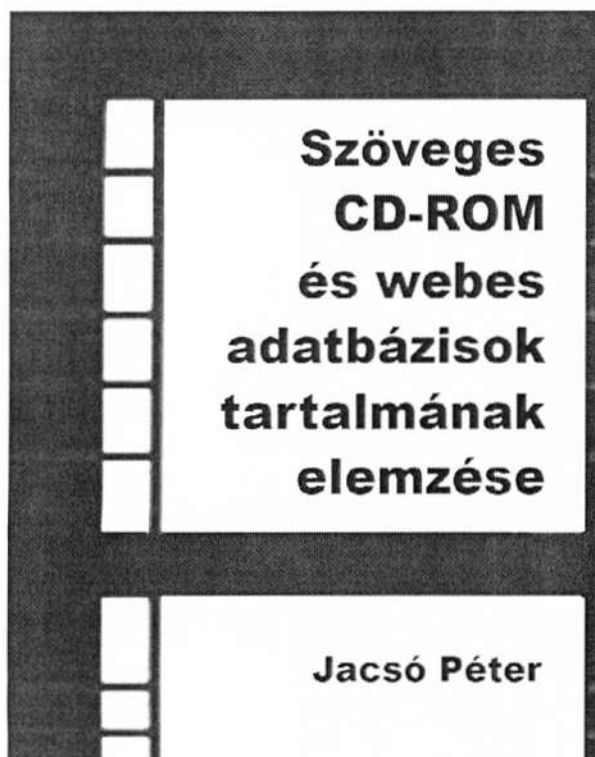




Szöveges CD-ROM és webes adatbázisok tartalmának elemzése



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

Szöveges CD-ROM és webes adatbázisok tartalmának elemzése [elektronikus dok.] / Jacsó Péter, ford. Murányi Péter. - Szöveg és képek. - Budapest : Könyvtári Intézet, 2005

Főcím a címképernyőről

ISBN 963-201-617-3 html

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

Figyelemgazdaságra épülő világunk – sok más mellett – a fogyasztóvédelem virágkorát is elhozta: ahogy erősödik és válik mind rafináltabbá a reklámpiac, úgy kell (kellene) egyre jobban ellenőrizni és minél szélesebb körben nyilvánosságra hozni – lehetőleg független – szakértőknek, hogy mi is van a hangzatos marketingszövegek és a vonzó csomagolás mögött. Jacsó Péter, a Hawaii Egyetem informatikai tanszékének professzora a nyolcvanas

évek vége óta foglalkozik az offline és online adatbázisok minőségének vizsgálatával, s számtalan cikkben, előadásban értékelte már ezeket a sokszor méregdrága termékeket és szolgáltatásokat, mutatott rá hibáikra, hiányosságaikra. 2001-ben egy angol nyelvű, szakmai körökben elismeréssel fogadott könyvben foglalta össze azokat az ismereteit és tapasztalatait, amelyek segíthetnek a gyorsan szaporodó adatbázisok közötti választásban, illetve minőségük felmérésében. A mű magyar változata 2005-ben készült el Murányi Péter szakszerű fordításában, és jelent meg a Könyvtári Intézet gondozásában, mégpedig kizárólag digitális formában kiadott könyvként, a Magyar Elektronikus Könyvtár gyűjteményében. (Kár, hogy a szöveg fordítás és lektorálás következtében néhány félkész, értelemzavaró mondat benne maradt, ami egy ilyen, a minőség fontosságáról szóló írásnál különösen kellemetlen.)

Bevezetőjében a szerző ezt írja: „Ez a könyv elsődlegesen azoknak szól, akik tanácsot adhatnak a végfelhasználóknak a megfelelő adatbázis kiválasztásában: iskolai, főiskolai, nyilvános és szakkönyvtárak könyvtárosainak és információs szakembereinek; továbbá azoknak, akik bírálják és értékelik az adatbázisokat vagy használati útmutatókat és címjegyzékeket készítenek az adatbázisokról...”. Könyvtárosok számára tehát többszörösen is érdekes ez a képekkel, esettanulmányokkal és irodalomjegyzékkel bőségesen ellátott könyv. S miközben borzongva olvasunk a jó vagy legalábbis hangzatos nevű adatbázis-előállítók és -kiadók „kredenceiben rejtőző csontvázakról”,* remélhetően tanulunk a profik hibáiból, és mi már nem követjük el őket, amikor saját OPAC-ot, bibliográfiai adatbázist, vagy teljes szövegű digitális könyvtárat tervezünk és építünk.

* „Searching for Skeletons in the Database Cupboard” – Jacsó Péter gyakran idézett kétrészes cikke a *Database* folyóirat 1993. februári és áprilisi számában.

A könyvnek mind a 12 fejezete az adatbázisok egy-egy jellemző tulajdonságát veszi górcső alá, módszertani útmutatót ad az alapos, szakszerű vizsgálatukhoz, de ügyes trükköket is bemutat „gyorstesztet” elvégzéséhez. Jacsó Péter ebben az írásában kizárólag csak a tartalommal foglalkozik, a felhasználói felület, a keresőnyelv vagy a kapcsolódó egyéb szolgáltatások csak érintőlegesen jelennek meg náhol. Az illusztrációként használt mintapéldák angol nyelvű (nagyreszt amerikai kiadású), bibliográfiai vagy teljes szövegű CD-ROM-ok, illetve webes szolgáltatások, köztük több fontos könyvtár- vagy információtudományi adatbázis, amelyek – a lyukas cipőben járó suszter esetéhez hasonlóan – sajnos többnyire a rossz példák között szerepelnek.

Az első fejezet a *fogalmakat és szakkifejezéseket* tisztázza: megtudhatjuk belőle, hogy hogyan lesznek az adatfájlokból szolgáltatásra alkalmas adatbázisok, hányféle közreműködő bábáskodik a megszületésük és a felhasználókhoz való közvetítésük körül, és milyen módokon szolgáltatják őket. A szerző szerint érdemes például megkülönböztetni a saját online szolgáltatásból webre vitt és az eleve weben született adatbázisokat. Mint a későbbi fejezetekből kiderül, utóbbiak – bár általában ingyenesek, vagy legalábbis jóval olcsóbbak – több esetben példamutatóak színvonalban, és a felhasználók valódi igényeinek kielégítésében.

A második rész a *tartalmi elemzés szempontjairól* szól, valamint áttekinti ennek a kritikai műfajnak a történetét és szakirodalmát. Az adatbázisok tartalmának vizsgálata tesztek, kísérletek és összehasonlító elemzések sorozatából áll. Ezt az idő- és költségigényes munkát nagymértékben felgyorsította és olcsóbbá tette a kifinomult keresőfelülettel ellátott CD-ROM verziók, illetve az utóbbi időben az átalánydíjas vagy ingyenes internetes változatok megjelenése. De egy alapos elemzés még így is komoly felkészültséget és türelmes vizsgálódást igényel; nem véletlen, hogy még a szaklapokban megjelenő recenziók írói és kiadói is gyakran megelégszenek a reklámszövegek és a felhasználói útmutatók alapján összeállított felületes ismertetőkkel.

A következő három fejezet az *adatbázisok által lefedett (szak)terület vizsgálatával* foglalkozik; vagyis hogy valójában milyen teljességgel és milyen mélységben dolgozzák fel azokat a témaköröket és kiadványfajtaikat, amelyeket meghirdetett céljuk szerint felvállaltak. Megtudhatjuk belőle, hogy a méret, vagyis a rekordok nagy száma önmagában

nem jelent értéket egy adatgyűjtemény esetében, még kevésbé garantálja annak használhatóságát, inkább csak marketingfogásként jó. (A könyv írásának idején kb. 11 ezer angol nyelvű weblapon szerepelt a „legnagyobb adatbázis” kifejezés.) Továbbá hogy a „nemzetközi” jelző esetleg csak annyit jelent, hogy az USA-ban megjelenő lapok mellett brit és a kanadai folyóiratokat is feldolgoznak a bibliográfia előállítói. Elgondolkodtató az is, amikor az *Information Science Abstracts*ben 2-3 találat van olyan témákban, amelyekről az ismertető azt állítja, hogy ezekben „az ISA az Ön elsődleges referenzsforrása”. Azon pedig már csak mosolyogni lehet, hogy a világ adatbázisait ismertető *World Databases*ben nincs nyilvántartva maga a „World Databases”. A nyomtatott folyóiratok megjelenési és terjesztési nehézségei az őket feldolgozó referáló szolgálatokra is hatással vannak: gyakran teljes lapszámok, vagy akár évfolyamok maradnak ki egészen alapvető kiadványok esetén is, s ez komolyan megkérdőjelezi azt a teljességet, ami ezektől a nagy múltú adatbázisoktól elvárható lenne. Pozitív fejlemény viszont, hogy a webes adatbázisok egyre nagyobb részét közvetlenül a fájlok előállítói szolgáltatják, ami javítja az átfutási időt a korábban egyeduralgó adatbázisközpontok vagy a CD-ROM-ok frissítési gyakoriságához képest.

A 7–9. fejezetek már az adatrekordok szintjén tárgyalják a *pontosság, a következetesség és a teljesség* kérdéseit. A szerző felhívja a figyelmet arra a gyakori hibára, hogy a számítógépről származó információknak nagyobb hitelt adunk. Pedig a bibliográfiai adatbázisok retrospektív rekordjait nyomtatott forrásokból rögzítették, továbbá az elmúlt 20-30 évben a nyomtatott kiadványok nagy része már számítógépen készült, vagyis a gépen levő adatok aligha lehetnek pontosabbak, mint a papíron megjelentek. A hibalehetőségeket tovább növelik a konverziók, amelyek az adatfájlok különböző online és offline változatainak előállításakor elkerülhetetlenek. A pontosságot viszont nagymértékben segíthetik az egységesített névalakok és tezaurusok, de ezeket még mindig nem minden adatbázis-előállító használja. Így fordulhatott például elő, hogy a *Mental Health Abstracts*ben az angol nyelv alá besorolt rekordoknál egyetlenegyszer sem sikerült helyesen leírni az „English” szót. A gépelési vagy karakterfelismerési hibáknál jóval veszélyesebbek azonban a ki nem töltött adatmezők, mert az ezekben való keresés vagy a velük való szűrés jelentős és nehezen észrevehető találatvesztéshez vezet. A hiányos adatelemek arányát a kiadók hajlamosak ugyanúgy eltitkolni a

felhasználók elől, mint a tematikus, földrajzi, nyelvi és időbeli lefedettségére vonatkozó információkat.

A következő két részben a szerző a hatékony vizsszakeresés szempontjából kulcsfontosságú *analitikus feltárás minőségével* kapcsolatos problémákat veszi sorra. A tárgyi indexelés, illetve a referátumok készítése, valamint ezek használhatósága, hasznossága persze meglehetősen szubjektív dolog, nehéz jól mérhető és összehasonlítható jellemzőket találni. De a tárgyszavazásnál mindenestre alapkövetelmény a tezauruszkészítés szabályainak megfelelő ellenőrzött szótár összeállítás és használata, valamint az adatbázis jellegéhez optimalizált mélységű, nem túl redundáns indexelés. A jól megírt kivonatok szintén sokat segítenek a releváns tételek megtalálásában, különösen a nagyon sok bibliográfiai rekordot, illetve a teljes szöveget is tartalmazó adatbázisoknál, és így nagymértékben növelhetik az értéküket. Sajnos jó néhány olyan adatbázis van, amelynek ugyan a nevében szerepel az „Abstracts” szó, de valójában csak a rekordok egy részénél vannak referátumok, és arra is van példa, hogy ezekre külön nem is lehet leszűkíteni a keresést.

Az utolsó fejezet az *árakkal* kapcsolatos megfontolásokat tartalmazza. A könyv általános tanulsága,

hogy az „azt kapod, amit fizetsz” mondás gyakran nem érvényes az adatbázisokra, különösen a web megjelenése óta. A korábban monopolhelyzetben lévő, esetenként drága információszolgáltatásoknak színvonalas, ügyesen felépített, másféle üzleti modelleken alapuló alternatívái jelentek meg, és megindult egy evolúciós verseny, melynek nyertesei remélhetően a felhasználók lesznek árban és minőségben egyaránt.

Az, hogy az ilyen kritikus szemléletű könyvek és folyóiratcikkek milyen hatással lehetnek erre a folyamatra, nehezen jósolható meg. (Jacsó Péter 1993-as javaslatát, mely az adatbázisok „felcímkésére” vonatkozott, hasonlóan az élelmiszerek összetételét mutató, ma már mindenhol kötelező tájékoztatókhoz, sajnos az adatbázisiparban mind eddig nem vezették be.) De kitartó munkával, az információközvetítők és az adatok végső felhasználóinak érdekeit képviselve, a minőség folyamatos hangsúlyozásával talán eljutunk majd az információs társadalomnak arra szintjére, amikor rangot jelent, ha egy adatbázis nyitólapján megjelenhet ez a mondat: „A könyvtárosok ajánlásával”.

Drótos László

(OSZK Magyar Elektronikus Könyvtár)

Könyvtár- és információtudományi folyóiratcikkek, felsőoktatás és nyelv

A könyvtár- és információtudományi folyóiratok különböző nyelveken jelennek meg, jóllehet az angol nyelvűek szolgálják leginkább a széles nemzetközi olvasóközönséget. A cikk egy kutatás eredményeit tartalmazza, különös tekintettel azokra az országokra, amelyeknek a nemzeti nyelvét viszonylag kevesen beszélik.

Módszerek

Az *Emerald* könyvtár- és információtudományi folyóiratait választották ki mintaként, mert az Emerald adja ki a legtöbb folyóiratot ezen a szakterületen, ugyanakkor a kiadó elkötelezett aziránt, hogy bátorítsa azokat a szerzőket, akiknek az anyanyelve nem angol. A 2004-ben is megjelenő, és a LISA adatbázisban is feldolgozott folyóiratokat jelölték ki, az Emerald 28 folyóirata közül 21 volt ilyen (5 nem volt kurrens, 2 pedig nem primer fo-

lyóirat). A LISA adatbázist azért választották, mert a benne szereplő adatokat könnyen lehet bibliográfiai szoftverbe letölteni (beleértve a kulcsszavakat és a referátumokat, szükség szerint lehetővé téve a mélyebb elemzést). A LISA adatbázisban való szereplés azt jelzi, hogy a folyóirat könnyen elérhető a könyvtár- és információtudományi hallgatók számára, mivel a LISA az egyik legfőbb referáló és indexelő szolgáltatás a szakterületen.

A kiválasztott 21 folyóirat valamennyi 2004-es cikkének adatait letöltötték az *Endnote* bibliográfiai szoftverbe. Ez a szoftver megkönnyíti az adatok kezelését, és a cikkek szakterületének elemzését. A mintában szereplő folyóiratok, zárójelben a belőlük 2004-ből felvett tételek számával a következők: *Aslib Proceedings* (30), *Bottom Line* (20), *Campus-Wide Information Systems* (22), *Collection Building* (22), *Electronic Library* (54), *Information Management and Computer Security* (30), *Information*