenek. A témában való tájékozódást segíti a könyv végén a független: további alkalmazási területek, neuronhálók matematikai alapjai, fogalomjegyzék, angol–magyar adatbányászati szótár, rövidítések jegyzéke, irodalomjegyzék, tárgymutató.

Az alkalmazási területek között olyan könyvtárosi figyelemre méltó szakaszok találhatók, mint webbányászat, weboldalak rangsorolása (gondoljunk a katalógustételek megjelenítési sorrendjére), a kattintássorok elemzése (pl. hány kattintással jut el az olvasó a számára fontos tételig), intelligens internetes keresés (amely a tájékoztató könyvtárosok hatékonyságát növeli), az üzleti intelligencia, online tartalom publikálása, tudáskinyerés szövegből, a használók elégedettségének mérése, vagy éppen hírszerzés. (A könyvtári alkalmazásról korábban referátum készült: *Nicholson, Scott: Adatbányászat a könyvtárban.* = *TMT*, 2004. 12. sz. p. 564–565.)

Mikulás Gábor
(GM Consulting)