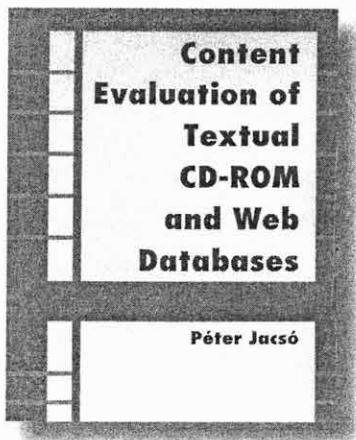




Mi van a megaladatbázisok keresőmaszkja mögött?

Könyvtári-információtudományi szakkönyvkiadásunkról egyáltalán nem mondhatjuk azt, hogy bővelkedik monográfiákban, s különösen nem szakirodalmi fordításokban. Ritkaságnak tekinthető hát *Jacsó Péter* 2005 őszén magyar nyelven, elektronikus formában közzétett könyve. A Magyarországról elszármazott szerző a Hawaïi Egyetem informatikai tanszékének professzora Honolulu-ban. A mű eredetileg angolul íródott¹, magyarra *Murányi Péter* ültette át *Szöveges CD-ROM és webes adatbázisok tartalmának elemzése* címmel. A kiadás a Könyvtári Intézet közreműködésével valósult meg, az elektronikus közreadást, a folyamatos hozzáférést pedig a Magyar Elektronikus Könyvtár (MEK) biztosítja, többféle digitális dokumentumformátumban. Oldalszámokkal ellátott változata kettő van: egyik Adobe Acrobat Readerrel olvasható *pdf*, másik *MsWord* formátumú. A két változat tördelése megegyezik, fájlformátumuktól eltekintve egyformák. Mindkettő alkalmas arra, hogy hivatkozható verzióknak tekintsük őket: az elektronikus megjelenésű mű szöveghelyeire a hagyományos módon, az oldalszámok megadásával utalhatunk.² A harmadik, *html* változatban közzétett szöveg fejezetenkénti szövegtestekben olvasható a képernyőről, hagyományosan nem hivatkozhatunk rá, ellenben e változatban a hipertext hivatkozások elevenek, így a külső forrásokat könnyűszerrel bevonhatjuk a könyv szövegtörzsének bűvkörébe. A mű tömörített csomagolású változatokban is letölthető. 336 illusztrációja különálló tömörített mappában, *gif* képekben is rögzítve van.



Content evaluation of textual CD-ROM and Web databases (magyar)

Szöveges CD-ROM és webes adatbázisok tartalmának elemzése [hálózati elektronikus mű] / Jacsó Péter; ford. Murányi Péter. – Budapest : Könyvtári Intézet: Magyar Elektronikus Könyvtár, 2005.

URL: <http://mek.oszk.hu/03000/03040>

Formátum-változatai:

Html. – ISBN 963-201-617-3

URL: <http://mek.oszk.hu/03000/03040/html/>

Pdf. – 299 p. – ISBN 963-201-618-1

URL: <http://mek.oszk.hu/03000/03040/03040.pdf>

MsWord. – 299 p.

URL: <http://mek.oszk.hu/03000/03040/03040.doc>

További zip-pel csomagolt verziók: html, pdf, MsWord, a képanyag gif fájlokban.

A teljes művet érteni szándékozó és elemezni kívánó olvasáshoz természetesen elkerülhetetlen az egy híján háromszáz oldal szöveg kinyomtatása, összefűzése, azaz könyvszerűvé tétele. E recenzió így egy, a pdf fájl műanyagspirál-gerincű könyvvé fűzött, költség-takarékossági megfontolásokból, két oldalt egy A4-es papírméretre printelt példánya, valamint az eredeti, angol nyelvű nyomtatott könyv kézbevétele alapján készült.

A Jacsó-könyv műfaját tekintve is ritkaság: monografikus adatbázis-kritikának, adatbázis-elemzésnek nevezhetjük. A szerzőt több mint egynegyed százada foglalkoztatják az adatbázisok progressziói, anomáliáik diagnosztikája és e vizsgálatok eredményeinek publikussá tétele. A kórlapokra – mint ahogyan azt a könyv bevezető ajánlásából megtudjuk – sokszor maguk a szolgáltatók is kíváncsiak. Jacsó adatbáziselemzési-műveltség terjesztést is végez kitartó türelemmel: több folyóiratban találkozunk számára külön fenntartott rovattal. Ezek a Jacsó-írások rövidebb-hosszabb, frappáns, publicisztikus cikkek vagy rendszeresen végzett elemzések eredményeit bemutató, azokat összegző tanulmányok. Cikkeinek összesített számát nehéz fölbecsülni, elképzelhető, hogy számuk az ezer felé közelít. A *Peter's Picks and Pans* az *Online* folyóirat rovata, amelyben „Péter” – a produktumok minőségi jellemzőinek megfelelően – dicsér és korhol, a Jacsó által kedvelt alliteráció kedvéért szabadon így is mondhatnánk: „Péter dicsér és dorgál”. Az *Information Today*-ben *Cheers and Jeers for xx* címmel az elmúlt év három-három legjobbnak ítelt, illetve leginkább elmarasztalt adatbázisa kerül bemutatásra. A *Thompson Gale* nagyszabású elektronikus információ-, tartalomszolgáltató és kiadó honlapján, a *Peter's Digital Reference Shelf* rovatban, havonta olvashatók Jacsótól különféle ismertető, ajánlások, kommentárok a tartalomszolgáltatók és e-kiadók újdonságairól, fejlesztéseiről.³ Egy kis merészséggel azt is állíthatjuk, hogy olvasmányos, népszerűsítő, meghökentető metaforákkal megmosolyogtató, ugyanakkor széles körű empirikus adatok értékelésére alapozó folyóiratcikkei, sajátos retorikájukkal – a napilapok nagyra becsült, irodalmi értékeket magukon viselő tárcáinak analógiája nyomán – amolyan *szaktárcák*.

A sajátos Jacsó-stílus és -módszer az adatbázisokat elemző nagyobb méretű művébe, az angolul 2001-ben megjelent könyvébe is bevonult. A könyvtárosoknak, információs szakembereknek, adatbázis-kritikával foglalkozóknak ajánlott kiadványból az érdeklődő sok hasznos információt szerezhet a gigantikus adatbázis-monstrumok igazi, Janus-arcáról, jacsoi kifejezéssel: anatómiájáról. A szakember ötleteket szerezhet elemzési eljárásokra és a szisztematikus elemzés fortélyait is ellesheti, a bőséges szakirodalmi meg-

alapozottságra támaszkodva pedig kitekinthet a témakörrel kapcsolatos kutatásokra, eredményekre.

A tizenkét fejezetre tagolt könyv az adatbázisok létrejöttétől, általános jellemzőik feltárásán keresztül az egyre kisebb részletekre kiterjedő elemzési módszerekig öleli fel a minőségi vizsgálat folyamatát, alkalmazható módszerét. Tetszi mindezt annak érdekében, hogy a felhasználó helyesen dönthesse el, melyik méregdrága adatbázist érdemes megvásárolnia, ha pedig már megvásárolta, illetve előfizetett rájuk, akkor világosan megítélhesse, vajon megfelelnek-e elvárásainak és a kitűzött céloknak.

A fejezetek tartalmuk szerint a következőkre terjednek ki:

- az adatbázistermékek fajtái, létrehozói és kiadói;
- a tartalmi elemzés általános szempontjai kutatástörténeti áttekintéssel;
- a szakterületi hatókör;
- az adatbázis dimenzióinak feltárása (méret; időbeli kiterjedés és mélység; frissesség);
- a feltárt források (a különféle dokumentum-típusok reprezentálásának módja; a magfolyóiratok repertorizálásának mértéke, rendszeressége; földrajzi feltártság; nyelvi megoszlás);
- a rekord tartalma (bibliográfiai adatok minősége; az ún. „értéknövelő” információtartalom);
- pontosság;
- az indexfájlokba rendeződő keresőelemek alaki konzisztenciája (egységesítés, autorizálási adattörzsek);
- a teljesség mértékének mérése;
- a tárgyi indexelés minősége;
- a referátumok minősége;
- az ár.

A kezdő szakaszokból érdemes kiemelnünk az érintett fogalmak (*adatfájl, adatfájl-előállító, adatbázis, adatbázis-kiadó, adatbázis-készítő, keresőszoftverek, adatbázis-szolgáltatások*)

megkülönböztetését a termékek jellemzőinek leírásán keresztül, az előállítás műveleteinek vázolásával és az e folyamatokban részt vevő piaci szereplők – szolgáltatók és szolgáltatásokat igénybe vevők – bemutatásával. Érzékelhetőkké válnak a tartalomszolgáltatás bonyolult, üzletitartalomipari szférákon belüli összefonódásai az adatfájl-előállítóktól az aggregált, több különböző szakterületi adatfájlt felvásárló és ömlesztő adatbázis-kiadóig és online szolgáltatóig. A bipoláris piaci szférában a szolgáltatók az előfizetési feltételek teljesítésétől függően „bérbe” adják a fogyasztóknak (a használóknak) a „bérelni” kívánt adatbázisrészeket. Fontos kiemelnünk: Jacsó ez utóbbiak, az adatbázishasználók nézőpontjából vizsgálódik.

A címben is jelölt megkülönböztetés – szöveges CD-ROM-ok és web adatbázisok – a későbbiek során tovább tagolódik: *szöveges CD-ROM-ok, webre vitt* (Web-borne) és *weben született* (Webborn) adatbázisokra, Murányi itt jól átültetett kifejezéseivel. Az első fejezet ismertető szakaszaiból világossá válik, hogy a CD-ROM adatbázisok virágkora az 1980-as évektől az 1990-es évek közepéig tartott. Az adathordozó megvásárlása és egy vagy néhány felhasználós licencelése jó alternatívát kínált fel akkor, amikor sok könyvtár számára a használati idő függvényében elszámolt online kapcsolatok még megfizethetetlenek voltak. Az internethasználat olcsóbbá válásával, a grafikus felületek elterjedésével, a szolgáltatók szerverein szédületes mértékben terebélyesedő adattárak webfelületeken váltak elérhetőkké. A szegmentált adatszerkezetű adatbázis-architektúrával épülő, webfelületen keresztül való tartalomszolgáltatást nevezi Jacsó *webre vitt, webre került* adatbázisnak. E típusnál jól alkalmazható például a mezőspecifikus keresés (a felhozott példák: DIALOG, Ovid, OCLC, SilverPlatter). *Weben született* adatbázisoknak nevezi azokat az adatbázisokat, amelyeket „először a weben tettek közre” (15. old.). Ezeknek példái között van többek között pozitív tulajdonságainak révén az *Internet Movie Database*, a kevésbé impozán-

sak között pedig az *Amazon.com* kereskedelmi adatbázis. A weben született adatbázisok általában silányabb képességeikről ismerhetők fel: „sokszor nincsenek böngészhető, mezőspecifikus indexeik, amelyek segítenének felderíteni az adatelemek pontosságát és következetességét, s általában nincs olyan adatelem, amely jelezné, mikor adtak új rekordokat az adatbázishoz.” (23. old.).

Az elemzett adatbázisok kategóriái: a címjegyzékek (címtárak), indexelő, referáló és teljes szövegű adatbázisok. A vizsgálatok zöme tizenhat adatbázisra terjed ki.

Az adatbázisok tartalmi elemzésének szempontjai a harmadik fejezettől különülnek el. A fejezetek sorrendje az adatbázisok módszeres átvilágításának lépései, fokozatai is lehetnek. Az adatbázis-használók köre egyre nő, ezzel a tendenciával viszont – erre Jacsó is figyelmeztet könyvének elején – nincs összhangban az adatbázisok *dimenzióinak* ismertsége. Az adatbázisok keresztmetszetéről, méretéről, minőségi jellemzőiről ugyanis csak megfelelő adatgyűjtések, tesztkeresések, elemzések, összehasonlítások révén szerezhethünk tudomást. Ezeknek a tényeknek, jellemzőknek a feltárása és ismerete viszont elengedhetetlenül szükséges akkor, amikor egy-egy intézménynek ki kell választania, melyik adatbázis-szolgáltatásra fizessen elő, melyik adatbázist vásárolja meg. Az információk után kutató felhasználónak is tisztában kell lennie azzal, hogy melyik adatbázis a legmegfelelőbb a számára, a kiválasztott forrás legáltalánosabb vonatkozásaiban a kitűzött célnak megfelel-e, és valójában mire is terjed ki a szolgáltatás szakterületi, földrajzi, nyelvi vonatkozásaiban. Jacsó részletesen bemutatja azt, hogy a gigantikus termékek önismertetéseire és reklámszövegeire a legkritikább esetben lehet megbízhatóan támaszkodni. A legbiztosabb az, ha kritikus vizsgálattal ki-ki maga is meg tud győződni támasztott igényeinek szempontjából egy-egy adatbázis tartalmáról és minőségéről, hiszen az impozáns számadatok és a jó benyo-

mást keltő szöveges ismertetőik mögött számtalan esetben bújnak meg megtévesztő, torzított adathalmazok, hamis állítások.

Nem könnyű feladat megállapítani és belátni az elektronikus hordozókon megjelenő adatbázistermékek méretét, kiterjedését. Amikor Jacsó a negyedik fejezetben az adatbázisok nagyságának és dimenzióinak felmérésével foglalkozik, a felmérésre használt módszereit is bemutatja a tesztelések ismertetésével. Kérdések sora merül fel a mérettel kapcsolatban: mit értünk adatbázis-méreten, mekkora az adatbázis, milyen mértékű és rendszerességű a weben közzétett adattárak új tételekkel való bővítése, mi képezi a leírás alapegységét. A statisztikai görbék alakulása szempontjából ugyanis nem mindegy például, hogy címváltozás esetén új entitásként kezelnek-e egy folyóiratot, a több műről készült kritika hány rekordra tagolódik, de nehézséget jelent a valódi nagyság körvonalazásánál a duplum rekordok kiszűrése is. Vizsgálati szempontot képez az időbeliség dimenziója is, szoros összefüggésben a mélység perspektívájával: lényeges meggyőződnünk a magfolyóiratok feltárásának kezdetéről, a retrospektív feldolgozottság teljességéről, az időbeli kiesésekről és a hézagosságok mértékéről is. A legsúlyosabb végtel az, ha a szakterületen fontosnak tartott források nem vagy csak szórványosan kerülnek be az adatbázisba. Magától értetődő dolog a magfolyóiratok feldolgozásának egyenletessége iránti felhasználói elvárás, és az is, hogy egy adott szakterületet az indexelő és referáló szolgáltatás valóban le is fedjen. A vizsgálatok sok esetben mégis rapszodikusán váltakozó feltártságról árulkodnak, legyen szó akár csak egy-egy folyóiratról vagy az évenként bekerülő rekordszámról, a frissességi kritériumokról. A felsorakoztatott esetek ismételt tanulságul szolgálhatnak arra is, hogy a kiemelt fontosságúnak tartott folyóiratok feltártságát legjobb egyenként is ellenőriznünk, ugyanis nem egy fontos magfolyóirat esetében mutatható ki hosszabb időszakra is kiterjedő fehér folt a feldolgozásban.

Az adatbázisok dimenzióinak és minőségének feltárását célzó, a háttérben megbújó jacsói kérdések összefoglalva:

- Mit mondanak a szolgáltatók saját termékeikről az adatbázis-leírásokban?
- Mit tudnak az önleírásokban felsorakoztatott tényanyagokból megmutatni?
- Mi ezekhez képest a valóság?

A jacsói vizsgálati módszerek:

- a szolgáltatók önbemutató állításainak összevetése a feltárható adatokkal;
- tesztkeresések eredményeinek bemutatása és elemzése;
- az eredmények és a levont tanulságok eseteleírásokkal, esettanulmányokkal való alátámasztása.

Az adatbázisok valóságának részletekbe menő vizsgálata további rejtőzködő összefüggéseket tár fel. A nyelvi, illetve a földrajzi mutatókból könnyen fény derülhet például arra, hogy a „nemzetközi” adatbázis milyen mértékben tartalmaz nem angol nyelvű dokumentumokra vonatkozó információt, illetve mennyiben tár fel az angol nyelvterületen kívülről is anyagot. Az eredmények sokszor lesújtóak. E vizsgálatok természetesen csak akkor végezhetőek el egyszerű adatlelviasos módszerekkel, ha a mutatók az adatbázis-méretet számszerűen is ábrázolni tudják, és ha a nyelvi vagy földrajzi kódok következetesen fel is vannak tüntetve a rekordokban. További érdekes jellemvonások – csökkenő, illetve növekvő, a nemzetköziség mértékét jelző tendenciák – deríthetők ki az ilyen szempontú feltártság éves összehasonlításából. A dokumentumokat feldolgozó adatbázisokban a „feltártság mértéke lehet széles, de sekély, a cikkek feltárása lehet túlzottan válogató, és néhány alapvető folyóirat hiánya [...] kizárhat egy vizsgált adatbázist abból, hogy komolyan számításba vegyék.” (96. old.) Jacsó ezt az állítását is példa- és esetengeteggel támasztja alá. Ebben a vonatkozásban a legszélsőségesebbnek ítéltető az, amikor a szakágazatok legmagasabb impakt faktorú

folyóiratai közül számos nem szerepel a szakterület egyébként rangosnak tartott referáló adatbázisában. Több esetben hiányzik maga a szakterület határozott körvonalazása és a lefedettség folyamatos biztosítása: az adatbázisgyártó vállalkozások ingatag „válogatási politikát” folytatnak, minek következtében a feldolgozottság folytonosságában hiatusok keletkeznek, és a szakterület minőségi folyóiratokkal való lefedettsége is kérdésessé válik. E hiatusok következhetnek abból is, hogy az egymást váltó tulajdonosok vagy főszerkesztők mást-mást gondolnak magfolyóiratnak, de lehetnek egy-egy gazdasági-menedzselési időszak válságát jelző „túlélési” tünetek is. Sokszor önkényesnek és indokolatlannak tűnik a fontos folyóiratok indexelésének elhanyagolása, és homályban marad a feldolgozottsági mutatók nem csekély hanyatlásának oka. Jacsó megállapításaihoz hozzátesszük: mindezen jelenségek azért veszélyesek, mert a felhasználó mélyreható adatbázis-analízis nélkül nem lehet tudatában a szövevényességeknek és hiatusoknak. Hamis illúziók között végzi keresését, és nem sejt, hogy a kapott eredmények milyen (nagy) mértékben lehetnek esetlegesek.

A rekordok tartalmi jellemzőinek vizsgálatkor Jacsó a rekordba kerülő adatok két nagy csoportját különbözteti meg: a bibliográfiai alapadatokat és a kiegészítő, ún. értéknövelő információkat nyújtó adatok csoportját. Behatóbban az értéknövelő adatokkal kapcsolatban vizsgálódik, ugyanazon cikkek más-más adatbázisból vett rekordjainak összehasonlításával. Ezek az információk vonatkozhatnak a nyelvre, dokumentumtípusra, a szerző affiliációira, a tárgyalásmódra, a célközönségre (tudományos, széles felhasználói közösségnek szánt stb.), a cikk terjedelmére (sorok, szavak száma), ábrákra, grafikonokra, az osztályozási jelzetekre vagy tárgyszavakra. Számunkra sem vitás, hogy a referátum megléte is jelentős értéknövelő tényező. Teljes-szövegű szolgáltatások esetében egy sor vizsgálati szempont

vethető ki, ilyen például a nyomtatott dokumentum elektronikus reprodukciójának és megjelenítésének kérdése. A hipertext és más adatkapcsolatok kiépítése további „hozzáadott érték” lehet az adatbázisban. E lehetőségek köztudottan széles körűek, Jacsó csak röviden ír le néhány példát: a kapcsolatok létrehozhatók adatbázisok között, rekordok között, a bírálóat összefűzhető a bírált művel, a szerzőre, műre vonatkozó hivatkozásokkal. Az értéknövelő információkat hordozó adatfészeségek váltakozó minőségét mutatja ki a következő fejezetekben mind az egyes adattípusok, mind jelölésmódjuk, mind következetességük szempontjából. Félrevezető viszont a nyolcadik fejezet címe, amely az angol eredetit – „Format and Content Consistency” – követve jelenik meg a magyar fordításban is „A formátum és a tartalom következetessége” formában. A címet látva az olvasó első gondolata az, hogy talán a metaadat- és szövegformátumokról eshet valamiféleképp szó a fejezetben. Gondolhat a rekordok adattartalmára is, a rekordformátum megkülönböztetett adatmezőinek következetes kitöltésére. Nevezetesen, az a gyakori felhasználói előítélet juthat erről eszünkbe, hogy a keresőmaszkban is felkínálkozó mezők, amelyek egy-egy adattípus keresését teszik lehetővé, ezekben az előítéletekben úgy jelennek meg, mintha az adatbázis minden rekordjának minden, a keresőfelületen is felkínálkozó mezőjében kötelezően ott lenne valamilyen adat, vagyis a felhasználó azt hiszi: *minden* mezőben *mindig van* adat. Ezt a problémát – mint utóbb a könyv szövegéből kiderül – a kilencedik fejezet tárgyalja „Completeness” – „Teljesség” címmel. A szövegformátumokról érintőlegesen egy korábbi fejezetben (a hatodikban), az értéknövelő tényezők között történik említés, a nyomtatott cikk elektronikus megjelenítésének problémaköréhez kötődően. A „*van* adat”, és minden esetben a kereső mentális ismeretével megegyező „*helyes* adat *van*” tételezés csapdait Jacsó érinti is néhány helyen, például amikor az ÉS Boole-operátor használatáról tesz említést:

a kilencedik fejezetben a szakirodalomból említ egy vizsgálati eredményt, amely kimutatta, hogy azok a felhasználók „*akik kettőnél több deskriptort használnak keresőkérdésükben a Boole-algebra ÉS operátorával összekapcsolva, nagy mértékben csökkentik a visszakeresés lehetőségét.*” (224. old.) Viszont a nyolcadik fejezetben nem ezekről, a fejezetcím alapján elgondolható kérdésekről van szó, hanem azokról a problémákról, amelyeket a könyvtári gyakorlatban az *egységesített tétel-fejek* alkalmazásával oldottak meg a hagyományos cédulakatalógusokban, a névvariánsokról (például rövidebb – hosszabb alak; szervezetek, cégek nevének a rövidített vagy a többnyelvű változatok stb.) az adott keresőeszköz rendeltetésének megfelelően egyetlen kiválasztott, kitüntetett alakra utalva. A különféle keresőelemek, a név-, címformák, dokumentumtípusok elnevezésének szabályozottságáról, egységesítéséről, autorizálásukról (authority control) van tehát szó. Tény, hogy ezek a keresőelemek valóban tekinthetők erősen *formalizált* adatelemeknek – helyettük akár számkódok vagy ikonok is állhatnak. Velük kapcsolatban viszont nem a „formátum”-ukról („format”), hanem *formájukról* (*form*), általában *nyelvi formájukról és alakjukról, morfológiai* (alaki) szerkezetükről beszélünk. A címben szereplő „tartalom” („content”) szó jelentése nem világos, a fejezet szövegéből sem lehet egyértelműen következtetni, mire is szeretett volna a szerző utalni. Jó lenne tudni, hogy talán arra az entitásra gondolt-e, amelyet nyelvi-írási képpel jelölünk: egy személy, intézmény, egy dokumentumtípus stb., és amelyek jelölésének inkonzisztenciáját tapasztalhatjuk az adatbázisokban, vagy talán általában az adatbázisok szöveges tartalmáról lenne szó, a tulajdonképpeni, rekordokba zárt dokumentumreprezentációkról, amelyek jó része kiesik a keresés látóköréből az inkonzisztens adatok miatt? Eltekintve a fejezet megfejthetetlen címétől, a szakasz tartalma mindenesetre arra figyelmeztet, hogy igen kevés indexelő-referáló

mamutszolgáltató fáradozik olyan koordináló, a keresőrendszer makrostruktúrájának fő konzoljához kapcsolódó *autorizálási adat-törzsek* (authority file-ok) kidolgoz(tat)ásán, gondoz(tat)ásán és következetes alkalmazásán, amelyek lényegesen jobbíthatnák a visszakeresés hatékonyságát. Arra is rávilágít, hogy sokszor rejtve marad az, hogy a sokféle adatbázist egyetlenegy nagy rendszerbe ömlesztő tár indexei honnan, melyik adatmezőkből generálják betűrendes mutatóikat.

Döbbenetes az adatbázisokban a szerző, cím és más, a „kontrollált” lista látszatát keltő indexekben a félregépelések vagy más figyelmenlenségek folytán keletkezett pontatlanságoknak a rendkívül nagy hibaszázaléka. A legkirívóbbak közé tartozik az ENGLISH, nyelvet jelölő szó változatainak sorjázása a nyelvi indexben, az EMGLISH-től az ENGLUSH-on át hosszan sorakozó hibavariálódással. Szerencsésebb esetben az egyes hibás vagy nem egyeztetett névalakok a betűrendes mutatóban közel, rosszabb esetben viszont messze kerülnek egymástól, a betűsor egészen más tájékaira. A hozzájuk kapcsolódó tételek így teljességgel kizáródnak a keresési folyamatból és sem a karakterhelyeségen alapuló keresési-megfeleltetési mechanizmus, sem a betűrendes lajstromoztatás során nem kerülhetnek elő.

A pontatlan írásmóddal rögzített keresőszavakat – hacsak az ugyanazon entitást jelölő szavak nem kerültek el messze egymástól –, viszonylag könnyű észrevenni az egyes mezők szövegéből generált indexek pásztázásával. Sokkal nagyobb veszélyt jelent az, ha bizonyos adattípusok mezője *adattartalom nélkül* marad, illetve váltakozik bennük az információk megléte vagy hiánya. A pontatlanságok és következtelenségek az indexek listáiból esetleg feltűnhetnek, egyegy adat meglétének és hiányának volta viszont *rejtve* marad. Nehezíti a hiányosságok feltárását az a tény is, hogy adatbázisonként változó konvenció, milyen adatelemek kerülnek a rekordba kötelező jelleggel. A mezők adattarta-

lommal való ellátása, kitöltése egyes esetekben hallgatólagos megegyezés vagy dokumentálatlan hagyomány. Megfelelő tájékoztatás nélkül a felhasználónak erről természetesen nem lehet tudomása. A mezők adat-hiányai egy-egy adat-feltöltési időszak rovására is írhatók.

A felhasználó szerencsésebb helyzetben van, ha az adatkategóriák jegyzésének gyakorlatáról maga a szolgáltató nyújt korrekt tájékoztatást. Az esetek többségében viszont ez a tájékoztatás vagy nem felel meg a valóságnak, vagy el is marad. Ilyenkor egy másik vizsgálati módszerhez, az indextételek számadatainak és arányainak vizsgálatához folyamodhatunk. Az adatok összevetéséhez természetesen ismerni kell a teljes adatbázis alaptételeinek rekordszámát, ehhez viszonyítva állapítható meg a rekordokba beírt vagy hozzákapcsolt adatkategória, mint például az alapértelmezett nyelv és országcód vagy a dokumentumtípus, a kiadási év megléte, illetve hiánya. Összetettebb feladatot jelent a hiányok megállapítása akkor, ha bizonyos dokumentumtípushoz (például kritika) kötendő, de esetenként hiányzó adatról van szó (a kritikaként megjelölt rekordban a nyelvi kód hiánya).

Jacsó nyomatékosan figyelmeztet arra a veszélyre, hogy végzetesen félrevezetheti a keresőt, ha egy-egy ilyen adatfélése az adatbázis rekordjainak csak töredékében van jelen, ennek ellenére a felhasználói nézetben olyan hozzáférési pontként kínálkozik fel, amely a teljes adatbázisra vonatkozó szűrést képes elvégezni. A felhasználó könnyen eshet abba a tévhitbe, hogy keresésével a teljes adatbázist pásztázta át, holott az adatkategória meglétének hiánya eleve leszűkíti a keresési tartományt. A keresési szempontok sokfélesége a tapasztalatlan bűvárkodó számára veszélyeket rejtget és könnyen válhat esetlegesség-növelő tényezővé.

Jacsó kiemelt hangsúllyal tárgyalja a tárgyi indexelés minőségét, amelynek négy szempontját emeli ki: az *alaposságot*, a *specifikusságot*, a *pontosságot* és a *következetességet*. Értékelé-

si eljárása magának a szabályozott szótárnak – a tárgyszólistának vagy a tezauszznak a vizsgálatával kezdődik. Fontos kritériumként emelkedik ki a szótár alkalmassága a szakterület lefedésére, a deskriptorcikkek részletessége és relációinak követési lehetősége, a hierarchiaszintek optimális tagoltsága, a szótár korpuszának szakirodalmi megalapozottsága – ugyanakkor a felhasználói ismereteknek való megfelelés elvárása is, többnyelvűség esetén a fogalmak adekvát megfeleltetése. További soktényezős vizsgálati aspektus a prekoordinált jegyzékek fogalmainak szerepeltetése a rekordokban, a könyvtári szakemberek előtt jól ismert problémákkal, mint például a tárgyszavazás *szubjektivitása*, az *alulindexelés – túlindexelés* dilemmái a tételekhez csatolt fogalmak mennyiségének vonatkozásában vagy a *túl tág – túl specifikus* tárgyszavazás rejtett veszélyei. Az esettanulmányokban alkalmazott szempontok és módszerek egyben más szituációkra és igényekre is ráhúzható vizsgálati módszerek know-how-ja.

Megkülönböztetett figyelmet érdemel a referátumok minőségéről szóló fejezet. Más-más minőségi elvárások támaszthatók az *indikativ* (tartalomra utaló), az *informativ* (a tartalmat összefoglaló), más a *kritikai* (szakértői kommentárral kiegészített) referátumokkal szemben. Az itt közölt értékelési kritériumok egyik irányadója az American National Standards Institute (ANSI) és a National Information Standards Organisation (NISO) közös, 2000-ben átdolgozott szabványa, amely tartalmazza az alapelveket, amelyek szerint a referátumok készítenők. Egy másik jelentős irány a sokfelé elterjedő felcímkezett, ún. strukturált referátumok és szemlék értékelési elveit próbálja meghatározni. Egy harmadik módszer a makrostruktúrák mintáit állítja fel, és megvizsgálja, hogy annak részei, mint a bevezetés, módszerek, eredmények és tárgyalás, miképpen vannak jelen, valamint azt kutatja, hogyan függ össze a makrostruktúra a referátum típusával. A minőségi kritériumok között vizsgálják a

követett referálási szabvány elemeinek meglétét és a nyelvtani szempontokat (például a harmadik személyű közlés érvényesülését), a tömörséget stb. Minőségi vizsgálat tárgya lehet az újraírt referátum – a változtatás, az újraírás mértékének függvényében, valamint az olvashatósági és informativitási szempontok is.

Murányi Péter szaktárgyi látóköre nyilvánvalóan hozzájárult ahhoz, hogy fordításából a jacsói mondanivaló lényege jól kiszűrhető. A fordítás erőnei között kell megemlítenünk mindenekelőtt a szakkifejezések megfelelő használatát, átültetését. Kevés olyan terminus technicust találunk, amely nem adekvát vagy nincs lefordítva. E kivételek olyan kifejezéseket érintenek, amelyeket esetleg a szerző sem használt találoan (ld. a nyolcadik fejezettel kapcsolatos észrevételeket, hozzáfűzve azt is, hogy helyes lett volna a fejezetcímhez egy fordítói megjegyzést tenni), vagy olyanokról van szó, amelyek csak a külföldi szakirodalomban ismert terminusok, bevezetésük, meghatározásuk, elterjedésük, a velük kapcsolatos honi szakirodalmi konszenzus kialakulása hosszabb folyamat függvénye. Elfogadható ilyenkor az a megoldás, hogy a fordító nem akar előzmények nélküli, kierőszakolt magyarázást eszközölni, az idegen szót változatlanul hagyja. Egyes esetekben a használt terminusra vagy kifejezés-együttesre más javaslatok is felmerülhetnek. Az adatbázisokat összeíró adattárak esetében pl. a címjegyzék adatbázis / útmutató adatbázis (directory) helyett a kontextustól függően inkább az *adatbázisjegyzék / adatbázistár / adatbázisok katalógusa, regisztere* javasolható. Egy másik példa: a „cover-to-cover coverage” fogalmát Murányi a „borítótól borítóig feltárt” (59. old.) összetétellel fordítja. Annak ellenére, hogy e tükörfordításnak publicisztikai írásokban is van előfordulása – annak érzékeltetésére például, hogy valaki egy könyvet folyamatosan, az elejétől a végéig, az elsőtől az utolsó sorig elolvas – itt mégis gördülékenyebb lenne a magyar nyelvi fordulatokhoz és a szaknyelvhez is

közelebb álló megoldás: „*minden egyes cikk repertorizálása*”, vagy: „*válogatás nélküli feltárás / cikkfelvétel / repertorizálás / dokumentálás / indexelés*”. Helyesen marad lefordíthatatlanul pl. a *buildware, dataware, searchware* kifejezés (11–12. old.), a hozzájuk kapcsolódó definíciók átültetett szövege érthetően megvilágítja jelentésüket. Napjaink alapvető problémája érhető tetten e jelenségekben: a szakmai és tudományos megnevezésekről, recepciójukról, a megnevezések magyar nyelvű átültetésének és befogadásának nehézségeiről és összetettségéről van ugyanis szó, amelynek folyamatát *Tolcsvai Nagy Gábor* a következőképpen magyarázza: „*a megnevezés nem egyszerű kijelölés, hanem összetett megismerés. A megismerést innen tekintve vannak esetek, amelyek egyszerű, „buta” feldolgozásúak, azaz valóban csak valamely tárgy kijelölése történik meg [...], míg a skála másik szélsőbb tartományában az erőteljes feldolgozás valószínűsíthető*”⁴. A korábbi korszakok szaknyelvének alakulása, *Tolcsvai* leírását alapul véve legalábbis, mintha egyszerűbb lett volna a mainál: az idegen nyelvű minták átvételének módszere a gyengébb vagy erőteljesebb magyarítási törekvésekkel változott. Az 1990-es évektől a „*bizonytalanság, orientálatlanság a legfőbb jellemző.*” *Összetettebb folyamattal, „belsőleg tagolódo sztenderdel” kell számolnunk, miközben „a szakmai és tudományos megnevezés nem puszta forma (nyelvi forma) kérdése, hanem a megismerés-folyamatok nyelvi leképezésének tömör változata.*”⁵ Egy könyvtár-informatikai szakszöveg megfelelő átültetése – még ha „csak” fogalmi szintű leírásról van is szó benne – ezért sem jelent manapság egyszerű feladatot.

A *Jacso*-könyv kapcsán elvárásaink között azonban már joggal szerepelhet az, hogy a fordításon átüssön a *Jacso*i retorika esszenciája, az a könyvnyedségre való törekvés, amellyel a szerző nyilvánvalóan az olvasó dolgának a megkönnyítését – a befogadást, a megértést szeretné könnyíteni. A szerző a kiterjedő adatelemzéseket a szak-

rodalmi eredményekbe és esettanulmányokba ágyazottan, populisztikus–publicisztikus stíluslemekkel jeleníti meg. Ez a jellegzetes retorika, vélhetően a szerző előadásainak tapasztalatai során alakulhatott ki, mégpedig amerikai hallgatósághoz igazodva. Az eredeti szöveget az előnyelvi effektusokra hangsúlyosan támaszkodó prezentációs előadásmód jellemzi. A fordítás szó szerint szóról szóra történik: lelkiismeretesen megtalálunk benne minden lexikális elemet, minden tagmondatot, minden hasonlatot, minden idiomatikus és metaforisztikus szólást, sőt, minden *Jacso*i retorikai túlzást. Egy végletes példát ragadok ki ennek illusztrálására:

Az eredetiből (7. old.):

„*Without the capability of viewing the index, finding all misspellings is like trying to find a black ring in a dark tunnel on a moonless night.*”

A fordításban (16. old.):

„*Az index megjelenítésének lehetősége nélkül az összes elírás megtalálása olyan, mintha egy fekete gyűrűt próbálnánk megtalálni egy sötét alagútban egy hold nélküli éjszakán.*”

A eredeti szöveg szavainak egymásutánjához való ragaszkodás végig nehézkessé teszi a magyar szöveget, amelynek a magyar nyelv természetéhez és stílár kifejezési lehetőségeihez, valamint a magyar befogadói publikumhoz kulturálisan is idomított változatát érdemes lenne még kidolgozni, mégpedig úgy, hogy a bekezdések, gondolategységek, fejezetek függvényében, a lexikális egységek ekvivalenciája helyett a referenciális tartalom visszaadására kerülne nagyobb hangsúly. Közben természetesen a nyelvtani, szórendi, egyeztetésszerű tévesztések és értelemzavaró hibák kiküszöbölésére is figyelni kellene. Bizonyára nem könnyű, de megoldható feladat ezeknek az elvárásoknak az érvényesítése közben érzékeltetni a jellegzetes, nyelvi eszközökkel játszadozó, populisztikus és szlenges elemek belopásától sem visszariadó *Jacso*i argumentációt és retorikát.

Az olvasóban óhatatlanul felmerül az a kérdés, hogy érdemes-e és mikor érdemes egy gyorsan elavuló tényanyagra alapozó, a dinamikusan változó információs technológiai környezethez kötött tematikájú könyvet lefordítani más nyelvre. A könyv számára végzett tesztek 1999 júliusa és 2000 novembere között folytak le, amelyek közül egyeseket 2001-ben ellenőriztek. A magyar kiadás négy év múlva készült el, 2005 őszétől tölthető le a Magyar Elektronikus Könyvtárból. Fél évtizednyi időszak egy-egy megalapító adatbázis életében akár milliós nagyságrendű rekordgyarapodást is jelenthet. A vizsgált adatbázisok magját képező, zömmel amerikai adatbázisok nem mindegyike érhető el legalább egy ponton Magyarországról, egy-egy honi intézményünkben pedig legfeljebb csak néhány, a recenziált könyvben is vizsgált adatbázis áll a magyar kutatók rendelkezésére. Mondhatnánk ez alapján azt is, hogy a könyvben felsorakoztatott tények, cégek, szolgáltatók, szolgáltatásaik és termékeik elemzése csak kis részben érdeklik a magyar közönséget. Hogy ezt mégsem indokolt állítanunk, magyarázhatom azzal is, hogy az adatbázisok mérete, a rekordok száma ugyan bizonyára több tízezerrel, több százezerrel nőtt, viszont, azt a kérdést, hogy változott-e és miben változott mindeközben a hasonló szolgáltatások minősége, épp a jacsói *extenzív szempontrendszerrel és módszerrel* végzett alapos, korábbi eredményeket is egybevető elemzésekkel tudnánk megállapítani. Erős a gyanúm, hogy nem sok, az adatbázisok szöveges tartalmát érintő minőségi változás volna kimutatható. Ezen kívül érdemes elgondolkoznunk azon is, hogy az adatbázisok folytonosan változó, történetiségüket elvesztő valóságát talán nem hiábavaló többdimenziós keresztmetszeteket rögzítő pillanatfelvételekkel, hagyományos, szövegesen leíró, képi illusztrációkkal ellátott közléssel is megörökítenünk az utókor vagy közeli jövőbeli, saját elemzéseink számára. Mégpedig kötött könyvtestű, igazi papír-hordozón is⁶. A mű egy ilyen keresztmetszetet rögzít, amikor az adat-

bázisok minőségi állapotát mutatja be, elemzési szempontrendszerrel állít fel és elemzési módszereket dolgoz ki. E szempontrendszer és módszer méltán beépülhet a könyvtárosok és információs szakemberek szemléletébe, mindennapi munkájába, és fontos, hogy belekerüljön az oktatók segédanyagába is. Nem mellékes az a körülmény sem, hogy az adatbázis-analízis fortélyainak feltárása közben Jacsó elszántan lobbizik a felhasználó érdekeiért.

Záradékkomban *Mezei Balázs* filozófus ide illő, találó megjegyzését idézem föl a szakfordítás céljairól és indítékairól. Szerinte egy művet többek között akkor indokolt lefordítani, ha „*a lefordított mű annyira fontos a szakma szempontjából, hogy hozzáférhetővé tétele minden bizonnyal a célnyelv szakmai kultúrájának a megtermékenyítéséhez vezet.*”⁷ Az adatbázisok elemzése, értékelése új feladat és nem egyszerű feladat a könyvtárosok, információs szakemberek számára. Jacsó könyvének magyar fordításától éppen ezt, a magyar könyvtár-informatikai szakmai kultúra megtermékenyítését szeretnénk remélni, mindenekelőtt azt, hogy segíti ennek az újfajta szakmai műveltségnek a kialakulását. Titkon azt is, hogy az itt bemutatott magyar fordítás-kötet nem egyszeri eset marad, hanem a külföldi szakirodalom legfontosabb monográfiáinak magyar szakfordítási könyvsorozatát is elindítja.

Dudás Anikó

Jegyzetek

1. JACSO PÉTER. Content evaluation of textual CD-ROM and Web databases. Englewood : Libraries Unlimited, 2001. (Database searching series). A Database Searching Series sorozat első tagja 1989-ben jelent meg, azóta mintegy 7-8 kötetem került kiadásra a könyvsorozat keretében.
2. Ezt a hivatkozási módszert követem ebben a recenzióban is. A zárójelben feltüntetett oldalszámok – ha ezt másképp nem jelzem – a magyar, oldalszámozott kiadásokra vonatkoznak.
3. Jacsó hálózati tájékoztatópultjának hozzáférése: <http://reviews.gale.com/index.php/digital-reference-shelf>
4. TOLCSVAI NAGY GÁBOR. Alkotás és befogadás a magyar nyelv 18.

század utáni történetében. Bp.: Áron, 2004, p. 155.

5. I.m., p. [156.]
6. Érdekességként megemlítem – ha már a könyvben az adatbázisárak mérlegelésének szempontjairól is szó esik az utolsó fejezetben –, hogy az elektronikusan kiadott magyar fordítás egy példányának ki-nyomtatása és bespiráloztatása, átlagos 20 Ft-os oldalankénti árkal- kulációval számítva, több mint 6000 Ft-ba kerül, kétszer annyiba, mint egy hasonló terjedelmű, szerény kivitelezésű nyomtatott könyv,

amely ráadásul valószínűleg tartósabb és esztétikusabb darabja le- het egy könyvgyűjteménynek, mint a házikészítésű, printelt, spirálo- zott változat. További költségekkel bekötött példány elkészíttetése is elképzelhető. E megjegyzés közben természetesen nem vonom kétségbe az elektronikus közzététel és a nyílt, távoli hozzáférés köz- ismert előnyeit.

7. MEZEI Balázs: Magyar lexikográfia. Magyar Filozófiai Szemle. 1995. 3-4. sz., p. 599.



MINERVA eredmények

A MINERVA Plus projekt, amelyben hazánk is vezető szerepet vállalt, kitűnő minősítést ka- pott a digitalizálás európai szintű összehangolásában kifejtett tevékenységéért. A projekt irányelvei több országban elősegítették a digitalizálás alapelveinek kidolgozását, vala- mint a kiadványok használatát.

A MINERVA Plus legfontosabb eredménye az Európai Unió új tagállamainak integrálása, valamint a nemzetközi munkakapcsolatok elősegítése volt, amely a „Jó példák” széleskörű cseréjéhez és a MINERVA kiadványok fejlődéséhez vezetett.

Ennek következtében a MINERVA Plus Projekt kitűnő minősítést kapott az Európa Bizottságtól: vagyis az eredmények „felülmúlták az elvárásokat”.

Végül az Európa Bizottság kifejezte azt a kívánságát, hogy a projekt folytatódhasson a MINERVAeC kezdeményezés keretében, valamint hogy az nemzeti képviselők a saját országukban hozzanak létre munkacsoportokat a kulturális tartalmak digitalizálásának elősegítése érdekében.

<http://www.mek.oszk.hu/minerva> <http://www.minervaeurope.org>

(Szalóki Gabriella (OSZK) tájékoztatása, 2006. márc. 23.)

Javaslat egy Magyar Internet Archívum létrehozására

Az OSZK MEK Osztálya javasolja egy Magyar Internet Archívum létrehozását. Az interneten megjelenő tartalom a folyamatos és rohamos növekedése mellett egyben folyamatosan és rohamosan pusztul is. Digitális kultúránk megőrzése nagyon fontos, de nálunk még fel nem ismert, meg nem oldott feladat: komplett webhelyek nyilvános gyűjteménye nem létezik még Magyarországon. Pedig ilyen archívumok a világ sok országában létrejöttek az elmúlt években, ahol vagy a teljes nemzeti webteret, vagy válogatott site-okat mentenek le rend- szeres időközönként, és tesznek kutathatóvá a jelen és a jövő számára. A feladat nagysá- ga és komplexitása miatt ez mindenképpen több intézmény együttműködését igényli. Töb- bek között a következő feladatokat kell megoldani: az archiválendő kör lehatárolása, az automatikus időszaki mentés és gazdaságos tárolás technológiájának kidolgozása, aján- lások készítése a jól archiválható web-site struktúrákhoz, metaadatok hozzárendelése az archivált anyaghoz, intelligens keresőprogramok beépítése, a webarchívumok jogi stá- tusának tisztázása...

(F: Drótos László, mekdl@iif.hu, OSZK-MEK (KIT Hírlevél, 2006/26. jún.28.)