

Jacsó Péter

**Szöveges CD-ROM és webes adatbázisok
tartalmának elemzése**

Könyvtári Intézet

Budapest, 2005

Eredeti cím:
Content Evaluation of Textual CD-ROM and Web Databases
Libraries Unlimited – A Division of Greenwood Publ. Group., Inc
Englewood, Colorado 2001

A fordítás elkészítéséhez a kiadó
(www.greenwood.com) hozzájárult

Fordította:
Murányi Péter

Lektorálta:
Bánhegyi Zsolt

Elektronikus szerkesztés:
OSZK Magyar Elektronikus Könyvtár Osztály
Elter András

ISBN 963 201 617 3 (HTML)

ISBN 963 201 618 1 (PDF)



NEMZETI KULTURÁLIS ÖRÖKSÉG
MINISZTERIUMA

Készült
a Nemzeti Kulturális Örökség Minisztériuma -
Nemzeti Kulturális Alapprogram
támogatásával



TARTALOM

Köszönetnyilvánítás	4
Bevezetés	5
1. Adatbázis termékek, készítőik és kiadói	9
Adatfájlok és előállítói.....	9
Adatbázis-készítő és -kereső szoftverek és készítőik.....	11
Adatbázis-kiadók.....	12
Adatbázis szolgáltatási lehetőségek.....	13
2. Az adatbázisok tartalma elemzésének szempontjai	16
Az adatbázisok értékelésének folyamata.....	18
A tartalom elemzésének története és szakirodalma.....	20
3. Az adatbázis szakterületi hatóköre	24
4. Az adatbázis dimenziói	39
Méret és összetétel.....	39
Időbeliség.....	52
Mélység.....	58
Frissesség.....	71
5. Az adatbázisok által feltárt források	87
A források típusai.....	87
Folyóiratbázis.....	96
Földrajzi feltártság.....	108
A feldolgozott anyag nyelvi megoszlása.....	117
Esettanulmány - Egy adatbázis szakterületének és feltárt forrásainak (disz)harmóniája.....	124
6. A rekord tartalma	157
Bibliográfiai adatelemek.....	157
Értéknövelt adatelemek.....	161
7. Pontosság	172
8. A formátum és a tartalom következetessége	194
9. Teljesség	203
10. A tárgyi indexelés minősége	221
Fogalmak és szakirodalom.....	221
Az indexelési minőség kiértékelésének folyamata.....	228
Esettanulmány a tárgyi indexelés minőségének értékeléséről.....	233
11. A referátumok minősége	243
Fogalmak és irodalom.....	244
A referátumok értékelésének folyamata.....	248
Esettanulmány a referátumok minőségének értékeléséről.....	253
12. Az árral kapcsolatos megfontolások	269
Különbségek a tartalomban.....	269
A szoftverek különbségei.....	272
Az árak anomáliái.....	274
Online információs szolgáltatások versenye.....	275
Ábrák jegyzéke	279
Bibliográfia	289

Köszönetnyilvánítások

Hálás vagyok Carol Tenopirnak, aki nemcsak bátorított arra, hogy megírjam ezt a könyvet, hanem írásom szerkesztésére is vállalkozott. Sok fontos megjegyzéssel és javaslattal segített, nagyon értékes volt számomra támogatása és ösztönzése a könyv megírásának három éve alatt.

Rich Lane-nek meg kell köszönnöm, hogy kérlelhetetlenül átvezette a könyvet a szerkesztési folyamaton, Joan Torkildsonnak pedig azt, hogy a nyomtatásra előkészítette a kéziratot, s megóvott azoktól a pontatlanságoktól és következetlenségektől, amelyeket magam az adatbázisokban kritizálok. Rendkívül értékes volt Susan Johnson segítsége a tények ellenőrzésében. Ha maradt bármilyen hiba, az csak az én felelősségem.

Nagyra értékelem Tom Hogan, az *Information Today* elnöke támogatását, aki lehetővé tette számomra, hogy rendszeres rovatom lehessen újságjaiban és magazinjaiban, s 1990-től az adatbázisok minőségével foglalkozó workshopokat tarthassak a National Online Meetingek során. Nagyra tartom azért, hogy még akkor sem próbált cenzúrázni engem, amikor rendszeresen kritizáltam egy olyan adatbázist, amelyet cége megvásárolt. Ő nemcsak okos üzletember, hanem tiszteletreméltó úriember is.

Köszönetet mondok azoknak, akik az *Information Today*, *Computers in Libraries*, *Database*, *Online & CD-ROM Review* és az *Online Information Review* című lapoknál rovataim szerkesztői voltak a múltban vagy most is azok, s különösen azoknak, akik oroszlánrészt vállaltak legtöbb rovatom szerkesztésében: Paula Hane-nek, Marydee Ojalanak, David Hoffmannak és John Eichornnak. A Gale Groupnak azt kell megköszönnöm, hogy a digitális referenzs forrásokról szóló havi szemléimnek helyet adott a weben és szerkesztői támogatást is nyújtott hozzá. Ezek az írások segítettek abban, hogy folyamatosan vizsgáljam ezeket a fontos kérdéseket.

A könyv megírása közben nagy örömmel gondoltam legjobb diákjaimra, különösen azokra, akik hozzám hasonlóan érdeklődtek az adatbázisok elemzése iránt: Lois Kiehl, Susan Johnson, Rich Gazan, Pam Cahn, Jennifer Winter, Marcia Kemble, Kimberley Scott, Justin Goo, Beth Tillinghast, Jennifer Jackson, Jessica Gehle, Nancy Kleban, Ellen Peterson, Evangeline Alexander és Marilyn Reppun. Állandó készenlétre ösztönöztek okos és provokatív kérdéseikkel és megjegyzéseikkel, ösztönzően aktívak voltak az adatbázisok minőségének és értékelésének kérdéseivel foglalkozó kurzusaim, workshopjaim és szemináriumaim során.

Elismerés illeti Don Hawkinst, az Information Science Abstracts főszerkesztőjét, korábbi vezető technikai szerkesztőjét. Számos szerkesztőségi cikke az ISA weboldalán és hagyományos kiadványokban megjelenő, egy adott adatbázis értelmezéséről és félreértelmezéséről szóló írásai ösztönöztek a könyv megírására. Az, ahogyan egy fájl készítőjeként látja a minőséggel kapcsolatos kérdéseket, gyakran különbözik az én, a használók oldaláról megfogalmazott véleményemtől, s ezzel ki is egészíti azt.

A köszönetmondás nagyon kevés, amikor feleségem, Tiszai Judit rendkívüli támogatását kell elismernem. Nagyszerű partner volt a projekt minden fázisában, az online és offline kutatásban, az eredmények elemzésén alapuló illusztrációk elkészítésében, és a kézirat kezelésének és a bibliográfia összeállításának prózai feladataiban. Még sokkal fontosabb, hogy nagyszerű partnerem maradt életünk munkán kívüli részében is.

Bevezetés

Az adatbázis fogalma általánosan elterjedté vált az elmúlt évtizedben. Az Amerikai Egyesült Államokban és a többi fejlett országban háziasszonyok és otthonülők, pénztárosok és autószerelők, zenetanárok és pszichológusok, a középiskolából kibukottak és a hajléktalanok is használnak adatbázisokat. Minden egyetemi és főiskolai hallgató és a legtöbb általános iskolai és középiskolai tanuló több adatbázist használ tanulmányai során. Sok vállalat alkalmazottaitól várják el, hogy ismerjenek néhány adatbázist. Másoktól azt követelik meg, hogy új adatbázisokat vizsgáljanak meg, amelyekkel saját és munkatársaik munkája hatékonyabbá válhat.

Mindezt természetesen az internet tette lehetővé, a weben található sok jó minőségű, ingyenes adatbázis. Gomba módra szaporodnak a kiváló és gyenge adatbázisok a weben, mert ott viszonylag könnyen közzé lehet őket tenni. Az arányokat tekintve épp annyi gyenge termék van a hagyományos fizetős online és CD-ROM adatbázisok között, mint a weben születettek között. (Itt kereshető weben született adatbázisok százaira utalok, nem egyszerűen web-helyekre.)

Rendkívüli mértékben növekszik a száma azoknak a végfelhasználóknak (könyvtári olvasók, kutatók, otthoni használók), akiknek ismerniük kell az általuk használt adatbázisok korlátait, erősségeit és gyengeségeit. 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 (vagy mindezeket együtt végzik).

A könyvtárak maradnak az információkkal kapcsolatos útmutatás központjai akkor is, ha nem ők maguk az információgyűjtemények központjai. A könyvtárosok és információs szakemberek továbbra is a legmegbízhatóbb, kompetens források a végfelhasználók információs igényeinek kielégítésében, legyen szó akár fizetős, akár ingyenes szolgáltatásokról.

Mostanában mindenféle magazinban találhatunk adatbázisokról „kritikákat”, a *Better Homes and Garden*-stől kezdve a *Horses and Houndson* keresztül a *PC Magazine*-ig. Ezek gyakran csupán a kiadók reklámjainak kissé átfogalmazott ismételései. A web sok használója érez sürgetést arra, hogy összeállítsa és közzétegye a legjobb adatbázisok listáját, anélkül, hogy tisztában lenne az adatbázis-elemzés alapjaival, bármit is tudna az adatbázisok elemzésének legalapvetőbb kérdéseiről. Egyes könyvtárosok és információs szakemberek, akik korábban soha nem írtak kritikát referenzforrásokról, most kényszerítve érzik magukat arra, hogy megosszák nem túl éleselmjű véleményüket a weben született adatbázisokról.

Nem véletlen, hogy a legjobb kritikák azoktól származnak, akik a szakmát még a nyomtatott világban sajátították el. Jim Rettig, Cheryl LaGuardia, Mary Ellen Bates, Reva Basch, Marydee Ojala, Barbara Quint, Ken Kister és Ken Black a legjobb kritikusok közé tartozik. Megvan annak az oka, hogy miért a könyvtár- és információtudományi folyóiratok, mint a *Database*, az *Information Today*, a *Choice*, a *Reference Book Bulletin* és a *Library Journal* az adatbázis kritikák legjobb forrásai. Nem meglepő, hogy könyvtárosok segítettek létrehozni a leghasznosabb webes címjegyzékeket és webkalauzokat, mint a SignPost, a Librarian's Index és az Internet, a BUBL és az INFOMINE, amelyek kiemelkednek az útkeresők tömegéből.

Ez a könyv, amely könyvtárosoknak, információs szakembereknek és adatbázisok kritikusainak szól rendszerező módon, összefoglalja az adatbázisok tartalmának elemzésére használható szempontokat. Az adatbázisok elérésére szolgáló szoftverek és hardverek szempontjaival nem foglalkozunk, bár feltétlenül fontos szerepet játszanak abban, hogy mennyire hatékonyan lehet megtalálni az információt.

Az adatbázisok tartalmának jellegzetességeit írjuk le és magyarázzuk meg. Pozitív és negatív példákat használunk az adatbázisok legfőbb jellegzetességeinek illusztrálására. Az információs szakemberek által használt adatbázisok kiválasztása az illusztrálásra magától értetődő választás volt. A példaként használt többi adatbázis közül sok széles körben népszerű, és az adott szakterületre vonatkozó képzettséggel nem rendelkező emberek is megérthetik. Egyes adatbázisokat azért választottunk ki, mert rendkívül jól illusztrálják a hiányosságokat. Ez a mintaként kiválasztott csoport adatbázisok különböző kategóriáit reprezentálja (címjegyzékek, indexelő, referáló és teljes szövegű adatbázisok). A leggyakrabban az alábbi adatbázisokat használtuk. Másokat csak egyszer vagy kétszer említünk meg, amikor egy meghatározott szempontot nagyon jól illusztrálnak.

Adatbázis	Tartalomszolgáltató
Library and Information Science Abstracts (LISA)	G. K. Saur
Information Science Abstracts (ISA)	Information Today
Library Literature (LibLit)	H. W. Wilson
Internet & Personal Computing Abstracts (IPCA)	Information Today
Social Science Search (Social SciSearch)	Institute for Scientific Information
Psychological Abstracts (PsycINFO)	American Psychological Association
Mental Health Abstracts (MHA)	IFI/Plenum
Ulrich's International Periodicals Directory (Ulrich's)	R. R. Bowker
The Serials Directory (TSD)	EBSCO
ISSN Online International Serials	International Serials Data System
Gale's Directory of Databases	Gale Corporation
World Databases	G. K. Saur
Bowker's Complete Video Directory (BCVD)	R. R. Bowker

A cégek és az adatbázisok neve folyamatosan változik. A könyv megírásának idején a UMI elhagyta nevének feloldását (University Microfilm International), felvette anyacégének nevét, egy minősítő kiegészítéssel (Bell & Howell Learning and Information), azután pedig, amikor a könyv nyomdába ment, hatékony információkereső rendszerének neve után ProQuestre változtatta nevét.

Az egyesülések, megvásárlások és közös kiadások szintén fejfájást okozhatnak. Még könyvem megírása előtt az R. R. Bowker céget az Elsevier megvásárolta. Saját neve alatt számos gyorstájékoztató adatbázisa volt, azután a G. K. Saurral közösen adott ki adatbázisokat Bowker-Saur név alatt, majd épp könyvem befejezésekor szétvált a német cégtől, s amikor ön olvassa a könyvet, elképzelhető, hogy az Elsevier már tovább is adott a vállalatnak, vagy legalábbis néhány digitális kincsén.

Néhány fájlkészítő, adatbázis és cég neve meg fog változni addigra, mire ez a könyv megjelenik. Az alábbiakban összefoglaljuk a legfontosabb változásokat, amelyekről tudomást szereztünk:

Régi név	Új név
University Microfilm International (UMI)	Bell & Howell Information and Learning*
Information Access Company (IAC)	Gale Group (része)
Library Literature & Information Science Full Text (LibLit)	Library Literature

* 2001 júniusában újra megváltoztatta nevét, Proquestre, népszerű szoftverének neve alapján.

Az információs rendszerek egy adatbázist különböző adatbázis (fájl) szám vagy rövidítés alapján azonosítják. Ilyenkor magyarázó leírást használunk, például „az Ulrich’s DIALOG-os változata” vagy az „Ulrich’s Bowker CD-ROM-os változata”, ahelyett, hogy 480-as fájlnek vagy Ulrich’s Plus on Discnek neveznénk őket. Más esetekben, amikor egy adatbázisnak csak egy változatát használjuk illusztrálásra az egész könyvben, az a név szerepel, ahogy azt az online rendszerben használják, mint például Social SciSearch az ISI hivatkozási adatbázisának DIALOG-os változatában. Az ábrákon helytakarékosági okokból szerepelhet rövidebb változat (mint például SocSci), de a kapcsolódó magyarázó szöveg az ajánlott hosszabb formát fogja használni. A keresési képernyőképek esetében az adatbázis szolgáltató által használt név vagy kód jelenik meg, mint például File 7 a Social SciSearch esetében. Ebben az esetben is a környező szöveg segít pontosan megadni, miről van szó.

Az adatbázisok tartalmának jellegzetességeit vizsgáló tesztek közül sok szoftverfüggetlen. A DIALOG nyújtja a legtöbb és messze a legjobb lehetőséget a tesztelésre, ezért szerepel legtöbbször a könyvben. A könyvtár- és információtudományi tanszékek is ezt a szolgáltatást veszik igénybe a legszélesebb körben. Vannak azonban példák az adatbázisok más változataira is, beleértve az Ovid, SilverPlatter, Bell & Howell Information and Learning, H. W. Wilson online és CD-ROM keresőszoftverét, és számos másikat (többségében a webre tervezett változatokat), amelyeket egy adott adatbázis számára egyedileg fejlesztett ki az eredeti tartalomszolgáltató, a Barnes & Noble, az Amazon.com vagy az All Movie Guide. (A szövegben az ingyenes webes adatbázisoknak csak az URL-jét adjuk meg, s csak akkor, ha a specifikus példa szükségessé teheti az olvasó számára, hogy megnézze azt a webhelyet vagy annak meghatározott oldalát. A webhelyeket könnyű megtalálni itt szereplő nevük alapján, különböző webes címjegyzékeket használva.)

Nem biztos, hogy az olvasónak van ideje, lehetősége vagy érdeklődése, hogy valamennyi, a könyvben szereplő szempont szerint értékelje az adatbázisokat. Egyes szempontok esetleg nem alkalmazhatók arra az adatbázisra, amely iránt érdeklődik. Ennek ellenére azonban a szempontok segítséget jelenthetnek abban, hogy összeállítsák a kérdések listáját, amelyet feltehetnek az adatbázisok kiadóinak, különösen fizetős szolgáltatások esetében, amikor joguk lehet arra, hogy részletes választ kapjanak olyan kérdésekre, mint az, hogy miként dolgozzák fel a magfolyóiratokat, milyen időkülönbség van a primer dokumentumok megjelenése és rekordjaiknak az adatbázisban való bekerülése között, a rekordok mekkora részében szerepelnek meghatározott adatelemek (például dokumentumtípus, az LC osztályozási jelzete).

Az adatbázisok tartalma folyamatosan változik, legtöbbször jobb lesz, bár vannak kivételek. Az Information Science Abstractset 1998-ban szerezte meg az Information Today. Bár bizonyos szempontokból fejlődött, minőségileg nem közelítette meg - ahogy várható lett volna - a korábban Microcomputer Abstracts néven ismert Internet & Personal Computing Abstracts (IPCA) adatbázist, amely az Information Today zászlóshajója.

Bár az 1998 után bekerült rekordok között vannak jobb minőségűek (bizonyosan azok, amelyeket az IPCA „adott tovább” nekik), az adatbázisba csak feleannyi rekord került be, mint egy évvel korábban, kihagyott alapvető információtudományi folyóiratokat vagy a minimumra

csökkentette feldolgozottságukat, ahogyan azt a 4. és 5. fejezetek illusztrálják az adatbázisok dimenzióiról és az adatbázisokban feltárt forrásokról. Lehetséges, hogy más adatbázis-kiadók időközben megoldották a könyvben tárgyalt problémák némelyikét. Értékelése során önnek az adatbázisok akkori változatát kell vizsgálnia. Ha csak másfajta utalás nem történik rá, a teszteket a könyv számára 1999 júliusa és 2000 novembere között végeztük, egyeseket 2001-ben újra elvégeztünk a könyv korrektúrájának olvasásakor.

Konzultációs szolgáltatásokat végeztem a könyvben említett cégek közül többnek a számára, az *Information Today* és a Gale Group állandó munkatársa vagyok, amelyek ugyanakkor fájlok készítői, egyes esetekben adatbázisok kiadói. Mindent megtettem, hogy ez a tény ne befolyásolja adatbázisaik minőségéről alkotott ítéletemet.

1. Adatbázis termékek, készítőik és kiadók

Érdemes ezt a könyvet azzal kezdeni, hogy megmagyarázunk néhány olyan fogalmat, amely többször előfordul majd a szövegben. Egyesek általánosan használatosak, de az olvasók esetleg különbözően értelmezhetik őket. Másoknak speciális jelentése lehet, definíciót és magyarázatot kívánnak a kétértelműség elkerülése érdekében. Az is fontos, hogy világosan lássuk a különbséget az adatbázisok kiadásának módszerei és médiumai között.

Egy *adatbázis* intellektuális és technikai műveletek láncolatának eredménye. Különböző szervezetek és egyének kapcsolódnak be ezekbe a műveletekbe. Esetenként az adatbázis egy lelkes és hozzáértő személy munkájának eredménye (ez volt a helyzet a kiváló *Encyclopaedia of the Orient* esetében), máskor emberek kis csoportjéé (egy remekmű, az *Internet Movie Database* a példa erre). Az adatbázisokat gyakran nagyobb szervezetek vásárolják meg, amelyek adaptálják, és saját szolgáltatásukba integrálják azokat. Ennek a könyvnek a szempontjából az adatbázisokhoz kapcsolódó legalapvetőbb fogalmak meghatározásának legjobb módja az, ha végigkövetjük a műveleteket és bemutatjuk a szereplőket és a termékeket a kezdetektől a közbeeső állomásokon keresztül a végső fázisig.

A folyamat az *adatfájl*-nak nevezett, géppel olvasható fájl előállításával indul. Évtizedeken keresztül az adatfájl volt az a forrás, amely lehetővé tette, hogy a számítógép előállítsa a referáló és indexelő szolgáltatások, teljes szövegű folyóiratok és újságok, szótárak és enciklopédiák nyomtatott változatát. Az adatfájl általában mágnesszalagon van. Ahhoz, hogy adatbázist készítsenek belőle, indexfájlokat kell létrehozni, az adatfájlt át kell szerkeszteni és keresőszoftverrel kell kiegészíteni. Tipikus esetben az adatfájl előállító az adatfájl tartalmát harmadik fél részére bérbe adják, hogy az adatbázist készítsen az adatfájl tartalmából; innen származik a gyakran használt *tartalomszolgáltató* (content provider) elnevezés. A szakirodalomban az adatfájl előállító pedig *adatbázis-készítő*ként (database producer) is ismertek.

A licenc birtokosai feldolgozzák az adatfájlokat és CD-ROM-on vagy online módon publikálják: ők lesznek az adatbázis kiadói. Az adatbázis-kiadók kifejleszthetik saját indexelő és kereső szoftverüket, vagy licencelhetik azt egy szoftverfejlesztőtől.

Egyes adatbázis-kiadók (ismertek *online vendorként* vagy *online szolgáltatóként* is), mint például a DIALOG, tartalomszolgáltatók százaitól vesznek licencbe tartalmat. Egyes adatfájlokat, mint például a *Pediatric Renal Diseases*-t, csak egyetlen adatbázis-kiadónak adnak át; míg másokat, mint például a *MEDLINE*-t, adatbázis-kiadók tucatjainak. Egyre elterjedtebb gyakorlat, hogy a fájlok előállító maguk is publikálják saját adatfájljukat vagy adatfájljaikat. Az H. W. Wilson és a Bell & Howell például széles körben bérbe adják adatfájljaikat másoknak, de ugyanakkor saját webhelyükön is közzéteszik azokat.

Adatfájlok és előállítóik

Az adatfájlok előállító válogatják, gyűjtik, rendszerezik és rögzítik a könyvekkel, cikkekkel, konferencia-előadásokkal, disszertációkkal, cégekkel, emberekkel, szolgáltatásokkal stb. kapcsolatos információkat. Az adatfájlok előállító a rögzített adatokat valamilyen szerkezetbe rendezik s általában ki is egészítik olyan információkkal, amelyek nem érhetőek el közvetlenül a forrásból. Ez az információ növeli az adatfájl értékét olyan elemek hozzáadásával, mint osztályozási jelzetek, tárgyszavak, referátumok, vagy ezek kombinációja. Mintaként használt csoportunkban az Amerikai Pszichológiai Társaság (American Psychological Association, APA) csupán adatfájl előállító volt, amely harmadik fél részére tartalmat adott át,

amelyből klasszikus referáló és indexelő forrásuk, a *Psychological Abstracts* különböző változatai kiadásra kerültek. (1999-ben az APA is adatbázis-kiadó lett, mivel elkészítette saját adatbázis változatát, a PsycINFO Directet, amelynek van egy ingyenes részhalmaza demonstrációs célokra a <http://www.psycinfo.com/demo/> címen.) A *tartalomszolgáltató* szűkebb fogalom, mint a *fájl-előállító*, mert lehet, hogy az előbbi nem készít számítógéppel olvasható fájlt, csak licenceli a tartalmat, például egy nyomtatott útikalauzt, változatlan formában, egy harmadik félnek.

Másik oldalról egyes adatfájl előállítók megszerzik a licencet ahhoz, hogy mások létező nyomtatott forrásait feldolgozzák és adatfájllá konvertálják. Például a Microsoft megszerezte a *Zagat Restaurant Guide* tartalmát, s ebből előállított egy adatbázist saját Sidewalk web-oldalán, amelyet később eladott a Ticketmasternek, amely Citysearchcsé alakította át (<http://www.citysearch.com/?brand=sidewalk>). A Zagat volt a tartalomszolgáltató vagy fájl-előállító (ha fájlként adta át a tartalmat a Microsoftnak), a Microsoft az adatbázis kiadója. A Zagat később adatbázis-kiadóvá is vált, miután saját tartalmát megjelentette a weben, a <http://www.zagat.com> címen.

Könyvünkben az adatfájl mindig egy géppel olvasható formában levő információgyűjteményre utal. Az adatfájl lehet tisztán szöveges (például egy bibliográfiai adatfájl), szöveges-numerikus (például cégek címjegyzéke, amely leíró és pénzügyi információkat is tartalmaz), döntően numerikus (például egy statisztikai adatfájl). A szövegen kívül az adatfájl tartalmazhat számítógépes grafikát, álló- és mozgóképeket, hangokat és videoklipeket, ez jellemzi az összes, a nagyközönségnek szóló multimédia enciklopédiát. Az adatfájl lehet eredeti dokumentumok hasonmásainak, faksimiléinek gyűjteménye, mint például az adóhivatal, az Internal Revenue Service (IRS) formanyomtatványai. A hasonmás tartalmának értékét nagy mértékben növelhetik az adatbázis előállításának folyamata során azzal, hogy indextételeket készítenek az egész szöveghez, mint például az International Women's Periodicals fájl vagy a teljes ProQuest adatbáziscsalád esetében.

Az adatfájlok előállítói felajánlhatják licencelésre a teljes tartalmat vagy annak egy részhalmazát. Például az Amerikai Egyesült Államok Nemzeti Orvosi Könyvtára (National Library of Medicine, NLM) a MEDLINE fájl különböző részhalmazait ajánlja fel licencelésre. A részhalmazok létrehozhatók nyelv, időszak vagy a források típusa alapján. Például a MEDLINE CORE nevű adatbázis-változat csak a legalapvetőbb orvosi folyóiratok Brandon-Hill listáján szereplő folyóiratokból származó rekordokat tartalmazza. A lista a <http://www.nlm.nih.gov/psr/outreach/branhill.html> címen volt elérhető. 2004-ben abbahagyták, utódja a Doody's Core Titles in the Health Sciences (<http://www.doody.com/dct/>). Mások csak a MEDLINE utolsó néhány évét bérlik, és nem az egész adatbázist, amely 1966-ig megy vissza.

Egy másik példában az Education Resources Information Center (ERIC) adatfájlt egyesek teljes egészében licencelik, mind a folyóiratok, mind a kutatási jelentések részhalmazát, az ERIC Digestek teljes szövegével együtt. Bérlezhető például csak a folyóiratok részhalmaza. A Psychological Abstracts adatfile ClinPsych részhalmaza csak azoknak a cikkeknek a rekordjait tartalmazza, amelyek a klinikai gyakorlatot tárgyalják. A Bowker cég mind a szemlékkel, mind azok nélkül licenbe adja Books in Print rekordjait. Egy harmadik változat csak olyan könyvek rekordjait tartalmazza, amelyek már nem kaphatók. Elképzelhető, hogy az adatbázisok neve nem tükrözi ezeket a különbségeket. A DIALOG például a Books in Print nevet használja, bár az ő változatuk csaknem egymillió rekordot tartalmaz már nem kapható (out of print) könyvekről, s könyvkritikák tízezreit is.

Egyre több adatfájl előállító választja azt a lehetőséget, hogy maga is publikálja saját tartalmát a weben. Ezek a változatok gyakran a legjobbak közé tartoznak, mert ők ismerik legjobban a tartalmat, és az annak megfelelő szoftverrel a lehető legtöbbet hozzák ki a fájlból. Az online szolgáltatóknak adatbázisaik tucatjainál vagy alkalmanként akár százainál is használt egységes interfészre, keresési űrlapra és output formátumra való tekintettel kell létrehozniuk újabb adatbázisaikat. Azok a fájlleállítók, akik csak néhány adatbázist tesznek közzé a weben, tartalmukhoz igazíthatják azokat, annak kényszere nélkül, hogy meg kelljen felelni a meglévő tervezési keretnek, ahogy az a nagy online információszolgáltatók esetében történik.

Például a Responsive Database Services (RSD) számos adatbázis-kiadónak és információszolgáltatónak licencelte négy adatfájlját, de a sajátjuk a legjobb változat. Az ERIC Assessment and Evaluation Clearinghouse-ának (Mérési és Értékelési Ügynökségének, címe <http://www.ericae.net>) ingyenes ERIC változata összehasonlíthatatlanul jobb, mint a térítéses CARL változat. (A címen most csak az az információ szerepel, hogy az USA Oktatási Minisztériuma 2003. december 13-ával bezárta az összes ERIC Clearinghouse-t, így az ERICAE címe sem működik már, egyes szolgáltatásai más címen megtalálhatók (így a SearchERIC a <http://SearchERIC.org> címen, más szolgáltatásokat teljesen megszüntettek.) A Restaurant Guide adatbázis esetében a Zagat saját változata (<http://www.zagat.com>) mérőfeldekkal jobb az interfész, keresési és output lehetőségek szempontjából, mint amit a valamikori Microsoft Sidewalk változat (<http://national.citysearch.com/?brand=sidewalk>) ajánlott, annak ellenére, hogy a Microsoft igazán tudja, hogy kell az adatbázisokat megtervezni, s kétségtelenül megvan ehhez a szoftvere és személyzete egyaránt.

Adatbázis-készítő és -kereső szoftverek és készítőik

Ahhoz, hogy egy adatfájlból adatbázis készüljön, sokféle szoftverre van szükség. Két alapvető, egymástól jól elkülöníthető szoftver összetevőt használnak ennek a folyamatnak a során. Az egyik összetevő az adatbázis-építő eszköz, a *buildware*, amelyet az adatfájl adatai elemének átszerkesztésére, konvertálására és a különböző indexek előállítására használnak. Ha egyszer az adatfájlt feldolgozták a buildware-rel, az eredmény a *dataware*, az eredeti tartalom módosított változata. Az adatfájllal ellentétben, amely szabványosított fájlformátum (megcímkézett, vesszőkkel és idézőjelekkel körülhatárolt, MARC [Machine Readable Catalog, format, géppel olvasható katalógus formátum], stb.), amelyet számos különböző program képes feldolgozni, a dataware olyan kizárólagos formátum, amelyet csak a *searchware* komponens képes olvasni, nem használhatja más szoftver. A searchware tartalmazza az interfészt, a navigációs rendszert, a súgó fájlokat, a parancs értelmezőprogramját és a keresőmotort. A *dataware* és a *searchware* együtt alkotja az adatbázist. Ezek integrálása hozza létre az adatbázist az adatfájlból a használók számára, de van egy harmadik összetevő is: a *buildware*.

Azoknak, akik keresnek az adatbázisban, nem kell tudniuk a *buildware*-ről, az adatbázis-előállító szoftverről. Ennek azonban fontos hatása van arra, hogyan lehet használni az adatbázist. A buildware határozza meg, melyik adatalemeket választják ki az adatfájlból (melyeket teszik böngészhetővé), hogyan hozzák létre az indexeket, hogyan szegmentálják a rekordot, melyik adatalemek kereshetők stb.

A *buildware*-t és a *searchware*-t általában független szoftverfejlesztők fejlesztik ki, s tőlük licencelik őket azok, akik adatbázisokat akarnak készíteni az adatfájlokból. Ez különösen érvényes a CD-ROM adatbázisokra és azokra a webes adatbázisokra, amelyeket maguk a fájlok előállítói tesznek közzé. A CD-ROM adatbázisok esetében például az Online Computer Systems *OptiWare* szoftvere nagyon népszerű. Ezt licencelte többek között az R. R. Bowker teljes adatbázis-családja számára, sok nemzeti könyvtár nemzeti katalógusa CD-ROM válto-

zata számára és a Public Affairs Information Services PAIS adatbázisa számára. A Dataware Technologies *CD-Answer* szoftverét használják legszélesebb körben CD-ROM adatbázisokhoz.

Az adatbázisok weben való publikálására számos lehetőség van (Jacsó 1998d). Ezek sokszor létező adatbázis-kezelők továbbfejlesztett változatai. Az egyik legolcsóbb és legegyszerűbb webes adatbázis-kiadó szoftver a *Reference WebPoster* a Research Information Services cégtől. A *Reference WebPoster* bármely, a ProCite, Reference Manager és EndNote által készített adatbázist képes jól kereshető webes adatbázissá átalakítani. A jó minőségű szoftverek között megemlíthető az Inmagic *Web Publisher* szoftvere, amely a DB/Textworks által előállított adatbázisokat alakítja át kifinomult webes adatbázisokká. A CGI cég olcsó, teljes szoftvert ajánl, a *Biblioscape*-et, amely nem pusztán egy kiegészítő program. Teljes szolgáltatást végez, nemcsak arra képes, hogy meglevő adatfájlból adatbázist hozzon létre, de saját maga is képes adatfájlt készíteni. Ez Paul Chen nagy hozzáértéssel és lelkesedéssel tervezett munkája, aki olyan gyakorisággal aktualizálja és fejleszti szoftverét, hogy azt a legfőbb szoftverfejlesztő cégek is megirigyelhetik. Ez a fajta egyéni teljesítmény a web által kínált lehetőségeknek köszönhető.

Az online információszolgáltatók általában maguk fejlesztették ki és tartják karban buildware-üket és searchware-üket. A DIALOG, LEXIS-NEXIS, Bell & Howell, Ovid) nem adja át másoknak, bár egyesek a keresőszoftvert, a searchware összetevőt átadhatják egyetemeknek és cégeknek a dataware helyi felmásolására. A SilverPlatter esetében elég egyedi a forgatókönyv, mivel WebSPIRS szoftverét átadta az H. W. Wilson cégnek, amely egyik vetélytársa a webes adatbázis-kiadói küzdőtéren. Az OCLC elegáns SiteSearch buildware és searchware szoftverét helyi megvalósításra bármely adatfájl számára licenceli.

Adatbázis-kiadók

Az adatbázis-kiadók azok, amelyek integrálják a dataware-t és a searchware-t, továbbá dokumentációval és segítséggel szolgálják a használókat. A legtöbb adatbázis kiadó (pl. a DIALOG, Ovid, SilverPlatter) nem állít elő saját tartalmat (kivéve a segédfájlokat, mint a DIALOG Finder adatbázisok); helyette másoktól licencelik a tartalmat. Más adatbázis-kiadók csak saját tartalmukat publikálják. Ez a helyzet az H. W. Wilson cég esetében, és ez jellemezte a Bell & Howellt 1998-ig, amíg ez utóbbi kezdett licencelni néhány adatfájlt az H. W. Wilsontól. Egyes kiadók keverik saját tartalmukat másokéval. Az EBSCO Publishing és az OCLC a legjobb példák erre a változatra. Az EBSCO-nak megvannak a saját referáló és indexelő, teljes szövegű és címjegyzék adatbázisai, amelyeket mások által szolgáltatott tartalommal együtt tesz közzé. Az OCLC többek között saját közös katalógusát publikálja, továbbá sok olyan adatbázist, amelynek az adatfájlját másoktól licenceli.

Egyre több adatfájl előállító, tartalomszolgáltató vált maga is kiadóvá, legalábbis a webes platformon. Továbbra is licencelik tartalmukat adatbázis-kiadóknak, de maguk is a webre teszik tartalmukat. Ennek a megközelítésnek vannak hagyományai. A Bell & Howell, az Information Access Company (most a Gale Group része), az H. W. Wilson és a Gale széles körben népszerű tartalomszolgáltató és adatbázis-kiadó egyaránt. Ebben az esetben az arány az, ami új. Azok, akik csak egyetlen vagy néhány fájl adtak ki, most meglovagolták a konjunktúrát. A Sociological Abstracts az egyik példa erre. Referáló és indexelő fájljukat legalább egy tucat online és CD-ROM kiadónak adták licenche, és 1998-ban a cég (amely most a Cambridge Scientific Abstracts része) megjelentette saját változatát a weben. Elég szokatlan a fájl-előállítók körében, hogy visszavonják fájljukat az adatbázis kiadóktól akkor, amikor saját változatukat megindítják. Az még szokatlanabb, hogy ezt úgy teszik, hogy csak

ingyenes változatot szolgáltatnak, az American Society of Civil Engineers azonban megtette ezt ASCE adatbázisával (Jacsó 1997d).

Az 1990-es évek elejéig a tartalom átadása CD-ROM-os adatbázisok számára kizárólagos volt. A SilverPlatternek volt a legnagyobb használói bázisa amiatt, mert kizárólagos hozzáférést biztosított néhány nagyon népszerű bibliográfiai adatbázishoz, mint például a PsycLIT és a Sociofile. Amikor az exkluzív szerződés lejárt, a CD-ROM-ok licencelése is követte a nem exkluzív licencelési modell mintáját. 2000-re a licencelési stratégiák a lehető legliberálisabbakká váltak, használói piac alakult ki. A Bell & Howell és az H. W. Wilson cégek a legjobb példák a keresztsben való licencelésre. A Bell & Howell kiadja zászólshajói, az ABI/INFORM, a Dissertation Abstracts, a Periodical Abstracts és a Magazine Express adatbázisok online és/vagy CD-ROM változatait, és ugyanakkor licenche adja sok adatbázis kiadónak is. Ők is licencelik az H. W. Wilson cégtől annak néhány teljes szövegű adatbázisát, közreadva őket webhelyükön és/vagy CD-ROM-on. Referáló és indexelő, teljes szövegű, képi és teljes szövegű és képi változatait is felkínálják fájljaiknak CD-ROM-on és webhelyükön is, saját szoftverüket, a ProQuestet használva. Ezen felül rendelkeznek az INSPEC CD-ROM változatával, valamint az H. W. Wilson cég néhány referáló és indexelő fájljának CD-ROM és online változatával, köztük az Applied Science & Technology adatbázissal.

Az H. W. Wilson cég adatfájljait sok jelentős online szolgáltatónak adja át, köztük van az Ovid, a DIALOG, az OCLC, a SilverPlatter és a Bell & Howell. Az utóbbi kettő ugyancsak közread H. W. Wilson fájlokat CD-ROM-on. Az H. W. Wilson cég rendelkezik összes indexelő, referáló és teljes szövegű adatbázisának CD-ROM és online változatával. A CD-ROM szoftver saját fejlesztésű, a webes szoftver azonban a WebSPIRS szoftver adaptált változata. A fájlok előállítói, szoftverfejlesztők és adatbázis-kiadók közti poligám viszony nagyon előnyös az ügyfelek számára, mivel választhatnak bármilyen formát, amit kedvelnek, olyan szoftverrel, amelyet a legjobban ismernek végfelhasználóik.

Adatbázis szolgáltatási lehetőségek

A cégeknek és a felsőoktatási intézményeknek számos választási lehetőségük van ahhoz, hogy adatbázisaikat elérhetővé tegyék alkalmazottaik és diákjaik számára. Arra is megvan a lehetőségük, hogy teszteljék a lehetőségek hatékonyságát, mert a licenc megállapodások általában egy vagy két évre szólnak. Az adatbázisokat CD-ROM-on vagy online szolgáltatathatják.

Hagyományos adatbázisok

Az 1980-as évek végén, az 1990-es évek közepén a CD-ROM nagyon népszerű alternatíva volt. Áttörést jelentett sok könyvtár számára, amely nem engedhette meg magának (s nem is volt képes felügyelni) a hagyományos, a használat mértékén alapuló online szolgáltatások igénybevételét, amelyek díjat fizettek a kapcsolati időért és a megjelenített vagy kinyomtatott tételekért. Ezek az online szolgáltatások kapcsolt vonalas telefon összeköttetésen vagy bérelt vonalakon keresztül voltak elérhetőek, amelyek növelték a keresés költségeit. Hogy ez ne nőjön még jobban, az adatbázis szolgáltatásokat képzett közvetítők használták, nem a végfelhasználók.

A CD-ROM alternatíva egy vagy több számítógépen rögzített előfizetési díj mellett biztosítottak hozzáférést az adatbázisokhoz. Több használó párhuzamos hozzáféréséhez az adatbázis hálózati változatára volt szükség - általában jelentős többletköltségért. Az ilyen hozzáférés gyakran a válaszadási idő nagy mértékű növekedését eredményezte, s szabályrendszerre volt szükség az adatbázisok használati idejének beosztására. A hálózat működtetése jelentős telepítési és karbantartási erőfeszítéseket követelt meg a könyvtártól. A CD-ROM adatbázisok

használata adott helyre korlátozódott. A viszonylag lassú hozzáférés eleve kizárta a távoli kapcsolódást az ilyen adatbázisokhoz, bár voltak sporadikus próbálkozások arra, hogy a CD-ROM adatbázisokat távoli számítógépekről használják. Ezek között a korlátok között hatékonyságuk nem csökkent olyan nagy adatbázisoknál, mint a ProQuest több száz CD-n levő adatbázisai, amelyek dokumentumok millióinak teljes szövegét tartalmazzák, sok közülük szöveg és grafika és képi formában.

A merevlemezek kapacitásának drámai növekedése és az ilyen tárolási eszközök árának hasonló mértékű csökkenése azonban a CD-ROM változat mellett az online változatoknak is utat nyitott. Ezt a folyamatot felgyorsította a gyors internet kapcsolat széleskörű elérhetősége. Az online modellhez való visszatérés kapcsán a legtöbb információszolgáltató elkezdett fix összegű előfizetési díjat kínálni. A díj általában az egyszerre hozzáférést kapó, párhuzamos használók számától függ - ezt automatikusan figyelik -, vagy egy cég alkalmazottainak vagy egy főiskola hallgatóinak teljes létszámától. A hozzáférést jelszóhoz vagy IP címhez köthető.

Az online változat lehet az adatfájl helyileg felmásolt változata, amelyet egy cég, főiskola vagy iskola licencl. Az Online Public Access Catalog (OPAC) programokat olyan lehetőségekkel ruházták fel, amelyek lehetővé teszik adatbázis építését, s olyan választási lehetőségeket kínálnak, amelyek nem voltak az eredeti online katalógus szoftverek részei, mint olyan keresési operátorok használata, amelyekkel meg lehet határozni a szavak sorrendjét és távolságát egy keresőkérdésben, hogy megkülönböztessük egymástól például az *information industry* és az *industry information* (információipar és ipari információ) vagy a *library school* és a *school library* (könyvtárosképző iskola és iskolai könyvtár) kifejezéseket. Az ilyen lehetőségekre szükség van olyan adatbázisokban, amelyekben hosszú, alapos referátumok vannak, s különösen teljes szövegű adatbázisok esetében. Alternatívaként egy intézmény azt a lehetőséget is választhatja, hogy előfizet távoli online adatbázisra (vagy adatbázisok csoportjára), amely(ek) valamelyik online információszolgáltatónál érhető(k) el. Ilyen esetekben a rendszer működtetésének és karbantartásának felelőssége az információszolgáltatóé.

Az 1990-es évek elejéig a legtöbb online szolgáltatás, amely az Interneten keresztül biztosított hozzáférést, a korlátozott Telnet szoftvertől függött, amely nem tette lehetővé a grafikus interfészt, a legtöbbször tisztán szöveges információkra korlátozódott - ez visszalépést jelentett a CD-ROM változatokhoz képest. Az 1990-es évek közepe óta a online adatbázisokhoz a web böngészőkön keresztüli hozzáférés vált dominánssá, amelyek grafikus interfészt és grafikus tartalmat ajánlanak.

Tenopir és Neufang (1995) vizsgálta azt, hogyan változtak több főiskolai könyvtár preferenciái az adatbázis szolgáltatási lehetőségek között. A tartalom elemzésének szempontjából ezek a szolgáltatási lehetőségek akkor kezdenek szerepet játszani, amikor a változatos tartalomhoz való hozzáférés biztosításának költségeit és hatékonyságát mérlegeljük.

A webre vitt és a weben született adatbázisok

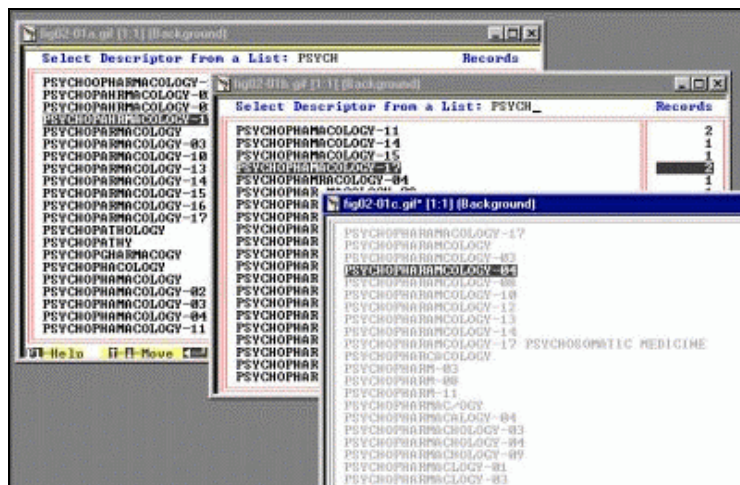
Sok a félreértés a *webes adatbázisokkal* kapcsolatban. Sokan azt hiszik, hogy a webes adatbázisok ingyenesek. Semmi sem állhat távolabb az igazságtól, mint ez az állítás. Valójában több pénzes adatbázis van a weben, mint ingyenes. Nem segítette a kérdés megvilágítását, amikor a DIALOG korábbi legfőbb vezetője, Dan Wagner azzal a túlzó állítással jött, hogy a DIALOG-nak több információja van, mint ami az egész weben található. Mivel maga a DIALOG is a weben található, ez az állítás kétségtelenül hibás. Talán azt jelenti, hogy a DIALOG-nál több információ érhető el, mint a web többi részén. Ez sem igaz. Ha csak a Lexis-Nexist nézzük, amely szintén elérhető a weben, s még néhányat a weben található legnagyobb pénzes adatbázis szolgáltatások közül, s összeadjuk a rekord számokat, akkor az is mutatja, hogy mennyire téves ez az állítás.

Amikor egy információforrás elérhető a weben, akkor helyesen webes adatbázisnak nevezhetjük - függetlenül attól, hogy ingyenes vagy pénzes. Ha egy létező adatbázist visznek át a kiadó saját telefonos rendszeréből (ilyen volt a legtöbb szakmai adatbázis), akkor ebben a könyvben az adatbázis *webre vitt* változataként utalunk rá. Ez illik a könyvben tárgyalt legtöbb online adatbázisra. Másrészt azonban, amikor egy adatbázist először a weben tettek közre, a könyvben *weben született* adatbázisként utalunk rá. Például a Books in Print adatbázis, amelyet a DIALOG, Ovid, OCLC, SilverPlatter és más hagyományos online szolgáltatók vittek a webre, webre került adatbázis. A Barnes & Noble, Amazon.com és Fatbrain adatbázisai azonban weben született adatbázisok.

2. Az adatbázisok tartalma elemzésének szempontjai

Az adatbázisokat számos szempont szerint lehet értékelni. Az adatbázist először interfészén és keresőgépen keresztül észleljük. A leginformatívabb, legmegbízhatóbb és legpontosabb adatbázis sem lehet sikeres, ha szoftver-összetevője nem intuitív és felhasználóbarát. A sugó informatívága, valamint a használók számára készült útmutatók minősége szintén szerepet játszhat az adatbázisok értékelésében. A szempontok másik csoportja a teljesítmény értékeléséhez kapcsolódik, mint például az átlagos válaszadási idő.

Ez a könyv nem foglalkozik ezekkel a szoftver kérdésekkel, bár kétségtelenül hatásuk van arra, hogyan ítélik meg az adatbázis tartalmát. Például a keresőszoftver indexböngészési képességei több esélyt ad a használónak, hogy képet kapjon az adatbázisról. Ha az adatbázisnak szokatlan, következtelen vagy elírt tárgyszavai vannak, de A szolgáltatás lehetővé teszi a böngészést használói számára, míg B nem, valószínű, hogy az A szolgáltatás adatbázisa jobb értékelést kap, mert az elírt kifejezések felismerhetők a használók számára böngészés közben. Kevésbé frusztráló látni, mennyi elírt változata van a *psychopharmacology* tárgyszónak a Mental Health Abstracts (MHA) adatbázisban (2.1. ábra), mint nem jutni hozzá fontos rekordokhoz vagy megpróbálni kitalálni az összes lehetséges elírt változatot egy olyan mezőben, amelynek a feltételezések szerint minden adatbázisban szigorú egységesítettnek kellene lennie. Az index megtekintésének lehetősége nélkül az összes elírás megtalálása olyan, mint ha egy fekete gyűrűt próbálnánk megtalálni egy sötét alagútban egy hold nélküli éjszakán. Ha a szoftver azt is lehetővé teszi, hogy az indexből egyszerre több fogalmat is kiválasszunk, a használó legalább jobban érezheti magát attól, hogy egy csapásra kiválaszthatja az országnevek összes variációját és pontatlan formáját az Information Science Abstractsból (ISA) (2.2. ábra). Az ISA DOS-os változatának szegény használói tájékozatlanok maradnak ezen adatbázis ország és nyelvi mezőinek sok pontatlanságáról és következtelenségéről, mert ezek az indexek nem böngészhetők.



2.1. ábra: A deszkriptorok nagymértékű elírása az MHA adatbázisban.



2.2. ábra: Több indextétel kiválasztása egyszerre az Information Science Abstractsból.

Az adatbázisok egyes hibáiért kompenzációt nyújtanak olyan keresési lehetőségek, mint például a keresési fogalom csonkolása, amely lehetővé teszi a különböző végződésű variációk és elírt formák visszakeresését. A szabadalomtulajdonosok nevének következtelen helyesírási formáit az ISA adatbázisban (2.3. ábra) viszonylag könnyű megtalálni egy csapásra az utolsó közös karakternél csonkolva, mint például PA=MITSUBISHI?. Ez a csonkolás természetesen nem segít visszakeresni azokat a rekordokat, amelyekben egy szó elején van elírás, mint például a Mistibushi, Misubishi, Mitshubishi erre a cégre ebben az adatbázisban.

Dialog Response		
Ref	Items	Index-term
E1	1	PA=MISTUBISHI
E2	1	PA=MISTUBISHI DENKI (JP)
E3	1	*PA=MISUBISHI
E4	1	PA=MISUBISHI DENKI K K (JP)
E5	3	PA=MIT
E6	1	PA=MIT TECH CORP.
E7	6	PA=MITA
E8	1	PA=MITA INDUST. CO. (JP)
E9	4	PA=MITA INDUSTRIAL CO. LTD. (JP)
E10	1	PA=MITA INDUSTRIAL CO., LTD., JP
E11	1	PA=MITECH
E12	1	PA=MITECH CORP.
E13	1	PA=MITEL
E14	1	PA=MITEL CORP (CA)
E15	3	PA=MITRE
E16	3	PA=MITRE CORP.
E17	2	PA=MITSHUBISHI
E18	1	PA=MITSHUBISHI DENKI (JP)
E19	1	PA=MITSHUBISHI DENKI K K (JP)
E20	246	PA=MITSUBISHI
E21	1	PA=MITSUBISHI CHEMICAL CORP. (JP)
E22	1	PA=MITSUBISHI CHEMICAL INDUSTRIES LTD. (JP)
E23	1	PA=MITSUBISHI CORP. (JP)
E24	2	PA=MITSUBISHI DENK K K (JP)
E25	3	PA=MITSUBISHI DENKI
E26	9	PA=MITSUBISHI DENKI (JP)
E27	190	PA=MITSUBISHI DENKI K K (JP)
E28	1	PA=MITSUBISHI DENKI K K (JP)
E29	1	PA=MITSUBISHI DENKI K.K.
E30	3	PA=MITSUBISHI DENKI K.K. (JP)
E31	1	PA=MITSUBISHI DENKI KABUSHIKI (JP)
E32	1	PA=MITSUBISHI DENKI KAIBUSHIKI KAISHA (JP)
E33	1	PA=MITSUBISHI DENKI KK
E34	12	PA=MITSUBISHI DENKI KK (JP)
E35	4	PA=MITSUBISHI DENKI KK, JP
E36	1	PA=MITSUBISHI DENKI, JAPAN
E37	1	PA=MITSUBISHI DENKI, KK, JP
E38	1	PA=MITSUBISHI ELEC. INFO. CENTER AMERICA, INC.
E39	1	PA=MITSUBISHI ELEC. INFO. TECH. CNT. AMERICA, INC
E40	1	PA=MITSUBISHI ELECTRIC CO. LTD. (JP)
E41	1	PA=MITSUBISHI ELECTRIC CORP.
E42	1	PA=MITSUBISHI ELECTRIC CORP. (JP)
E43	1	PA=MITSUBISHI JUKOGYO K K (JP)
E44	1	PA=MITSUBISHI K K (JP)
E45	1	PA=MITSUBISHI KK (JP)
E46	1	PA=MITSUBISHI LTD. (JP)
E47	1	PA=MITSUBISHI SEMICONDUCTOR AMERICA INC.
E48	1	PA=MITSUBISHI SEMICONDUCTOR AMERICA, INC.
E49	1	PA=MITSUBISHI, D. K K (JP)
E50	1	PA=MITSUBISHI, DENKI, JAPAN

2.3. ábra: Részlet az ISA szabadalomtulajdonosi indexéből.

Az adatbázisok értékelésének folyamata

Ez a könyv az adatbázisok tartalmának mennyiségi és minőségi összetevőire koncentrál. Az adatbázisok tartalmának szisztematikus értékelése az egyik alapvető tevékenység az adatbázisok minőségének értékeléséhez. Bár bizonyos benyomásokat eseti használat során is lehet szerezni az adatbázisról, helyes döntést arról, hogy kiválasszunk egy adatbázist egy másik rovására, vagy válasszunk egy adatbázis különböző változatai közül, csak tényekre és kézzelfogható adatokra szabad alapozni. Ezeket a fájl-előállítók vagy az adatbázis-kiadók nem feltétlenül közlik maguktól.

Az adatbázisok tartalmának elemzése szükségessé teszi tesztek, kísérletek, elemzések és egyéb kutatási műveletek és módszerek sorát. Ezek segítenek megerősíteni a fájl előállítóinak pontatlanul megfogalmazott állításait az adatbázis tárgyával, összetételével, lefedettségével, időhatáraival és frissességével kapcsolatban, valamint az adatbázisokat alkotó rekordok adatelemeinek pontosságáról, következetességéről és teljességéről.

A tartalom elemzése akkor kezdődik, amikor a forrásdokumentumok kiadói feltehetőleg korrigált, hiba nélküli információt adnak át a referáló és indexelő szolgáltatásnak vagy - teljes szövegű adatbázisok esetében - közvetlenül az adatbázis-kiadónak. Akkor végződik, amikor a közvetítők vagy a végfelhasználók megvizsgálják a keresés eredményét. A két pont között sok potenciális problématerület van, amely az értékelést indokolttá teszi. Williams (1990) kimutatta, hogy az adatok minősége bárhol befolyásolható lehet az információ létrejöttétől az adatbázis használatáig terjedő láncolatban. Figyelmeztet arra, hogy bár érték adódik hozzá a szerzőtől vagy alkotótól kezdve a végfelhasználóig, annak is megvan a lehetősége, hogy bekerüljenek hibák a végrehajtás, kihagyás, újrendezés vagy kiemelés során.

Nem biztató látni, hogy a fájlok előállítóinak és kiadóinak milyen lovagi attitűdje van az adatbázisok tartalmának minőségével kapcsolatban, ahogyan a jogi nyilatkozatok pajzsa mögé bújnak. Online szolgáltatások használatára vonatkozó kikötéseket és feltételeket vizsgálva Tenopir (1995, 122.) arra a következtetésre jutott, hogy nyelvezetük „azt sugallja, hogy senki sem felelős a tartalom minőségéért”. Valóban, a Dun & Bradstreet adatbázis DIALOG-os változatának felelősség-elhárítási nyilatkozata (2.4. ábra) úgy hangzik, mint azok a nyilatkozatok, amelyeket a fogságba esett katonáknak kell tenniük piárra éhes országok tv kamerái előtt. Sok fájl-előállító és adatbázis-készítő esküszik a minőség iránti elkötelezettségére, s nyomatékosan ígérnek olyan akciókat, amelyeket sohasem teljesítenek, de termékeikhez kapcsolt jogi nyilatkozatok nem erősítik meg ezeket az ígéreteket. Bizonyos mértékig érthető, mert a fájlkészítőknek és az adatbázis-kiadóknak következmények nélkül sikerül megúszniuk a gondatlanságot és a félrevezető állításokat, amelyek sok szakmában a káros gyakorlat és a megtévesztő hirdetések miatti büntetésekhez vezethetnek.

4. You will not knowingly provide information to a customer located in South Africa.
C. Terms Which Apply to Both End Users and Information Professionals
1. YOU ACKNOWLEDGE THAT THE AMOUNT YOU PAY TO The Dialog Corporation FOR THE INFORMATION RECEIVED FROM D&B WILL IN MOST INSTANCES REPRESENT A SMALL PORTION OF YOUR OVERALL COST OF THE PROJECT, TASK OR FUNCTION FOR WHICH THE INFORMATION WILL BE USED. YOU ALSO ACKNOWLEDGE THAT THE TYPE OF INFORMATION TO BE PROVIDED BY D&B TO The Dialog Corporation AND FURNISH TO YOU WILL CONTAIN A DEGREE OF ERROR. FINALLY, YOU ACKNOWLEDGE THAT THE PRICES WHICH D&B CHARGES The Dialog Corporation FOR INFORMATION ARE BASED, IN PART, UPON D&B'S EXPECTATION THAT THE RISK OF ANY LOSS OR INJURY WHICH MIGHT BE INCURRED BY YOU (AND/OR YOUR CUSTOMER IN THE CASE OF AN INFORMATION PROFESSIONAL) IN RELIANCE UPON THE INFORMATION WILL BE BORNE BY YOU (AND/OR YOUR CUSTOMER IN THE CASE OF AN INFORMATION PROFESSIONAL). FOR THESE REASONS, YOU AGREE THAT YOU ARE RESPONSIBLE FOR DETERMINING THAT ALL INFORMATION PROVIDED BY D&B TO The Dialog Corporation AND FURNISHED TO YOU IS SUFFICIENTLY ACCURATE FOR YOUR PURPOSES. NEITHER D&B NOR The Dialog Corporation MAKES ANY REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION, INCLUDING, BUT NOT LIMITED TO, ITS CORRECTNESS, COMPLETENESS, CURRENTNESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR WITH RESPECT TO THE MEDIA ON WHICH THE INFORMATION IS PROVIDED AND NEITHER D&B NOR The Dialog Corporation SHALL BE LIABLE FOR ANY LOSS OR INJURY ARISING OUT OF OR CAUSED, IN WHOLE OR IN PART, BY D&B'S OR The Dialog Corporation's NEGLIGENCE, ACTS OR OMISSIONS IN REPORTING, COMMUNICATING OR DELIVERING THE INFORMATION, DIRECTLY OR INDIRECTLY, TO YOU.
2. YOU AGREE THAT NEITHER D&B NOR The Dialog Corporation's WILL BE LIABLE FOR CONSEQUENTIAL DAMAGES, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

2.4. ábra: A Dun & Bradstreet adatbázis felelősség elhárítási nyilatkozata.

Elég sokat mond másik oldalról, hogy sok weben született adatbázisnál található olyan szem-betűnő kérés, hogy a használók tudassák az adatbázis kiadójával, ha hibát találtak az adatbázisban (2.5. ábra), sok még olyan ugrópontot is tartalmaz, amelyen keresztül a használók közvetlenül kapcsolatba kerülhetnek velük. Ez az, amit Mintz (1990) javasolt 1990-ben, amikor azt kérte az adatbázis-kiadóktól, hogy vezessenek be „Fixit” (javítsd meg) parancsot, amely közvetlenül a keresési folyamatból működik. A NewsNet volt az egyetlen - ma már sajnálatra méltóan megszűnt - online szolgáltatás, amely ezt megvalósította. E könyv szerzőjének az adatbázisok címkézésének bevezetésére vonatkozó javaslatát, amely hasonló ahhoz, amit a Food and Drug Administration megkövetel az élelmiszereken (Jacsó 1993a) sehol sem valósították meg a hagyományos fájllelőállítók vagy adatbázis-készítők. Érdekes, hogy egyes weben született adatbázisok, mint az Internet Movie Database, az All-Movie-Guide és a Computer Science Bibliography közölnek statisztikai információkat, amelyek hasonlóak a javasolt címkéhez (2.6. ábra).

Quinn (1995) kimutatta, hogy a közvetítők és hivatásos keresők az adatbázis-ipar végső ellenőrző csapatát képezik. Hasonlóan fontos, hogy ők azok, akik tanácsokat adnak a többi hivatásos kereső számára az adatbázisok súlyos hiányosságait bemutató jelentéseikkel, ahogyan Pagell (1987), Orenstein (1989, 1993) és Mintz (1995) tették ezt emlékezetes cikkeikben vagy Basch, Bates, Ojala, Quint, Tenopir és a könyv szerzője teszik ezt rendszeres szerkesztőségi cikkeikben és rovataikban. Bár ezek az adatbázis-kritikák, kommentárok, az adatbázisokat dicsérő vagy lehúzó írások nem jutnak el az alkalmi végfelhasználókhoz, némi visszajelzést nyújtanak a fájlok előállítói és az adatbázisok kiadói számára. Ők viszonzásul időnként figyelembe veszik az ajánlásokat, és vagy fejlesztik termékeiket (ez volt a helyzet az Economic Literature Index esetében), eltávolítják a kritizált adatbázist (ahogy a DIALOG tette ezt a klinikai halál állapotában levő Political Science adatbázissal), vagy megszüntetik a megállapodást az eredeti fájllelőállítóval (ez volt az eset az IFI/Plenum Data Corporationnel, amely korábban készítette az ISA adatbázist a Documentation Abstracts, Inc. számára, amely felbontotta a velük kötött megállapodást, miután megkapta e könyv szerzőjének az adatbázist értékelő jelentését).



2.5. ábra: Közvetlen ugrópont a hiba jelzésére az Internet Movie Database-ből.

Ennek az oldalnak a hibáit és hiányosságait jelenteni lehet az IMDb szerkesztőinek megnyomva a lenti gombot, ahol ezeket megvizsgáljuk, s ha elfogadjuk, a jövőbeni aktualizáláskor szerepeltetjük.

Subject Area	Number of references with		
	Abstracts	Keyword Fields	URLs
Others/Unclassifieds	19913	79288	23464
Theory/Foundations of CS	2303	32805	4914
Mathematics	1736	37299	7771
Artificial Intelligence	2947	11068	2966
Parallel Processing	11206	40224	4055
Computer Graphics	2434	40473	5077
Technical Reports	35060	5345	19680
Compiler	6013	25718	6411
Softw.Eng./Formal Methods	4713	5157	4391
Distributed Systems	2321	18603	5058
Databases	5452	9394	1170
Neural Networks	1051	1023	486
Human-Comp. Interaction	16680	5965	4709
Operating Systems	2236	3073	1198
Typesetting	1175	4363	423
Logic Programming	380	1566	97
Object-Oriented	596	3485	394
Wavelets	537	646	279
Total	116753	325495	92543

2.6. ábra: Adatbázis statisztika a Computer Science Bibliography adatbázisban (részlet)

A tartalom elemzésének története és szakirodalma

Azzal, hogy a web válik a világ legnagyobb olyan helyévé, ahol a szerzők költségén jelentetik meg a műveket, megnőtt az érdeklődés, ha nem is a tartalom elemzése, de legalább az iránt, hogy tanácsokat adjanak a használóknak a weben született adatbázisok veszélyeiről. Sokan szüntelenül figyelmeztetnek annak veszélyeire, milyen megbízhatatlan, pontatlan, elavult, nem teljes információk keringenek a weben. Úgy tűnik, kevesen tudják azt, hogy ezek a veszélyek leselkednek a szakmai adatbázisokban és elődeikben - a nyomtatott referáló és indexelő kiadványokban és a címjegyzékekben - évtizedek óta. Még kevesebben látszanak tudni, kik és milyen folyóiratokban, magazinokban, konferenciakötetekben publikáltak ezeknek a forrásoknak az értékeléséről. Legtöbbjük csupán felbőfögi a klisészerű figyelmeztetéseket, jobb esetben némi anekdotikus adalékkal fűszerezik. Egy alkalommal egy szerző szóvá tette a korábbi irodalom hiányát, annak ellenére, hogy cikkek bőséges számban tárgyalták a problémát. Hurst (1999) könyvében kijelentette, hogy „a legnagyobb felfedezésünk [sic], amit tudomásunk szerint egyetlen információtudományi vagy könyvtári publikáció sem tárgyalt, hogy gyakran többéves hiányok vannak a feltárt anyagban.”

Könyvünk 5. fejezetében, amely a folyóiratok feltártságának mértékével foglalkozik, sok hivatkozás található olyan cikkekre és könyvekre, amelyek éppen ezt a kérdést tárgyalják, köztük Pagell (1987), Orenstein (1993), Grzeszkiewicz és Hawbaker (1996), Tenopir (1995) és Jacsó (1995, 1997a) írásai. Egy cikkében, amely nagyrészt 1999-es könyvéből való kivonat, de amelyből a „tudomásunk szerint” kifejezést kihagyta, Hurst (2000) megerősítette és kiszélesítette megállapítását, azt igyekezve bizonyítani, hogy „senki sem tiltakozott hangosan a következtelen folyóiratcímek, az indexelésből kihagyott évek miatt, sem azért, mert egy folyóiratról gyakran csak a fontosabb cikkeket dolgozzák fel, nem a teljes számot. „

A következtelen folyóiratcímeket már régóta élesen kritizálják olyan információs szakemberek, mint Williams és Lannom már 1981-ben, majd Tenopir (1997), LaGuardia (1991) a *Philosopher's Index*, Jacsó (1998b, 1999b) az *Economic Literature Index* és a Pascal esetében, Johnson (1999) és Jacsó (1999c) számos könyvtár- és információtudományi publikációban.

A valóság az, hogy sokat publikáltak már az adatbázisok tartalmának elemzéséről. Nem csupán a szerző megtisztelő feladata volt (1997), hogy az *Annual Review of Information Science & Technology*-ban elismeréssel adózzon azok előtt, akik úttörők voltak az ilyen kutatásban. Ennek az is a célja volt, hogy háttérül szolgáljon mindazok számára, akik szeretnék megismerni ennek a kutatási területnek a mérföldköveit és legfontosabb részterületeit. Különösen fontos, hogy legalább elismerjük azokat, akiknek az elméleti és gyakorlati hozzájárulása a nyomtatott referáló és indexelő kiadványok korában és az online és CD-ROM adatbázisok úttörő napjaiban egyengette az utat és modellül szolgált az 1980-as és 1990-es évek során végzett hozzáértő elemzések számára.

Az adatbázisok tartalmát elemző mai tevékenységek többsége ezek nyomtatott formájú elődeinek értékeléséből származik. Goldberg (1992) hangsúlyozza azt, hogy „akár nyomtatott, akár számítógépes indexről van szó, sok gyakorlat és szabvány ugyanaz marad, és a nyomtatott indexekre vonatkozó értékelési szempontok hasonló módon alkalmazhatók a CD-ROM-okra is.” Ez kiterjeszthető a könyvtári gyűjtemények, referáló és indexelő folyóiratok vagy különböző nyomtatott enciklopédiák értékelési útmutatóira (amelyeket Katz (1987), Lancaster (1979) és Kister (1986) klasszikus tankönyvei tartalmazzák), s amelyek könnyen adaptálhatók egy online szolgáltatás adatbázis-gyűjteményeire vagy egyes adatbázisokra. Ezeknek a munkáknak vannak újabb kiadásai, de az eredeti kiadásokra hivatkozom, hogy jelezzem, milyen messzire megy vissza a tartalom elemzése még monografikus művekben is.

Martyn és Slater (1964) cikke a referáló folyóiratok teszteléséről és Lancaster (1971) nagy hatású cikke az indexek és referáló folyóiratok értékeléséről ma is érvényes és alkalmazható maradt szekundér információs szolgáltatások elektronikus verzióira. A helyesírási hibák első nagymértékű elemzése, amelyet Bourne végzett 1977-ben, mit sem veszített aktualitásából. A hivatkozásvizsgálatokkal foglalkozó, alapos és részletes szemléjében Smith (1981) arra figyelmeztetett, hogy a pontatlan hivatkozások hatással lehetnek hivatkozásvizsgálatokból levont következtetésekre, ez még igazabbnak hangzik ma, mint 15 évvel ezelőtt. Williams és Lannom (1981) ma is publikálhatná húszéves cikkét arról, hogy hiányzik a folyóiratcímek egységesítése az adatbázisokban, csak a statisztikáikat kellene aktualizálni, nem üzenetüket.

A referáló és indexelő szolgáltatások egyik klasszikus, kiterjedt vizsgálatában, amelyet Pauline Atherton Cochrane végzett a *Physics Abstracts*-et felhasználva, majd Stella Keenannel közösen megírta 1964-ben (Atherton Cochrane és Keenan, 1965), ma is modellként szolgálhat a hozzáértő, szakszerű értékelésre. Ez érvényes sok más ragyogó és megvilágító írásra, amelyek főszerepet játszottak abban, hogy könyvtárosságot választottam szakmámnak az 1970-es évek végén és az adatbázisok értékelését az 1980-as évek végén. Ennek a kutatásnak a történetét Atherton Cochrane felidézte a <http://www.libsci.sc.edu/bob/ISP/cochrane2.htm> címen.

Az adatgyűjtést az időbeli lemaradás és a kiadás helye közti kapcsolattal foglalkozó vizsgálathoz, amelyet Turtle és Robinson (1974) végzett el a Library and Information Science Abstractsre (LISA) és Library Literature-re vonatkozóan, nem segítették számítógépek, de módszertanuk, tömör és világos értékelésük, kritikai szemléletük példás bárki számára, aki ma az adatbázisok minőségével kapcsolatos kutatást végez, azok időszerűségére vonatkozóan. Ugyanez mondható el Tenopir (1982) kutatásáról, amelyet az adatbázisok által feltárt anyag elemzésének bibliográfiai és szakterületi módszereivel kapcsolatban végzett.

A japán információs hálózatra és tudományos és műszaki irodalom információs hálózatára és bibliográfiai ellenőrzésére vonatkozó, Gibson és Kunkel által 1980-ban végzett vizsgálat tartós érdeklődést keltett a japán folyóiratok, konferencia-előadások és szabadalmak feltártóságának vizsgálatára. Ewbank (1982) az elsődlegesen faktografikus adatbázisok kiválasztására vonatkozó útmutatójában olyan iránymutatással szolgált, amely megjelenik kritériumként a mai értékelési ellenőrző listákon, mint a tárgy, teljesség, a feltárt források köre, pontosság, elérhetőség, a jelentés formátuma és a technikai támogatás. A piszkos adatok átfogó problémáját jelezte Norton (1981) és Pemberton (1983) az 1980-as évek elején.

Míg általában az adatbázisok, s a könyvtár- és információtudományi adatbázisok különösen sok kívánnivalót hagynak maguk után még ma is, értékelésük az 1960 és 1990 közötti időszakban példaadó modellt szolgáltat az átfogó jellegre, módszertanra és a következtetésre.

Gilchrist (1966), Goldstein (1973), Edwards (1976), Gilchrist és Presanis (1971), Dansey (1973), Bottle és Efthimiadis (1984), LaBorie, Halperin és White (1985) és Stieg és Atkinson (1988) cikkei voltak a legalaposabb tanulmányok az adatbázisok és nyomtatott formájú elődeik tartalmával kapcsolatban. Közülük több alkalmazott újszerű módszertant a mérési kritériumokra és/vagy az eredmények bemutatására vonatkozóan.

Viszonylag kevés olyan tanulmány készült, amely nagyszámú tesztkeresést, mintarekordok nagy halmazát vagy párhuzamos értékelési módszereket használt. Az egyik vizsgálatot az 1980-as évek elején a Drexel Egyetem információs tanulmányi tanszékének (College of Information Studies at Drexel University) munkatársai végezték, hogy értékeljék a MEDLINE által feltárt anyagot és az orvosi viselkedéstudományok szakirodalmát (Griffith et al. 1986). Ez a tanulmány modellként szolgált számos további, sokkal kisebb méretű vizsgálathoz. Sparck Jones és Van Rijsbergen (1976) kimutatta, hogy sok visszakeresési teszt nem meggyőző jellege a használt tesztgyűjtés elégtelen voltának tulajdonítható, elsődlegesen a visszakeresési kísérleteknek megfelelő méretű fájl hiányának. Pao (1989) 48, a teszteknel használt adathalmazt vizsgált meg, és úgy találta, hogy 41 tartalmazott 5000-nél kevesebb tételt, 17-ben volt még 500-nál is kevesebb. Sajnálatos módon kevés vizsgálat alkalmazta az adatbázisok értékelésére a szakterületi profil és a bibliográfia szerinti megközelítés párhuzamos módszerét, amelyet Tenopir (1982) tesztelt és hasonlított össze a GeoRef és a GeoArchive adatbázisokat használva céladatbázisként. Sok adatbázis CD-ROM-on való elérhetősége két okból is új horizontot nyitott a tartalom elemzéséhez. A CD-ROM adatbázisok megváltoztathatatlan tartalma lehetővé teszi a kutatók számára, hogy korábbi kereséseket reprodukáljanak az eredmények igazolásához, vagy különböző teszteket végezzenek el ugyanabban az adatbázisban, időről időre egy változót módosítva. Az, hogy az adatbázisokat korlátlanul lehet használni fix licencdíjért, lehetővé teszi a kutatók számára, hogy reprodukáljanak kereséseket az eredmények igazolására, s kiterjesszék elemzésüket az egész adatbázisra - ez egy olyan lehetőség, amely online rendszerekben ritkán megvalósítható (Jacsó 1992a). A kicsi, nem reprezentatív minták problémáját valamennyire legyőzte a számítógépes szoftverek elérhetősége, amelyek képesek megfelelő nagyságú, jól kiválasztott minták feldolgozására. A lehetőségek ellenére egy adatbázis rekordjai teljes populációjának teljes körű elemzése (ahogy azt Hood és Wilson végezte ragyogóan 1994-ben a LISA CD-ROM-változatára vonatkozóan) még mindig nem tipikus jelenség.

Az adatbázisok minősége iránti megnövekedett érdeklődés részben annak köszönhető, hogy a géppel olvasható adatbázisok kényelmes célpontot jelentenek az értékelés számára, az adatbázissal együtt kapott keresőszoftverek maguk is gyakran hatékony eszközei lehetnek az értékelésnek. Ezek jelentős mértékben csökkentik a nagy minták elemzésének unalmas munkáját. A CD-ROM technika bevezetése még jobban felszította az érdeklődést. A CD-ROM-os licencszerződések általában nem a végtelenségig szólnak, hanem rövid távú elkötelezettségek előfizetés formájában. Ez a könyvtárosokat és az információs szakembereket még inkább hajlamossá teszi arra, hogy többet tudjanak meg ezekről az elektronikus információs forrásokról, s átgondolják választásukat az adatbázis tartalmának elemzése alapján.

Az ingyenes webes adatbázisok kiváló lehetőséget kínálnak arra, hogy a keresők megvizsgálják azok tartalmát. Szoftverük és adatszerkezetük azonban kevésbé mutatkoznak meg értékelő keresések számára, mint a hagyományos online és CD-ROM-os információforrásoknál. Például a weben született adatbázisoknak 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, ahogy az a DIALOG, az Ovid és a SilverPlatter adatbázisainak legtöbb rekordja esetében megtörténik. Lehetséges, hogy a szoftvernek nincsenek olyan vonásai, amellyel meghatározható például a rekordok teljes száma az adatbázisban. Ezen korlátok ellenére jobb és sokkal hatékonyabb lehetőségek vannak a tartalom elemzésére, mint egykor nyomtatott kiadványok esetében, amelyek hatalmas mennyiségű munkát követeltek.

3. Az adatbázis szakterületi hatóköre

Szélesebb értelemben az adatbázis tárgya határozza meg a benne feltárt anyagot. A cél, hatókör magába foglal olyan szempontokat, mint a szakterület, méret, összetétel, frissesség, a forrásdokumentumok és -tárgyak (folyóiratok, cégek, zenei hangfelvételek) feltárttságának mértéke, nyelv és földrajzi eredet szerinti feltárttság. Szűkebb értelemben a hatókör az adatbázis tárgyi lefedettségét határozza meg. Egyes feltételek, mint a szakterület, a méret, a források köre kölcsönös kapcsolatban vannak egymással és ebben az összefüggésben kell szemlélni őket.

Hogy milyen jó egy adatbázis hatóköre, szubjektív kérdés. Ki kell elégítenie egyéni használó, illetve egy könyvtár vagy intézmény használóinak érdeklődését. A weben született adatbázisok egyik legvonzóbb jellegzetessége, hogy mély és széles lefedettséget biztosíthatnak egy nagyon speciális témában, amely viszonylag kevés ember elsődleges érdeklődési területe. A Climbing Database (hegymászók adatbázisa) (<http://www.climbingboulder.com>) és a kiváló Biographies of Women Mathematicians (matematikusnők életrajzai), amely körülbelül 100 tudósról tartalmaz életrajzi esszét (<http://www.agnesscott.edu/lriddle/women/women.htm>), jó példák lehetnek erre.

Valamivel szélesebb a 4000 Years of Women in Science (A nők 4000 éve a tudományban) életrajzi adatbázis (<http://crux.astr.ua.edu/4000WS/4000WS.html>) szakterülete. Valamivel több tételt tartalmaz (körülbelül 140-et 1999 végén), de valamivel szűkebb időszakot ölel fel, mert elsősorban a 20. század előtti tudósnőket vesz fel. Az archINFORM adatbázis (3.1. ábra) körülbelül 8000 20. századi építészeti tervről tartalmaz információkat, s azt illusztrálja, hogy valóban a kicsi is szép lehet (<http://www.archINFORM.net>). Az ilyen adatbázis nem lenne megvalósítható egy nagy kereskedelmi online vagy CD-ROM kiadó számára korlátozott tárgya és kis célcsoportja miatt. Még akkor is, ha egy adatbázis kicsi, megvalósítható a weben, s méltányolni kell, amikor a tárgyra vonatkozó nyilatkozat világos és egyértelmű, vagyis megfelelő módon tájékoztatja a használót.

Például az AIAA Meeting Papers kereshető bibliográfiai adatbázisa (3.2. ábra) világossá tette tárgyát honlapján (<http://www2.aiaa.org/Research>), (A megadott címről ma már a <http://www.aiaa.org/content.cfm?pageid=413> címre irányítják át a használót, ahol az AIAA elektronikus könyvtára található, a <http://www.aiaa.org/content.cfm?pageid=2> cím alatt tájékoztató is olvasható az adatbázis használatáról.) Így tett a Pennsylvania Flora Project Database (<http://www.upenn.edu/paflora/dbsearch.html>) is. (Erről a címről ma már a <http://www.paflora.org/search.html> címre irányítják át a használót, ahol már nincs meg az adatbázis leírása.)

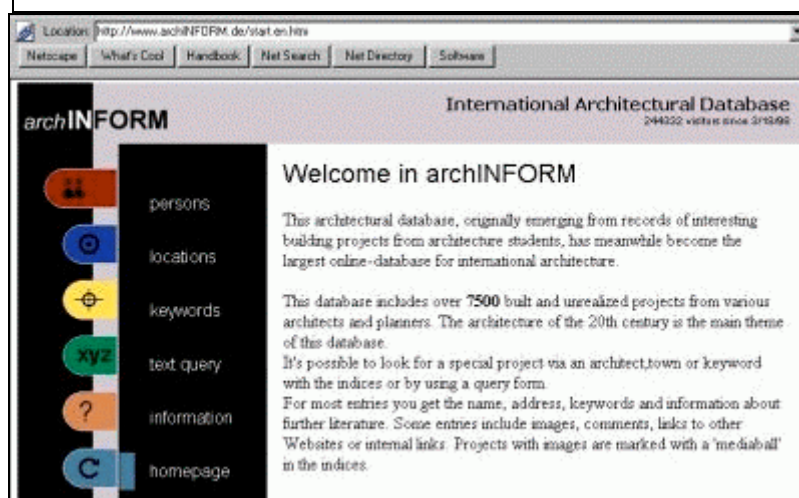
Üdvözljük az archINFORMnál

Ez az építészeti adatbázis, amely eredetileg építészhallgatók érdekes építészeti terveinek rekordjaiból alakult ki, időközben a nemzetközi építészet legnagyobb online adatbázisává vált.

Az adatbázis több mint **7500** megépített és meg nem valósított tervet tartalmaz különböző építészekről és tervezőkről. Az adatbázis fő témája a 20. század építészete.

Az építész, a város vagy kulcsszó szerint lehet terveket keresni az indexek segítségével vagy egy keresőúrlapot használva.

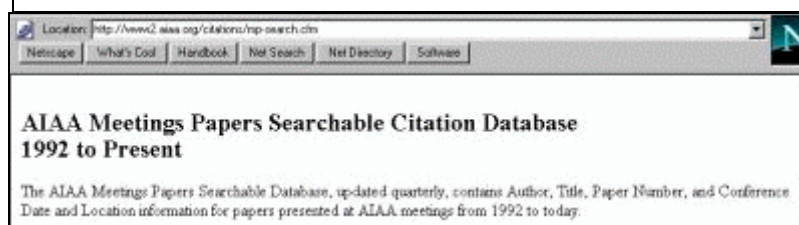
A legtöbb tétel esetében megkapják a nevet, címet, kulcsszavakat és a további irodalomra vonatkozó irodalmat. Egyes tételek tartalmaznak képeket, kommentárokat, más weboldalakhoz vezető vagy belső ugrópontokat. Azokat a terveket, amelyekben kép is szerepel, a „mediaball” jelzi az indexekben.



3.1. ábra: Az archINFORM adatbázis tárgyával kapcsolatos nyilatkozata

Az AIAA Meetings Papers Searchable Database 1992-től máig

Az AIAA Meetings Papers Searchable Database-t (konferencia-előadások kereshető adatbázisa) negyedévenként aktualizáljuk, az AIAA 1992-től máig rendezett konferenciáin tartott előadásokhoz kapcsolódóan azok szerzőire, címére, az előadás sorszáma és a konferencia dátumára és helyére vonatkozó információkat tartalmaz



3.2. ábra: Az AIAA magyarázata adatbázisának tárgyáról

Egy adatbázis neve vagy rövidítése gyakran utal a benne feltárt anyag tárgyára. Például a Library Literature, Library and Information Science Abstracts és a Mental Health Abstracts egyértelműen meghatározza az adatbázis szakterületét. A PsycINFO és a PsycLIT is utal tárgyára. Néha az adatbázis neve önmagában nem lehet elég informatív. Az Information Science Abstracts számos

könyvtartudományi és információtechnológiai folyóiratot is feltárt, ahogy azt összes verziójának implementációjában szereplő tárgyi megjegyzések is világossá teszik (3.3. ábra).

Csak tesztkeresések tudják igazán feltárni, mi van a névben vagy a név mögött. A feltártságra vonatkozó állítások kérésre opcionálisan megjelennek (mint például amikor rákattintunk az i szimbólumra az Ovid adatbázisaiban, vagy megadjuk a ?file 202 parancsot a DIALOG-ban) szintén tájékoztatással szolgálhat az adatbázis tárgyköréről. Figyelni kell arra, hogy fájl előállítói vagy az adatbázis kiadói honlapján szereplő reklámszövegek mellett, ezeket az állításokat is gyakran némi kételkedéssel fogadjuk (gyakran erős kételkedéssel).

Az 5. fejezet azt az esetet illusztrálja, amikor az Information Science Abstracts adatbázis szerkesztőjének állítása szűkíti az ISA tárgykörét, amely a szerkesztő szerint az információtudomány, szemben más adatbázisokkal, mint a Library Literature és a Library and Information Science Abstracts, amelyek a könyvtár- és információtudományt dolgozzák fel. Ez megzavarhatja a használókat, akik a DIALOG adatbázis-leírásának (bluesheetjének) a fájl leíró részében azt olvassák, hogy „az Information Science Abstracts bibliográfiai adatokat és referátumokat szolgáltat az információtudomány és a könyvtartudomány területéről” (<http://library.dialog.com/bluesheets/html/bl0202.html>), (Az ISA adatbázisa jelenleg nem érhető el a DIALOG-nál, így természetesen ez a leírás sem.) Másrészt vannak olyan adatbázisok, amelyek reklámszövegei teljes mértékben megfelelnek a valóságnak. Ilyen az H. W. Wilson adatbáziscsalád egésze (3.4. ábra).

Fájlleírás

Az Information Science Abstracts bibliográfiai adatokat és referátumokat szolgáltat az információtudomány és a könyvtartudomány területén. Anyagában nemzetközi, az adatbázis több mint 300 folyóiratról indexel és referál cikkeket, valamint könyveket, kutatási jelentéseket, kongresszusi kiadványokat és szabadalmi leírásokat. Minden rekord tartalmaz bibliográfiai adatokat és referátumokat, továbbá deskriptorokat

Information Science Abstracts

Last Loaded on Web: Friday, May 26, 2000
Last Update To Bluesheet: May 17, 2000

Bluesheet Contents [PDF version](#)

File Description	Dialog File Data	Special Features	Sample Record	Sort	Rates
Subject Coverage	Database Content	DIALINDEXOneSearch Categories	Basic Index	Rank	
Sources	Document Types Indexed	Contact	Additional Indexes	Map	
Print Counterparts	Geographic Coverage	Terms and Conditions	Limit	Predefined Format Options	

File Description [\(new\)](#)

Information Science Abstracts provides references and abstracts in the fields of information science and library science. International in scope, the database indexes and abstracts articles from over 300 journals, as well as from books, research reports, conference proceedings, and patents. Each record includes the bibliographic citation and abstract as well as descriptors drawn

3.3. ábra: Az Information Science Abstracts adatbázis tárgyának meghatározása

A Wilson Library Literature Database bibliográfia adatbázis, amely 354, az Amerikai Egyesült Államokban vagy máshol kiadott, alapvető könyvtár- és információtudományi periodikumot indexel. Könyveket, könyvfejezeteket, kongresszusi kiadványokat, könyvtári szakdolgozatokat és brosrákat is indexel.

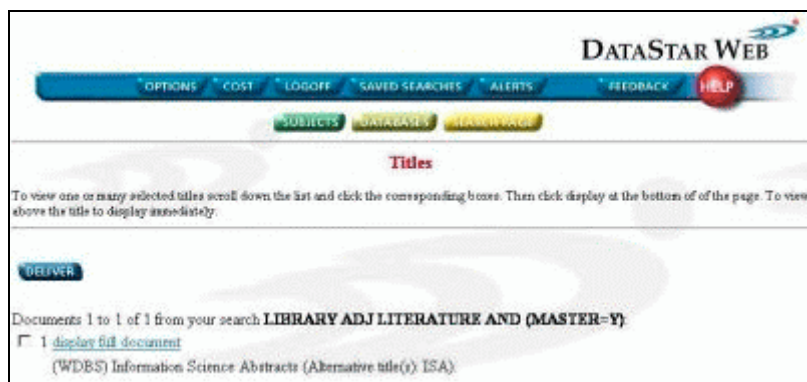
SCOPE:	
The Wilson Library Literature Database is a bibliographic database that indexes 354 key library and information science periodicals published in the United States and elsewhere. Books, chapters within books, conference proceedings, library school theses, and pamphlets are also indexed.	
SUBJECT COVERAGE:	
Automation	Information Brokers
Cataloging	Library Associations & Conferences
Censorship	Library Equipment & Supplies
Children's Literature	Online Searching
Circulation Procedures	Personnel Administration
Classification	Preservation of Materials
Copyright Legislation	Public Relations
Education for Librarianship	Publishing
Government Funding	
TYPE OF MATERIAL (FROM PERIODICALS) INDEXED:	
<ul style="list-style-type: none"> ◆ feature articles ◆ obituaries ◆ interviews ◆ notices of appointments and awards ◆ regularly appearing columns ◆ special theme issues ◆ editorials (of reference value) ◆ letters to the editor (of reference value) ◆ reviews of books in library and information science 	
OTHER TYPE OF MATERIAL INDEXED:	
<ul style="list-style-type: none"> ◆ books ◆ chapters or sections of books ◆ library school theses ◆ pamphlets 	

3.4. ábra: Egy Wilson adatbázis tárgyának meghatározása

Az Elsevier cég által kiadott Scirus címjegyzék esete világosan illusztrálja, hogy sem a nevet, sem a kiadó állításait az adatbázis tárgyáról nem lehet szó szerint venni. Az Elsevier minden képernyőn szembeűnő módon megjelenít egy logót, közhírré téve, hogy „Scirus, színtiszta tudományos információ”. E könyv szerzője, aki a Scirusban az összehasonlító vallástudomány referenzsforrásaira vonatkozó tudományos információt keresett, olyan weblapokra mutató linkeket kapott, amelyek a címlapjukon a legvulgárisabb szavakat tartalmazták. Egy másik, a négybetűs szavakra végzett keresés 38000 találatot hozott. Néhány lapnak a meglátogatása igazolta, hogy ez a címjegyzék nem kizárólag tudományos információt tartalmaz. (<http://www2.hawaii.edu/~jacso/extra/infotoday/scirus/scirus.html>).

Az átfogó szó a fájlok előállítóinak a fájl tárgyára vonatkozó állításaiban nagyon gyakran előjön, de ennek a szónak a jelentését nyilvánvalóan különböző módon interpretálják. A Bowker-Saur World Databases, amely körülbelül 10000 rekordot tartalmaz, átfogónak (mindenre kiterjedőnek) mondja magát kiadói reklámjában. A Gale cég Directory of Databases szintén átfogónak mondja magát, és 15000 adatbázisra vonatkozó rekordjával az állítás sokkal inkább igazoltnak tűnik. Ha ehhez a tényhez azt is hozzátesszük, hogy a World Databasesnek az adatbázisok minden egyes implementációjáról külön rekordja van, míg a Gale egyetlen rekordot készít médiatípusonként (azaz egy rekordot a MEDLINE online verziókra és egy másikat a CD-ROM-verziókra), a World Databases átfogó jellegére vonatkozó állításai még irreálisabbnak látszik. Az ebben a bekezdésben említett adatbázisokra vonatkozó tesztkeresések igazolják azt a benyomást, hogy a Bowker-Saur adatbázis messze van a teljességtől.

Nem voltak benne rekordok olyan adatbázisokról, mint a Library Literature vagy a Library and Information Science Abstracts, annak ellenére, hogy az utóbbi nem csak brit adatbázis, hanem gondozója - jól gondolta, kedves olvasó - a Bowker-Saur. A *library* és a *literature* szavakra vonatkozó együttes keresés egyetlen rekordot talált, az Information Science Abstractset, mert ez azt állítja magáról, hogy feldolgozza a könyvtári szakirodalmat (3.5. ábra). Az azokra a tételekre vonatkozó keresés, amelyekben a *microcomputer* és az *abstracts* szavak egymás mellett állnak, csak egy rekordot talál, a Computer Abstractset, s nincs rekord a Microcomputer Abstractsre vonatkozóan. (Erre a keresésre még azelőtt került sor, mielőtt az adatbázis Internet and Personal Computing Abstractsre változtatta volna a címét.) Nem meglepő, hogy a World Databases címjegyzéknek nincs tétele magára a World Databasesre vonatkozóan sem. Soha nem is aktualizálták.



3.5a. ábra: Nincs rekord a Library Literature adatbázisra.

A Gale Database of Publications and Broadcast Media azt állítja, hogy „átfogó adatbázis, amely részletes információt tartalmaz 61000 újságról, magazinról, folyóiratról, periodikumról, címjegyzékről, hírlevélről és rádió-, televízió- és kábel állomásról és rendszerről.” Levonva a rádió és televízióállomásokra és rendszerekre vonatkozó rekordokat, a maradék körülbelül 50000 periodikus kiadvány lehet. Az ilyen szűrés lehetővé teszi az összehasonlítást az Ulrich's Periodical Database-zel és az EBSCO Serials Database adatbázisával, mindkettő átfogó feltártságúnak mondja magát - és adataik körülbelül 210000 és 155000 periodikumról adnak információt.

A The Serials Directory (TSD) és az Ulrich's Periodical Database átfogó jellege mindjárt más értelmet kap annak ismeretében, hogy az ISSN Database-nek 850000 rekordja van, és a Kongresszusi Könyvtár időszaki kiadványokat tartalmazó részhalmazának pedig közel 900000 rekordja. A kisebb adatbázisok közül egyik sem említ tárgyi szűkítést, de sok ilyennek kell lennie, és némelyik közülük jogosan is alkalmaz korlátozást. Például az adatbázis tárgyának a dokumentumok nyelve és a kiadás országa szerinti korlátozása ésszerű, s megmagyarázhatja a kisebb méretet és a szűkebb tárgyat.



3. 5b. ábra: Nincs rekord a Microcomputer Abstracts adatbázisra.

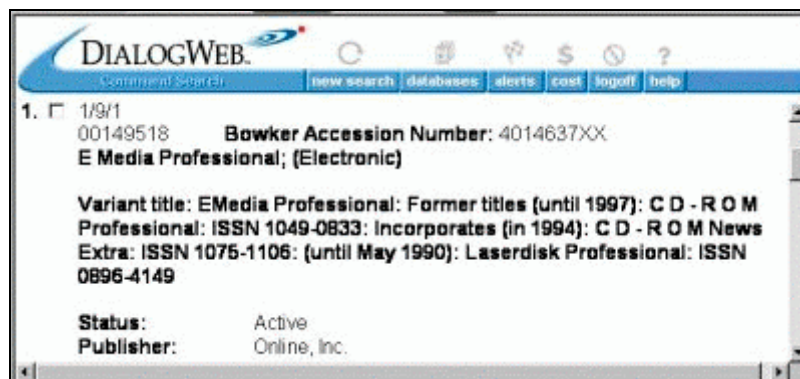
Lehetnek sokkal kevésbé nyilvánvaló okai is a kisebb méretnek és az adatbázis érzékelt tárgyának. Például a követő tételkatalogizálás gyakorlata (amikor a cím megváltozása után az EBSCO új rekordot hoz létre és a korábbi cím rekordja is megmarad az adatbázisban), az EBSCOt előnyösebb helyzetbe hozza az Ulrich'sszal szemben a címek teljes számát tekintve,

mert az Ulrich's a legutolsó tétel katalogizálásának elvét alkalmazza. E szerint az eljárás mód szerint az újabb címre vonatkozó rekord magába foglalja a korábbi címmel kapcsolatos információkat és a korábbi cím rekordját törlik.

A megközelítés e különbsége szerint az EBSCO-nak külön rekordja van a Laserdisk Professional, CD-ROM Professional és az EMedia Professional címekre (3.6. ábra), plusz egy további rekord az E Media Professional változatra (ISSN szám és ok nélkül) és egy rekord a CD-ROM News Extra-ra, amely a CD-ROM Professional rosszul tervezett és rövidéletű melléklete volt. Az Ulrich'snak egyetlen tétele van, csak a mostani cím alatt található meg, s ez a rekord tartalmazza a két korábbi címet (3.7. ábra). Figyelembe véve a periodikumok címváltozásainak rendkívüli mértékét, ez nagyon jelentős különbséget okozhat, amikor összehasonlítjuk az adatbázisok állítólagos tárgyát a címek szavai, a kiadó, deskriptorok, országnevek, nyelv stb. szerint tesztkeresésekkel.



3.6. ábra: Az EBSCO adatbázisában külön rekordok találhatóak egy folyóirat jelenlegi és korábbi címére



3.7. ábra: Az Ulrich'sban egyetlen rekord található egy folyóirat jelenlegi és korábbi címére

Az adatbázis tárgya pontos mérésének másik módja tesztkeresések elvégzése. A címek indexmezőjében egy szóra (*bibliometrics*) vagy egy összetett kifejezésre (*artificial intelligence*) végzett tesztkeresés gyors és könnyű módja annak, hogy néhány témára vonatkozóan benyomást szerezzünk az adatbázis által feltárt anyagról. Néhány egyszerű keresés eredménye a használat szkeptikussá teheti vagy bizalommal töltheti el azokkal az állításokkal szemben, amelyeket az adatfájl előállítója vagy kiadója tesz a feltárt témákkal kapcsolatban, különösen azokkal, amelyeket kiemelten említene meg a reklámanyagokban. A címekre vonatkozó keresések önmagukban nem lehetnek meggyőző erejűek (kivéve olyan extrém eseteket, amelyeket a következőkben tárgyalok). Azokkal az eredményekkel együtt kell értékelni őket, amelyek az adatbázis összetételére, kurrensségére, retrospektivitására és folyóiratbázisára vonatkozóan végeztek, valamint arra a gyakorlatra vonatkozóan, amely meghatározza a tételeknek az adatbázisba való felvételét. A kérdés kombinálható a kiadási évekkel, a keresés eredményét a legutolsó két vagy három évre korlátozva, hogy az adatbázis legfrissebb részét teszteljék. Egy ilyen kérdés így néz ki:

S supercomput?/ti AND PY=1998:2000 (a DIALOG-nál)

F supercomput* in ti AND PY=1998-2000 (a WebSPIRS-ben)

Néha a vizsgált adatbázisban végzett egyetlen teszt önmagában is kétségeket ébreszthet az adatbázis tárgyával kapcsolatban. Az előbbi keresés az ISA adatbázisban 2000 novemberében nulla találatot eredményezett. Bár az nyilvánvaló, hogy nem minden, a szuperszámítógépekkel foglalkozó cikk vagy konferencia előadás címében szerepel ez a szó, elképzelhetetlen, hogy egyetlen sincs benne egy olyan adatbázisban, amely a szuperszámítógépeket a 22 legfontosabb szakterület közé sorolja adatbázis-leírásában, amelyet 2000-ben két alkalommal is aktualizáltak. Egy, a cím mezőre való korlátozás nélkül végzett keresés (azaz a címben, a referátum szövegében és a deskriptor mezőben) egyetlen rekordot eredményezett a három legújabb évben; nem kifejezetten erőteljes feltártság.

Az egyes adatbázisokban végzett keresések fényt deríthetnek arra, hogy mennyire felelnek meg a valóságnak a tárgyukra vonatkozó állítások, de a több adatbázisban végzett tesztelés a dolgokat megfelelőbb perspektívába állíthatja. A csak a címek indexében való keresés egyenlő feltételek szerinti keresést biztosít, megszüntetve azokat a különbségeket, amelyeket a referálás és az indexelés különbségei okozhatnak. Nyilvánvaló, hogy az az adatbázis, amely a keresőszót deskriptorként tartalmazza, több találatot adna, mint az, amely egy szinonimát vagy egy szűkebb fogalmat használ deskriptorként.

Ennek a tesztnek az abszolút korrektsége érdekében a keresést ugyanarra az időszakra és ugyanarra (ugyanazokra) a nyelv(ek)re kell korlátozni, amelyek az állítások szerint szerepelnek minden adatbázisban. Ha lehetséges, a keresést ugyanarra a dokumentumtípusra kellene korlátozni, mint például folyóiratcikkekre vagy konferencia-előadásokra, megszüntetve a dokumentumtípusok feltártsága különbségeiből adódó eltéréseket. Ezt nehéz elvégezni, mert sok adatbázis vagy nem használ egy dokumentumtípus mezőt, vagy nagyon következtelenül teszi azt.

A LISA adatbázisban például nincs ilyen mező, így a cikkeket, konferencia előadásokat és szemléket nem lehet egyértelműen megkülönböztetni. Az ISA-ban van dokumentumtípus mezőt, de az nem megbízható. Például a konferencia előadások rekordjai „monográfia fejezete”-ként („monographic chapter”) jelennek meg az egyik évben, „monográfia”-ként („monographic”) a következőben, majd 1999-től „folyóiratcikk”-ként („journal articles”). Meg kell érteni, hogy a nyelv és dokumentumtípus szerinti feltártság, hasonlóan a retrospektivitáshoz, fontos előny lehet egyes használók számára, és ezeket is vizsgálni kell, amikor az adatbázis összetételét és a feltárt időtartományt tesztelik.

Ezek a megfontolásokon túl sem szabad figyelmen kívül hagyni, hogy néhány kevésbé nyilvánvaló feldolgozási jellemző torzíthatja az összehasonlító eredményeket, mint például azok a különbségek, ahogyan az indexeket építik. Ha olyan összehasonlításokat teszünk, amelyek a címek szavaira vonatkozó kereséseken alapulnak, fontos ellenőrizni, hogy a címindexeket ugyanabból (ugyanazokból) a mező(k)ből építik-e minden tesztelt adatbázisban. Ha az egyik adatbázis a címek indexét a címek és az alcímek mezőiből építi, az torzíthatja az eredményeket. Ha egy harmadik adatbázis a címkiegészítés mezőt vagy almezőt is használja, a különbségek sokkal jelentősebbnek látszhatnak, mint amilyenek valójában. (A címkiegészítések nagyon hasznosak, mert információkat adnak nem kifejező címekre, és segítenek a keresésben és a eredménylisták áttekintésében. Az H. W. Wilson cég például nagyon informatív címkiegészítéseket szolgáltat, ahogy azt a 6. fejezetben tárgyaljuk.) Az időszaki kiadványok címjegyzékei esetében, függően az adatfájl kivitelezésének módjától lehet, hogy a címindex csak a főcímet és a kulscímet tartalmazza az egyik adatbázisban, míg a másikban tartalmazhatja a korábbi, következő, alternatív, párhuzamos, gerinc- és eltérő címeteket.

Hasonló okok magyarázhatják a művészek neve alapján végzett tesztkeresések eredményeinek hatalmas különbségeit zenei címjegyzékekben vagy katalógusokban. A legendás zenészre, Mark Knopflerre vonatkozó keresés 14 találatot ad az Amazon.com-nál (3.8. ábra) és 112-t a Borders.com-nál (3.9. ábra). Az albumok hatalmas számának oka a Borders.com-nál az, hogy ebben megtalálni minden olyan albumot, amelyeken Mark Knopfler vendégművészként szerepelt, mert a művészek indexébe fölveszik a megjegyzések adatmezőt is, amely tartalmaz ilyen információkat. A referáló és indexelő és a teljes szövegű adatbázisok nem problematikusak ebből a szempontból, bár, ahogy korábban említettük, azok az adatbázisok, amelyek kiegészített címeket használnak, mint például az H.W. Wilson adatbázisai, feltétlenül előnyben lesznek a címek szavaira vonatkozó kereséseken alapuló összehasonlítások esetén.



3.8. ábra: A Mark Knopflerre mint művészre vonatkozó keresés eredményei az Amazon.com-nál



3.9. ábra: A Mark Knopflerre mint művészre vonatkozó keresés eredményei a Borders.com-nál

Az adatbázisok kiadói néha információt adnak arról, hogy a feltárt anyag szempontjából mit tartanak adatbázisuk különös erősségének. Ez azonban nem feltétlenül pontos. Az IFI/Plenum, az Information Science Abstracts korábbi kiadója honlapján (és a DIALOG adatbázis-leírásában és reklámkiadványában) 27 olyan témát sorolt fel, amelyekben „az ISA az ön elsődleges referenzs forrása” (3.10. ábra). Ez az állítás első pillantásra is valószínűtlennek és irreálisnak látszik. Olyan, mintha a Haworth Press azzal kérkedne katalógusában, hogy ő a dinamikus objektum modellekről és az állandó web kapcsolatról szóló könyvek elsődleges forrása.

A lista, amit most revízió alá vettek, csordultig volt divatos frázisokkal, amelyekkel elárasztanak a COMDEX sajtótájékoztatóin. Az adatbázis már nem. Egy keresés olyan divatos szavakra, mint a *firewall* (tűzfal) és a *listserver* (két szóként és többes számban is keresve) 2, illetve

3 rekordot adott az egész adatbázisban. A Library Literature-ben és a LISA-ban szintén csupán két rekordja volt, de azok előállítói sohasem állították, hogy listserverre vagy a firewallra vonatkozóan az ő adatbázisuk „az ön elsődleges referenz forrása”. Bár az adatbázis új vezetősége kicserélte a tárgyi feltártságra vonatkozó abszurd listát egy elfogadhatóbbal, az sem mentes félrevezető állításoktól, ahogy a korábbi *supercomputing* példa illusztrálja, és néhány további példa fogja illusztrálni. A kifejezést *supercomputing*ról *supercomputers*re változtatták az adatbázis átvétele utáni három revízió egyike során, de csak egyetlen árva rekordot adtak hozzá az adatbázishoz erről a témáról, ahogy azt egy későbbi keresés bizonyította, amely a cím, referátum és deszkriptor mezőkben keresett.

Könnyű észrevenni az ilyen extrém blöfföket egy gyors és egyszerű keresés elvégzésével a vizsgált adatbázisban. Több munkát igényel, amikor a téma szerinti keresés nagyszámú rekordot eredményez. A halmaz nagyságát megfelelő perspektívába kell állítani azáltal, hogy összehasonlítható adatbázisokban, összehasonlítható indexekben és összehasonlítható időtartamokra végzünk tesztkereséseket. Azonos feltételek biztosításához az *összehasonlítható* fontos szó itt. A kulcsszó szerinti keresés - vagy a DIALOG szóhasználata szerint az alapszótárban végzett keresés - egy teljes szövegű adatbázisban nagyságrenddel több találatot ad, mint egy referáló/indexelő adatbázisban. Ugyanakkor egy referáló/indexelő adatbázisban sokkal nagyobb az esély arra, hogy sokkal több, a keresési kifejezés(ek)e)t tartalmazó rekordot találunk, mint egy csak indexelő adatbázisban.

A deszkriptor mezőben való keresés szintén torzíthatja az összehasonlító keresés eredményeit. Aránytalanul azt az adatbázist részesíti előnyben, amely a keresési fogalmat deszkriptorként használja, szemben azokkal, amelyek bővebb, szűkebb vagy szinonim fogalmat használnak, vagy a tesztkérdésben használt szó helyett eltérő helyesírási formát alkalmaznak. Ez a helyzet például, amikor az egyik adatbázis használja a *CD-ROM* deszkriptort, míg a másik *optical disk* (optikai lemez) kifejezést használja, ahogyan az ERIC adatbázis. A hozzáadott értéket tartalmazó elemek, mint például a referátumok és a specifikusabb fogalmak használata nagyon fontos lehet természetesen, de saját jogán kell értékelni, a 10. fejezetben tárgyalt, az indexelés minőségére vonatkozó indexelési feltételek szerint.

Egyes témák esetében az ilyen keresések sem lehetnek tökéletesek. Az amerikai szerzők elsődlegesen a *postpartum depression* kifejezést használják, európai és brit nemzetközösségből származó szerzők gyakran a *puerperal depression* kifejezést preferálják. Ebből következően az előbbi kifejezés használata előnyös a MEDLINE számára az EMBASE-zel szemben (amelynek szélesebb az európai forrásbázisa), a második fogalom esetén fordított a helyzet. A *puerperal depression* kifejezés 352 rekordot talál az EMBASE-ben és csak 22-t a MEDLINE-ban. A *postpartum depression* kifejezés 149 rekordot talál az EMBASE-ben és 218-at a MEDLINE-ban.

Hasonló okokból a kölcsönzési rendszerre az *issue system* vagy *loan system* brit fogalmak használata (s nem az amerikai *circulation system*-é) a LISA-t favorizálná (372 rekord) az ISA-val szemben (87 rekord). A keresésnek a cím mezőre való korlátozása még nagyobb arányú eltérést eredményez: 74 rekordot a LISA-ban és 13-at az ISA-ban. Természetesen ez indirekt módon fényt vet az adatbázisokban feltárt források földrajzi megoszlására, de vannak más megközelítések is erre a célra. Jó kompromisszumot érhetünk el, ha mind a brit, mind az amerikai fogalomra és variációikra is keresünk.

ISA.

AZ ÖN ELSŐDLEGES REFERENZS FORRÁSA

Az Information Science Abstracts (ISA) a nemzetközi szakirodalom és a kutatások adatbázisa - a robbanásszerűen növekedő információ-tudományi ipar élvonalában van, áttekintve számos témát, köztük a következőket:

Szerzői jog
Digitális könyvtárak
Dokumentumszolgáltatás
Elektronikus és asztali kiadás
Szakértői rendszerek
Tűzfalak
Zsilipek
Képi rendszerek
Az információs ipar szabványai
Információs erőforrások menedzsmentje
Információs szolgáltatások
Intelligens ágensek
Interoperabilitás
Listaszerverek
Orvosi informatika
Multimédia és hipermédia
Hálózatosítás és hálózatok
Online keresés
Személyes információs rendszerek
Rekordkezelés
Szkenelési technika
Keresőmotorok
Szuperszámítógépek
Rendszer-rendszer interfészek
Távközlés
Virtuális valóság
World wide web

ISA YOUR PRIMARY REFERENCE SOURCE FOR

Information Science Abstracts (ISA)—a multi-format database of international literature and research references—is on the cutting edge of the exploding information science industry, surveying an extensive range of topics that include:

- Copyright
- Digital libraries
- Document delivery
- Electronic and desktop publishing
- Expert systems
- Fire walls
- Gateways
- Imaging systems
- Information industry standards
- Information resources management
- Information services
- Intelligent agents
- Interoperability
- List servers
- Medical informatics
- Multimedia and hypermedia
- Networking and networks
- Online searching
- Personal information systems
- Records management
- Scanning technology
- Search engines
- Supercomputing
- System-system interfaces
- Telecommunications
- Virtual reality
- World wide web

3.10. ábra: Az ISA divatos kifejezéseket tartalmazó oldala

Hogy lemérjük az állítólagos tárgyi feltártságot azokban a témákban, amelyekkel kapcsolatban az ISA azt állította, hogy „az élvonalban” vannak”, a DIALINDEX adatbázist használtuk 10 adatbázis összehasonlítására. A DIALINDEX-ben megtalálható az összes adatbázis index fájlja, de maguk a rekordok nem. Ez különösen hasznos adatbázis ahhoz, hogy megtaláljuk a legígéretesebb adatbázisokat bármilyen, néhány kulcsszóval leírható téma esetében. A DIALINDEX a DIALOG-ban található fajlszámok szerint azonosítja az adatbázisokat: INSPEC (2), COMPENDEX (8), ABI/INFORM (15), LISA (61), PASCAL (144), Trade & Industry Database (148), ISA (202), Microcomputer Abstracts (233), Computer Database (275) és Library Literature (438).

A DIALOG rangsorolási számokat ad az adatbázisoknak minden egyes kérdéshez kapcsolódóan. Az 1-es számot kapja az, ahol a legtöbb találat van az adott kérdésre vonatkozóan, 2-t a második legnagyobb, és így tovább. Ha döntetlen van két adatbázis között, akkor a DIALOG a fájlszámok növekvő sorrendjében rangsorolja őket. Például, ha a LISA és az ISA adatbázisokban ugyanannyi találat lenne, s holtversenyben állnának a 6. helyen, akkor a LISA (File 61) kapja a 6-os számot és az ISA (File 202) kapja a 7-est. Ez kis mértékben torzíthatja a rangsort, így az értékelőnek azt lehet tanácsolni, hogy törölje ezt az automatikus rangsorolást, s ugyanazt a ranglista számot adja azoknak az adatbázisoknak, amelyek ugyanannyi találatot adtak.

Az ISA szakterületi listáján szereplő divatos kifejezéseket csak minimális mértékben módosítottuk a kérdésekben, hogy szerepeljenek az egyes és többes számú formák, kötőjeles és egybeírt változatok és a legvalószínűbb helyesírási változatok. A listáról minden második kifejezést kiválasztottunk. Minél magasabb a ranglistaszám, annál rosszabb az adatbázis relatív pozíciója. Az összesített ranglistaszámot úgy kaptuk meg, hogy az egyes kérdések ranglistaértékét összeadtuk. A 3.11. ábra azt mutatja, hogy az ISA (File 202) sohasem érte el az 1-es, 2-es vagy 4-es ranglistaszámot; a 3-as, 5-ös, 6-os és 7-es számot egyszer érte el, a 8-ast kétszer, 9-est és 10-est háromszor. Ezzel a ranglista legaljára került ebben az összehasonlításban, ahol pedig abszolút mértékben rendelkezett a hazai pálya előnyével.

Elmerenghetünk, hogy ez az adatbázis vajon hogyan boldogul olyan szakterületeken, amelyek nem specialitásai. Az ilyen anomáliák természetesen az elemzőt is kíváncsivá teszik, hogy vajon ezek az alaptalan állítások érvényesek-e az egész (adatbázis) családban is. Az IFI/Plenum Mental Health Abstracts (MHA) adatbázisában végzett tesztkeresések alapján úgy tűnik, hogy ott is ez a helyzet.

DIALINDEX RANK ORDER										
	File Number									
	2	8	15	61	144	148	202	233	275	438
ti										
digital libraries	2	6	8	3	1	5	7	10	9	4
electronic and desktop publishing	3	10	6	5	8	2	9	4	1	7
fire wall	5	6	3	8	7	1	9	4	2	10
imaging systems	1	3	6	10	4	2	8	7	5	9
information resources management	5	9	7	4	8	1	3	10	2	6
intelligent agents	1	5	6	8	3	2	9	7	4	10
list servers	4	10	5	8	9	2	6	3	1	7
multimedia and hypermedia	3	5	6	8	7	1	10	4	2	9
online searching	5	7	9	2	6	5	3	8	10	1
record management	3	9	1	4	8	2	5	10	7	6
search engines	5	9	3	6	8	1	10	4	2	7
virtual reality	2	4	6	9	5	1	8	7	3	10
world wide web	3	7	4	5	8	1	10	6	2	9
Total title score*	41	90	70	80	82	26	97	84	50	95
Compound rank order	2	8	4	5	6	1	10	7	3	9

* The lower the better!

Összesített pontszám*

Összetett rangsor

* Minél alacsonyabb, annál jobb!

3.11. ábra: A divatos kifejezések eredményének mátrixa

Az MHA „verseny társai” közé a PsycINFO, az EMBASE és a MEDLINE adatbázisok tartoztak. A fájl előállítójának a honlapja itt is megemlíti specifikus szakterületeket, ahol „az MHA ismert kiváló feltártságról” (3.12. ábra)

IFI/Plenum Data Corporation

Mental Health Abstracts

Bibliográfiai adatok a viselkedésről, mentális egészségről és mentális betegségről.

Most elérhető CD-ROM-on.

A *Mental Health Abstracts* több mint 500000, a mentális egészséggel és mentális betegségekkel kapcsolatos cikk bibliográfiai adatait tartalmazza. Több mint 1000 periodikum cikkeinek referátumait és a hozzájuk tartozó indexfogalmakat tartalmazza, valamint könyveket, kutatási jelentéseket és kongresszusi kiadványokat.

Egyedi feltártság

Ebben az adatbázisban megtalálhatja sok olyan közlemény adatait, amelyek nincsenek benne más adatbázisokban. A Mental Health Abstracts különösen ismert arról, hogy kiválóan feltárja a következő területeket:

Pszichofarmakológia

Pszichiátriai kezelés

A mentális betegségek társadalmi és jogi aspektusai

Törvényszéki orvostudományi szakirodalom



3.12. ábra: A tárgyi feltártsággal kapcsolatos állítások a Mental Health Abstracts-ben

A tárgyi fogalmakat kis mértékben módosítottuk hogy lehetővé tegyük a szóvégződések variációit, mint például a *psychopharmacolog?*, amely visszakeresi a *psychopharmacology*, *psychopharmacologist(s)*, *psychopharmacological* stb. szavakat. Az MHA a négy terület közül háromból az utolsó helyen végzett, annak ellenére, hogy ezekben kellett volna a legjobban szerepelnie. Az MHA által két legfontosabbnak mondott szakterületre vonatkozó keresési eredményeket a 3.13. ábra illusztrálja. Az MHA találatainak száma feltűnően alacsony volt, figyelembe véve azt, hogy 1967-ig megy vissza az adatbázis, - sokkal régebbre, mint az EMBASE adatbázis. Amikor a keresést az utóbbi 10 évre korlátoztuk, az MHA számai még lehangolóbbá váltak.

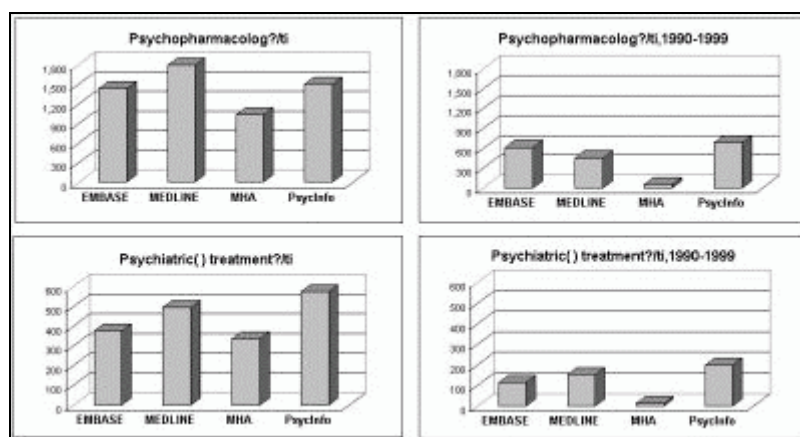
Bár a költségek témáját a 12. fejezetben tárgyaljuk, itt kell megemlíteni, hogy az MHA-nál 150%-kal magasabb volt a kapcsolati idő díja, mint a PsycINFO-nál, és a nagyon jól teljesítő MEDLINE adatbázis már elérhető volt ingyen is (bár nem a DIALOG-nál). Újra csak azt lehet mondani, hogy a régi bölcsesség, miszerint azt kapod, amiért fizetsz, gyakran nem

érvényes az adatbázisokra. Az sem jelenthet sokat, hogy a fájl előállítójának neve milyen elismertségnek örvend más területeken. Az IFI/Plenum anyacége például tekintélyes kiadó. Hasonló módon az Elsevier a tudományos folyóiratok legnagyobb kiadója, de Scirus címjegyzéke, amely minden oldalán azt állítja, hogy „csak tudományos információ”-ra szolgál, a használókat felületes és teljesen vulgáris weblapok tízezreihez vezeti el, ahogyan azt a <http://www2.hawaii.edu/~jacso/extra/> címen láthatjuk illusztrálva.

A szerzőnek ezen a weblapján található egy illusztrált útmutatót is a Claritas drága Population Demographics adatbázisának nonszensz adataihoz. A statisztikai adatok, amelyek csaknem 6000 nulla lakosú települést tartalmaznak, különös statisztikai mérésekre utalnak, hogy mértéktartóan fogalmazzunk.

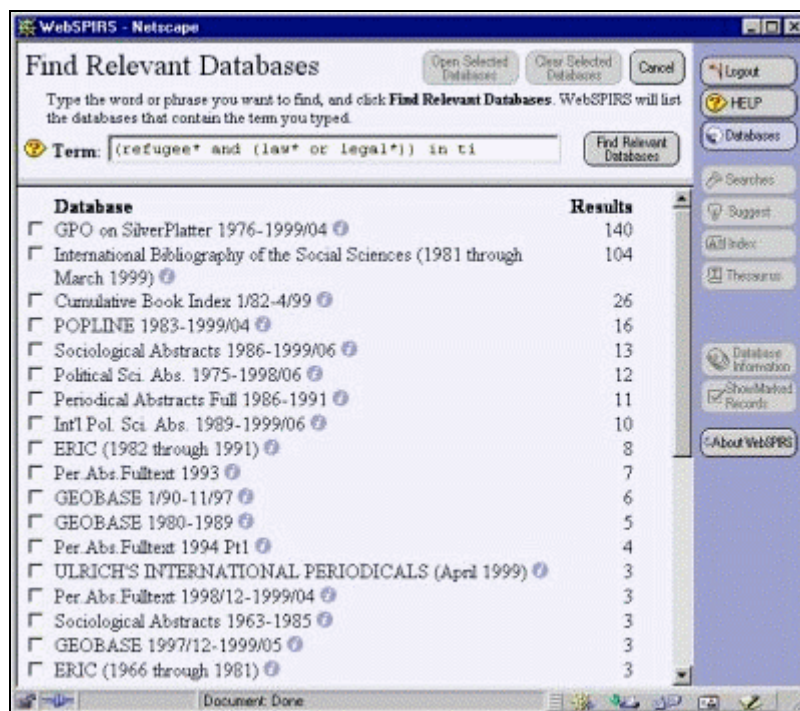
Ilyen tesztek több olyan témában elvégezhetőek, amelyek a célközönség elsődleges érdeklődéséhez kapcsolódnak. Ez azt a benyomást adhatja, hogy az adatbázis tárgyi feltártsága pertinens egy meghatározott használó vagy használói csoport számára - függetlenül attól, hogy mi a tesztelt adatbázis állítólagos tárgyi feltártsága. Ha valaki meghatározott szakterületen ismeri a használói közönség preferenciáit, a tesztkeresések könnyen korlátozhatóak időszakra, a kiadás országára vagy nyelvre.

Például a humán tudományok kutatói ismertek arról, hogy a számítógéptudósoknál jobban kedvelik a 10-15 évvel ezelőtt kiadott könyveket; ezért a tesztek módszereit a szakterületet és a megcélzott használói csoportok igényeihez kell igazítani. Ezt illusztrálják az 5. fejezetben az adatbázis dimenzióira és a feldolgozott forrásdokumentumokra vonatkozó részek a kiadványok megjelenésének országa és nyelve szempontjából.



3.13. ábra: A DIALINDEX-ben végzett keresések eredményei

Az ilyen típusú tesztelés a DIALOG, DataStar és WebSPIRS rendszerekben a legkönnyebb, mivel speciális parancsaik és adatbázisaik vannak arra, hogy egyszerre az összes adatbázis indexében keressenek. A DIALOG ezt DIALINDEX-nek hívja, a DataStar CROS-nak, s a WebSPIRS-nek is van speciális funkciója (Find Database) az ilyen keresések elvégzéséhez (3.14. ábra)

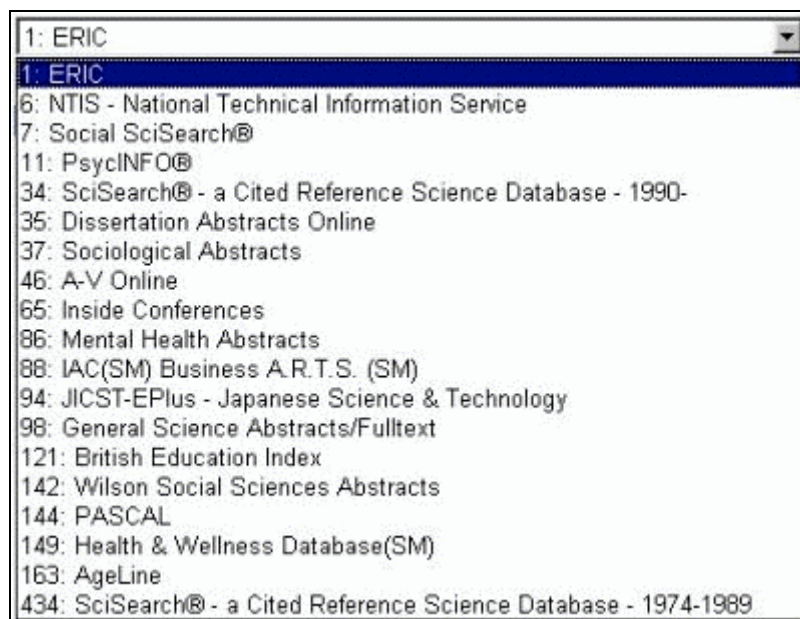


3.14. ábra: A Find Database funkció a WebSPIRS-ben.

A keresés elvégezhető több adatbázisban anélkül is, hogy a közös indexben keresnénk, de ez természetesen tovább tart és többbe kerül. A DIALOG megkönnyíti az ilyen több fájlban történő keresést azzal, hogy közel 500 adatbázisát egy vagy több tárgyi csoportba sorolja be. Ezeknek a csoportoknak saját neve van, mint például PSYCHOLOGY és LIBRARY AND INFORMATION SCIENCE. A csoportok nevének rövidített változata (PSYCH és INFOSCI a mi példánkban) használható mind a DIALINDEX-ben, mind a OneSearch keresésekben. Az utóbbi a keresést a csoport által meghatározott valódi adatbázisokban végzi el.

A felhasználókat figyelmeztetni kell arra, hogy ne bízzanak kizárólagosan a DIALOG előre meghatározott tárgyi adatbáziscsoportjaiban, mert azok gyakran tartalmaznak majdnem teljesen irreleváns adatbázisokat s hagynak ki nagy mértékben relevánsakat. A pszichológia szakterületén például a DIALOG PSYCH adatbáziscsoportja felsorolja a British Education Indexet, a National Technical Information Service (NTIS) adatbázist és a A-V Online-t - nem kifejezetten olyan adatbázisok, amelyekről az embernek eszébe jut a pszichológia.

A DIALOG ugyanakkor kihagyja ebből a csoportból a nagymértékben releváns MEDLINE és EMBASE adatbázisokat (3.15. ábra). Amikor valaki a józan eszt használja az adatbázisok kiválasztásában és a keresőkérdés megfogalmazásában, az ilyen tesztkeresések jó kiindulópontot adhatnak ahhoz, hogy továbbmenjünk az értékelésben és megvizsgáljuk az adatbázisok más jellegzetességeit, méretüket és összetételüket, feltárt forrásaikat és aktualitásukat.



3.15. ábra: A rosszul meghatározott PSYCH adatbázis csoport a DIALOG-nál.

4. Az adatbázis dimenziói

Egy adatbázisról szerzett első benyomásunk nem különbözik attól, amit egy emberről szerzünk. Ahogy elsőre észrevevesszük, milyen magas, milyen nagy és milyen idős az a személy, az első kiadói reklámból is megtudjuk, milyen nagy az adatbázis és időben meddig megy vissza. Ahogyan téves lehet egy személlyel kapcsolatos, rövid találkozón alapuló első benyomásunk, ez lehet érvényes egy adatbázissal kapcsolatos első benyomásunkra is. Ezek a jellemzők fontosak lehetnek, de csak akkor, ha az ember a maga összefüggésében látja, hogyan számították ki a gusztusos broszúrákban büszkén jelentett méreteket, s mi van azok mögött a számok mögött, amelyekkel egy adatbázist le akarnak írni.

Méret és összetétel

A rádiók és televíziók reklámjai, a nyomtatásban megjelent hirdetések és a *Guinness rekordok* könyve ahhoz szoktathatott hozzá sok embert, hogy azt higgye, hogy ami nagyobb, az jobb is. Nem véletlen, hogy az adatbázisokat hirdető anyagok is elsősorban az adatbázis méretével kapcsolatos állításokat állítják előtérbe. A józan ész azonban arra figyelmeztet, hogy a legnagyobb nem egyenlő a legjobbal. Lehet, hogy Oroszországnak van a legnagyobb kereskedelmi repülőgép flottája Európában, de légitársasága a legrosszabbak közé tartozik a biztonság, kényelem, a szolgáltatás minősége és a pontosság szempontjából. A legnagyobb tengeri különlegességeket hirdető étteremben nem szükségszerűen a legjobb tintahalat, osztrigát, garnélarákot szolgálják fel. Sok múlik természetesen az egyedi preferenciákon abból a szempontból, hogy mi teszi a tengeri specialitások éttermét a legjobbá, de az étterem mérete nem feltétlenül alapvető követelmény. Ugyanez érvényes az adatbázisokra is.

A legnagyobb orvosi adatbázis, amely nem dolgozza fel Kelet-Ázsia folyóiratait a megfelelő mértékben, nem lehet a legjobb annak a számára, aki ennek a régióknak a fejlődéséről vagy klinikai jelentéseiről keres információt. A legnagyobb cégjegyzék kevés értékkel bírhat egy olyan cég számára, amely üzletének nagy részét Kelet-Európa országaival bonyolítja le, ha az adatbázis nem tartalmazza Lengyelország, Magyarország, Szlovákia, Csehország és Románia cégeit.

Elég beszédes az adatbázis méretére vonatkozó megszállottsággal kapcsolatban, hogy a Barnes & Noble perrel fenyegette meg az Amazon.com-ot, mert az a világ legnagyobb webes könyvesboltjának mondta magát. Az Amazon.com valóban sok webhasználót megnyert a maga számára úttörő vállalkozásával, első osztályú szoftverével és kiváló szolgáltatásaival, sokkal hamarabb, mint hogy a Barnes & Noble-nak egyáltalán eszébe jutott volna a webre menni, bár most az utóbbi jogosan büszkélkedik az exkluzív „a világ legnagyobb online könyvkereskedője” szlogennel. Mihelyt a Barnes & Noble, ahelyett hogy a jogászai által kezelt dolgokat pénzelné, elkezdett áldozni a jó programozásra és figyelmét a méret helyett más kérdéseknek szentelte, rendkívüli mértékben növelte adatbázisa tartalmának funkcionalitását. Időközben az Amazon.com kiterjesztette tevékenységét sok területre, a játékoktól a konyhai edényekig, ugyanakkor könyvadatbázisának mérete sem menthette meg részvényárának zuhanásától.

Elég sokat mond el a mérettel kapcsolatos rögeszmékről az, hogy a Northern Light keresőgépében végzett keresés 11153 olyan weblapot talált, amely tartalmazza a *largest database (a legnagyobb adatbázis)* kifejezést. Az eredménylista első oldala alapján olyan weboldalak tartalmazzák a legnagyobb adatbázis kifejezést, mint 1, éttermek listája, 2, ígehirdető és tanító videók, 3, kölcsönzők Oregonban, 4, partiképes aggregények, 5, thaiföldi ingatlanok, 6,

magánhirdetések és 7, ejtőernyős vízisízéssel kapcsolatos információk. Maga a *largest* (legnagyobb) szó több mint 4,6 millió oldalon fordul elő, eszerint a - nos - legnagyobb keresőgép szerint.

Az adatbázis mérete ennek ellenére hasznos indikátor lehet, ha pontosan és megfelelő összefüggésben értelmezik. Ezt az információt általában könnyű megszerezni a kiadótól vagy magából az adatbázisból. Mindig jobb az adatokat az adatbázisból kinyerni, mint az esetleg már elavult dokumentációból. A legtöbb DIALOG adatbázisban a (tervezett) aktualizálási dátumot hozzáadják a rekordhoz; ezért egy teljesen csonkolt keresés (S UD=?) meg fogja mutatni az adatbázis pillanatnyi méretét. A SilverPlatter szoftverét használó adatbázisokban nincs meg a teljes csonkolási lehetőség, de a FIND UD>0 parancs meg fogja adni a rekordok teljes számát. A 4.1. ábra megmutatja annak a keresésnek az eredményét, amely meghatározza az Information Science Abstracts (File 202), Library and Information Science Abstracts (File 61) és Library Literature (File 438) méretét, az S UD=? keresőkérdés alkalmazásával. Az eredmények ennek a három adatbázisnak a méretét tükrözik 1999 utolsó hetében.

File	Database Name	Hits
<input type="checkbox"/> 61:	LISA (Library & Information Science Abstracts)	211345
<input type="checkbox"/> 202:	Information Science Abstracts	208136
<input type="checkbox"/> 438:	Library Literature	202849

There are **3** databases matching your statement 'S UD=?'.

4.1. ábra: Három könyvtár- és információtudományi adatbázis mérete

Egy multidiszciplináris adatbázisban is meg lehet határozni egy megfelelő részhalmaz méretét, ha van olyan adatelem, amely egyértelműen azonosítja a részhalmazt. Ez a helyzet a Social SciSearch adatbázis (File 7) és a Trade & Industry Database (File 148) esetében. A 4.2. ábra megmutatja, hogy ezeknek az adatbázisoknak az könyvtár- és információtudományi részhalmaza összehasonlítható a három előző adatbázissal. A Bell & Howell szintén jól feltárja az könyvtár- és információtudomány irodalmát az ABI/INFORM és a Periodicals Abstracts PlusText adatbázisokban, de egyik sem határozza meg a kategóriát, amelyhez a rekordok tartoznak, ennél fogva nem lehet megállapítani ezen adatbázisok könyvtár- és információtudományi részhalmazának méretét.

Ranked DIALINDEX® Results		
Your select statement is 'S IN=LIBRAR? OR SC=LIBRAR?'		
<input checked="" type="checkbox"/> Select All	<input type="checkbox"/> Clear Selections	Begin Databases
File	Database Name	Hits
<input type="checkbox"/> 148:	Gale Group Trade & Industry Database(TM)	224268
<input type="checkbox"/> 7:	Social SciSearch®	145152

4.2. ábra: Az információ- és könyvtártudományi
részhalmoz méretének megállapítása a Social
SciSearch és a Trade & Industry adatbázisokban

Azokban az adatbázisokban, amelyeknek nincs külön meghatározott aktualizálás mezője, más technikákat kell alkalmazni. A legtöbb adatbázisban általában van néhány olyan adatelem, amely nagy valószínűséggel minden rekordban megtalálható, s ezek értékeinek száma korlátozott. A Bowker cég címjegyzékeinek saját változatában használt szoftverben nincs aktualizálás mező, de vannak más adatmezők, amelyekről feltételezni lehet, hogy minden rekordban jelen vannak. Például az Ulrich's adatbázisban (Ulrich's International Periodicals Directory) a periodikum státusza aktív, megszűnt vagy nem igazolt lehet (active, ceased és unverified), ezért az SS=\$ keresés valószínűleg a rekordok teljes számát eredményezi. A Bowker Who's Who adatbázisaiban a felvett személyek neve jó keresési kritérium lehet az adatbázis méretének meghatározásához. Elég biztosan feltételezhető, hogy a benne szereplő személyek vagy férfiak, vagy nők, minden rekordnak van a nemre vonatkozó információja, még akkor is, ha esetleg ez az adat „nem igazolt” (unverified) vagy „meghatározatlan” (undetermined). A használatot figyelmeztetni kell arra, hogy sok adatbázisban rekordok tízezreiben nincs meg a feltételezett adatelem, ahogy azt majd a 9. fejezetben a rekordok teljességére vonatkozóan megtárgyaljuk.

A teljes csonkolás működhet olyan mezők esetében is, amelyek gyakorlatilag mindig jelen vannak, mint például a cím mező, de nagyon hosszú időbe telhet egy TI=\$ vagy hasonló keresés végrehajtása a Bowker saját CD-ROM implementációiban, mert az egyedi tételek száma egy nagy adatbázisban több millió is lehet. Elvégezhető azonban, s jelen szerző el is végezte egy gyors számítógéppel és a szoftver gyors windowsos változatával, ahogy azt a 9. fejezet illusztrálja. Előre figyelmeztetjük, hogy a DOS-os verzióban a képernyő lefagyhat meghatározott mennyiségű inaktív idő után; azaz a keresés befejeződhet, míg a kereső távol van.

Számos adatbázisban vannak mezők két értékkel, például angol és nem angol nyelvű, magán vagy állami cégek, vagy egy pszichológiai adatbázisban a populáció Human és Animal (emberre és állatra vonatkozó) lehet. Ha ezeket következetesen hozzáadják minden rekordhoz, akkor a méret meghatározható egy egyszerű OR-ral összekötött keresőkérdéssel ebben a mezőben: S LA=(English or Non-English). Azok a CD-ROM adatbázisok, amelyek a Dataware Technologies szoftverét használják (mint például a Historical Abstracts), általában képesek *none* és az *all* (egyik sem, mind) keresőfogalmat használni egy mezőre, s ezzel is meghatározható egy adatbázis mérete. Szoftverének buildware-összetevője generálja ezt a két speciális indextételt minden indexelt mezőre.

Ha egy rekordban mondjuk a kiadási év jelen van, akkor a kiadási év indexének *all* pszeudo indexében található tételek száma növekszik eggyel. Ha a mező hiányzik, akkor a pszeudo index *none* tételeinek száma növekszik eggyel. Ugyanez történik minden indexelt mező pszeudo index tételeivel. Mivel ezeket az index tételeket akkor generálják, amikor az

adatbázist létrehozzák, a kereséshez csak egy másodpercre van szükség, amikor előbb az *all*, azután pedig a *none* keresést végezzük el, akár még a cím mező esetében is.

A legtöbb olyan keresőszoftver, amelyet közvetlenül a web számára fejlesztettek ki (ahelyett, hogy átvitték volna a hagyományos online változathól), nem ajánl fel lehetőséget az adatbázis méretének meghatározására a fenti módszerek valamelyikével. Bár a kiadási év kereshető, általában nem alkalmazható magában, hanem csak korlátozó mezőként egy tárgy vagy szerző szerinti keresést kiegészítve. Például akkorának kell elfogadnunk az adatbázisok méretét, amit az Amazon.com és a Barnes & Noble állít magáról. Bár az Amazon.com esetében a Kiadó/Dátum keresési űrlap lehetővé teszi egy meghatározott év előtt, során vagy után kiadott könyvek keresését, amikor a 2001 előtt (before 2001) opciót önmagában választjuk, az az üzenet jön válaszként, hogy a kiadó nevét vagy egy kulcsszót is meg kell adni. Hasonló módon a Barnes & Noble is felajánlja a keresést három árkategória szerint, s ugyanígy média típusa szerint (papírkötésű, keménykötésű, nagybetűs, audiokönyv), amelyek tökéletes jelöltek lehetnek az adatbázis méretének meghatározására. Sajnos ezek a keresési feltételek csak egy címhez, szerzőhöz vagy tárgyhoz kapcsolódó korlátozóként használhatóak.

Önmagában nem elég, ha ismerjük az adatbázis méretét. Ezt a számot az adatbázis összetételének függvényében kell vizsgálni. Ahhoz, hogy egy adatbázis sokatmondó méretét megítéljük, tudnunk kell, hogy a rekordokat milyen egységek számára hozták létre. A G. K. Saur World Databases címjegyzéke (WDNBS) egy rekordot hoz létre minden adatbázis minden egyes változatára - plusz még egy mester rekordot. Ez felduzzasztja a rekordok számát a World Databases adatbázisban. A Gale cég Directory of Databases adatbázisában új rekordot képez minden egyes médiumra, amelyben az adatbázis megtestesül (azaz egy rekord van a mágnesszalag változat(ok)ra, egy a CD-ROM változat(ok)ra és egy az online változat(ok)ra, de nincs külön-külön rekord minden egyes CD-ROM vagy online változatra).

A Mental Health Abstracts adatbázis esetében a Gale címjegyzékében egy rekord van az adatbázis online és egy másik a CD-ROM változatára (4.3. ábra). A WDBS-ben külön-külön rekord van a CompuServe-en, a Westlaw-n és a DIALOG online adatbázis változaton, plusz egy mester rekord (4.4. ábra). Ez nagyon felhigítja az adatbázis méretét, különösen azért, mert a CompuServe és a Westlaw is a DIALOG-nak passzolja a keresőkérdést, hogy ezt az adatbázist szolgáltatassa. Mellesleg a WDBS nem tud az MHA adatbázis CD-ROM változatának létezéséről.

A másik komoly és fontosabb következmény, hogy sokkal drágább a G. K. Saur adatbázisában keresni, ahol minden egyes rekord megjelenítése négy dollárba kerül és az ingyenes találati lista semmi információt nem ad arról, hogy melyik rekord melyik változatára vonatkozik (4.4. ábra). Egy kitalálós játék az, ahogy megjeleníti az MHA adatbázis DIALOG, Westlaw vagy CompuServe változatának rekordjait.

Tovább növeli a gondot, ha az ember rájön, hogy mennyire hihetetlenül pontatlanok az információk a World Databasesben, s hogy nagyrészt ugyanazt a nonszensz információt viszik át egyik rekordból a másikba, egyenként négy dollárért. Ez a helyzet akkor, amikor minimális szerkesztői odafigyelés vagy hozzáértés elég lett volna a legszembetűnőbb hibák észrevételéhez, mint egy adatbázis részhalmazának kiadása esetében. Nagyon gyakran a rekordok közötti egyetlen különbség a fájl és a terjesztő neve.

DIALOGWEB Command Search new search databases

1. ☐ 1/TI,MT,DT,CO,VD,UF/1 (Item 1 from file: 230)
Medium Type: Online **Document type:** Database
Mental Health Abstracts
 (MHA)
Former Database Name: NIMH Data Base
Producer:
 IFI/Plenum Data Corporation
Updating: 300 records a month
File Label: 86; PSYC2
Vendor#: Vendor name
 2000002270: The Dialog Corporation
 2000011221: CompuServe Information Service
 Gale Dir Online

2. ☐ 1/TI,MT,DT,CO,VD,UF/2 (Item 2 from file: 230)
Medium Type: CD-ROM **Document type:** Database
Mental Health Abstracts
Producer:
 IFI/Plenum Data Corporation
Updating: Quarterly
Vendor#: Vendor name
 3100002270: The Dialog Corporation
 Gale Dir Online

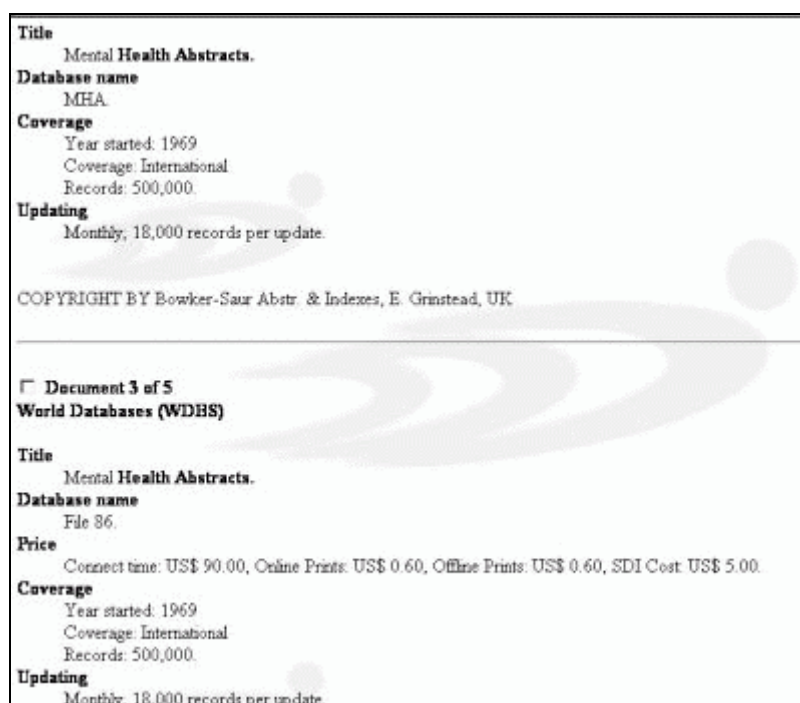
4.3. ábra: Az MHA adatbázis online és CD-ROM változataira egy-egy rekord található a Gale címjegyzékében

A WDBS azt állítja az MHA-ra vonatkozó mind a négy rekord esetében, hogy minden frissítéskor 18000 rekorddal egészítik ki az adatbázist. A szám önmagában is abszurd, s ezt propagálják olyan módon, hogy átviszik az MHA adatbázist leíró egyik rekordból a másikba (4.5. ábra). A pontos szám körülbelül 300 frissítésenként, ahogyan azt a Gale címjegyzéke megadja, s ahogy azt az MHA aktualizálási indexének böngészése is igazolja.

A leírás egységeire vonatkozó ésszerűbb és hagyományosabb döntések is jelentős különbségekhez vezethetnek az adatbázis méretében. Ahogy azt az előző fejezet illusztrálta, az Ulrich's adatbázisban a Bowker a legújabb tétel elvét alkalmazza egy rekord megalkotására minden egyes időszaki kiadvány esetében. Ez azt jelenti, hogy ha egy időszaki kiadvány címe megváltozik, a korábbi cím rekordját törlik és a releváns elemek (cím és ISSN) bekerülnek az új rekordba. Ezzel ellentétben az EBSCO az egymást követő tételek elvét követi, s külön rekordot alkot minden egyes időszaki kiadványra, amelynek a címe megváltozott, s megtartja a korábbi címhez tartozó rekordot is.



4.4. ábra: Az MHA online változatainak négy rekordja a World Databasesben.



4.5. ábra: Abszurd információt visznek át egyik rekordból a másikba a World Databasesben.

Jelentős különbség van abban, hogyan kezelik az adatbázisok azokat a szemlecikkeket, amelyek több műről tartalmaznak kritikát. Például az *EContent* (korábban *Database* címen volt ismert) „Péter's Picks and Pans” rovatában mindig három vagy négy adatbázist ismertet. Ezt a rovatot több adatbázis feldolgozza, közte az ABI/INFORM, a Trade & Industry Database, a Periodical Abstracts PlusText, a Social SciSearch és az ISA. 1997-ig a legtöbb adatbázis egyetlen rekordot készített az egész szemlecikkről.

1997 óta azonban a Social SciSearch kezdett külön rekordot alkotni minden egyes bírált műről, és sok más fájl előállító is átváltott erre a rendszerre (ABI/INFORM, Trade & Industry Database és Periodical Abstracts). 1999-től az ABI/INFORM visszatért a cikkenkénti egy rekord elvéhez. A Trade & Industry Database 1999-ben három rekordot készített az év első rovatának három kritikájához, de csak egyet a másodikhoz. Mások, mint az ISA és a LibLit,

megmaradtak a cikkenkénti egy rekord elvénél, amely hátrányba hozza őket a fájl méretének összehasonlításakor, különösen akkor, amikor egy folyóirat feltártságának mélységét hasonlítják össze.

Ez különösen igaz a Wilson adatbázisokra, amelyekben számos szemlecikk szerepel egyetlen rekorddal cikkenként, nem külön rekorddal a cikkben kritizált minden egyes műről. Például a Readers' Guide Abstractsben több mint 220000 rekord van könyvkritikákból. Sok szemlében több mint egy műről jelent meg kritika, de mivel az H. W. Wilson nem hígítja fel adatbázisa méretét azzal, hogy minden egyes kritizált műről egy rekordot hoz létre, a használóknak nem kell többet fizetnie az idő és különösen a tételenkénti megjelenítés és nyomtatás miatt minden egyes rekord visszakeresésekor.

A szemlecikkek különböző és következtelen kezelése magyarázza meg a keresési eredmények furcsaságait a 4.6. ábrán. A LibLit (File 438) és az ISA (File 202) 1999-ben az első két cikket egy-egy rekordként dolgozta fel. A Trade & Industry Database (File 148) három rekordot hozott létre az első, egyet a második cikkre. A LISA (File 61) nem dolgozott fel egyetlen szemlecikket és egyetlen cikket sem 1999-ben a *Database*-ből egészen az év júliusáig (ez volt az utolsó szám ez alatt a cím alatt). Az ABI/INFORM (File 15) egy rekordot készített 1999 mindhárom szemlecikkére. A Social SciSearch (File 7) összesen 9 rekordot alkotott, egyet-egyet mindhárom cikk három-három kritikájára. Ez apró pont az adatbázisok rekordjainak százezrei között, de amikor a feldolgozottságot a szerzők vagy a folyóiratok szintjén hasonlítjuk össze, a különbség jelentős lehet.

Search History				
Database Details				
Set	Term Searched	Items	File	
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	20		Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	2	438	Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	2	202	Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	4	148	Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	0	61	Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	3	15	Display
S1	AU=JACSO? AND JN=DATABASE? AND PY=1999	9	7	Display

4.6. ábra: A szemlecikkek feldolgozási különbségeinek hatása.

A rekordok teljes száma a Bowker's Complete Video Directoryban félrevezető lehet, ha nem tudjuk, hogy számos rekord van ugyanarról a filmről. (4.7. ábra). Egyes rekordok nyilvánvaló kihagyásain és következtelenségein (mint például az MPAA osztályzatai vagy Gene Wilder mint az elsődleges közreműködő a találati listán), az ár és a rögzítési formátum (VHS kontra Betamax) különbségein kívül néhány megháromszorozott vagy megnégyszerezett rekordban az eltérések ugyanakkor a filmnek fél tucat másik rekordjában csupán lényegtelen apróságok. Ezek a részletek csak videókereskedők számára lehetnek érdekesek, de nem egy audiovizuális könyvtár ügyfelének, ahol ezt az adatbázist gyakran használják. Még ha ezek a különbségek netán mégis fontosak (mint például a rögzítési formátum), jobb lenne egyetlen rekordot létrehozni, amely a különböző kiadások mindenféle sajátosságait szerepeltetné. A 0,00 dolláros ár nagyon vonzó, de nem igaz.

Ez a megközelítés nem felel meg az angol-amerikai katalogizálási szabályoknak (AACR2), de a címtár adatbázisoknak nem kell követniük azokat a szabályokat, amelyeket azért hoztak létre, hogy a könyvtárak adott állományához való hozzáférés eszközei legyenek. Dicséretre méltóan a Bowker elkezdte alkalmazni ezt a közös rekord koncepciót *Books in Print* (BIP)

adattábasának saját CD-ROM változatában, ahol a különböző kiadásokat és kötésváltozatokat egyetlen rekordon belül sorolják fel. Bár ez csökkentette az adattábas méretét, a használóknak előnyére szolgált, és - ismételjük -, az AACR2 szabályok nem érvényesek a címtárak tételeire. Az információ rendezése egyetlen rekordban egyszerűsítette a keresés folyamatát. Sajnos a BIP online verzióinak nem minden kiadója követte ezt a gyakorlatot, azon használók kárára, akikre a tranzakciónkénti díjszabás érvényes. Nekik minden egyes rekordért fizetniük kell még akkor is, ha a lényeges információk (cím, szerző, kiadó stb.) alapvetően ugyanazok.

Title	Contributor	MPAA	Price	Avail.	ISBN/AUPC
R Annie Hall	Allen, Woody		\$49.95	1977	
Annie Hall	Allen, Woody	PG	\$14.95	1977	
Annie Hall	Allen, Woody		\$19.95+	1977	
Annie Hall	Allen, Woody	PG	\$29.95+	1977	0752805313
R Everything You Always W	Allen, Woody	R	\$0.00+	1972	
Everything You Always W	Allen, Woody	R	\$0.00	1972	0752804325
Everything You Always W	Wilder, Gene		\$19.98	1972	
Annie Hall	Allen, Woody	PG	\$19.98		
Everything You Always W	Allen, Woody	R	\$19.98		
Everything You Always W	Allen, Woody		\$39.98		

4.7. ábra: Több rekord ugyanarról a filmről a Bowker cég Complete Video Directoryjában.

Az Internet Movie Database mutatja a legelegánsabb és legtisztességesebb megoldást ugyanazon film több változatának a kezelésére. Egy visszakeresett rekord van az *Annie Hall*-ról (4.8. ábra), s az értéknövelt információkat tartalmazó mezők tucatjai között van egy közvetlen linkkel kapcsolt alrekord, amely akkor jelenik meg, ha a film DVD kiadásának opcióját választjuk (4.9. ábra)

Annie Hall (1977)

Directed by **Woody Allen**

Writing credits **Woody Allen** and **Marshall Brickman**

Genre: **Comedy / Romance**

Tagline: **A nervous romance.**

User Rating: **★★★★★ 8.2/10** (3939 votes) **VOTE**

top 250: #75

Plot Outline: **Neurotic New York comedian Alvy Singer falls in love with the ditzy Annie Hall.**

User Comments: **Surprisingly Good (more)**

Cast overview, first billed only:

Woody Allen	... Alvy Singer
Diane Keaton	... Annie Hall
Tony Roberts (I)	... Rob
Carol Kane	... Albee
Paul Simon (I)	... Tony Lacey
Shelley Duvall	... Pam
Janet Margolin	... Robin
Colleen Dewhurst	... Miss Hall
Christopher Walken	... Duane Hall
Donald Serrin	... Dad Hall
Helen Mirren	... Grammy Hall

4.8. ábra: Az *Annie Hall* mesterrekordja az Internet Movie Database-ben.



4.9. ábra: Az *Annie Hall* DVD-specifikus információjának kapcsolódó rekordja.

Bár az adatbázis méretének felduzzasztása sokszor rosszul elgondolt vagy legalább vitatható koncepción alapszik, a duplikált rekordok jelentős száma a gondatlan vagy hozzá nem értő menedzsment (vagy mindkettő) jele az adatbázis építésében, és a használó figyelmen kívül hagyásáról árulkodik. A CD-ROM-os film címjegyzékek közül a *Cinemanía* messze a legjobb kiváló tartalma alapján, különösen Roger Ebert, Leonard Maltin és Pauline Kael elsőrangú szemléinek köszönhetően. A Microsoft nem duzzasztotta fel az adatbázis méretét azzal, hogy minden kritikához külön rekordot hozott létre. Ezek magának a filmnek az egyetlen mesterrekordjához vannak kapcsolva hiperlinkekkel.

A Corel cég gyatra film címjegyzékében büszkén hirdette magáról, hogy 90000 rekordot tartalmaz, ám elmulasztotta megemlíteni, hogy számos rekord duplikátum vagy akár harmadik rekord, eltérő információkkal a közreműködők listájáról, a film hosszáról és más adatelemekről. Ezen felül a filmek osztályzatai és kritikái szemmel láthatóan olyanoktól származnak, akiket az unalmas „Akarsz napi 300 dollárt keresni filmnézéssel?” szövegű, a buszmegállókba kiragasztott hirdetésekkel toboroztak. Ennek az adatbázisnak az alacsony színvonala annak tulajdonítható, hogy a Corelnek hiányzott a gyakorlata az adatbázis-építésben; néhány évvel később abba is hagyták ezt a tevékenységet.

Különösen frusztráló, ha a használó rájön arra, hogy egy referáló és indexelő adatbázis méretét nagy mértékben megduplázott és megháromszorozott rekordok növelik, ahol nincs is olyan lehetséges kifogás, amely a címjegyzékek esetében talán elfogadható. Még lehangolóbb volt, amikor egy professzionálisnak gondolt adatfájl készítő (például az IFI/Plenum) készített egy fájlt, a könyvtárosoknak és más információs szakembereknek szánt Information Science Abstractset, amely tele volt ugyanarról a cikkről második és harmadik rekordokkal. Erről az adatbázisról készült kritikájában Jacsó (1997e) becslése szerint - amely tesztkereséseken alapult - körülbelül 12000 duplikált pár volt az adatbázisban (és a nyomtatott változatban is).

Bár más adatbázisokban is vannak duplikátumok, ez az arány példa nélküli egy olyan adatbázisban, amely 200000-nél kevesebb rekordból áll. Nagyon nyugtalanító volt, hogy az adatbázis készítőjének fogalma sem volt a duplikátumokról. A cég alelnöke, aki hosszú időn keresztül az adatbázis szerkesztője is volt, azt állította „1987 óta, amikor az IFI átvette az előállítását és kifinomult duplikátum detektáló rendszert kezdett alkalmazni, a duplum és triplum rekordok előfordulása megszűnt” (Allcock 1997).

Valóban így történt? Egy egyszerű ellenőrzés (4.10. ábra, a DIALOG IDO (Identify Duplicates Only, csak a duplikátumok azonosítása) parancsa megmutatta, hogy ez az állítás nem alapulhatott tényeken vagy a valóságon. Az *ASLIB Proceedings*ből származó 67 1990-es cikkek közül 30 (a duplikált párok 60 rekordjából) volt duplikátum. 1991-ben 76 rekordból is 30 duplikátum volt, egy pedig háromszor szerepelt.



4.10. ábra: A duplikátumok detektálásának eredményei

A DIALOG duplikátumokat felderítő algoritmusa nagyon jó, de nem tökéletes, ennek eredményeit nem lehet teljes mértékben elfogadni. Csak a szerzői és a cím mezőt ellenőrzi. Ennek a módszernek a következményeként azokat a tételeket, amelyek egy rendszeresen megjelenő rovatához tartoznak, s így közös címük van, de nincs megkülönböztető alcímük (mint például a Savvy Searching), duplikátumoknak fogja tekinteni. A könyvkritikák rekordjain gyakran egyszerűen a „Book review” cím szerepel. Az ugyanattól a szerzőtől származó könyvkritikák így duplikátumnak fognak minősülni, pedig nem azok. Az emberi szem számára nyilvánvaló, hogy a kötet, szám vagy oldalszám egyértelműen megkülönbözteti az egyik szemlét a másiktól, de a DIALOG duplikátumokat felderítő algoritmusa nem ellenőrzi ezeket az adatelemeket. Az is lehetséges, hogy ugyanaz a szerző publikálta ugyanazt a cikket egy másik folyóiratban, ami nem igazán okos dolog, de mindkettőre külön rekordot kell felvenni. Másrészt a valódi duplikátumokat a program nem találja meg, ha a címet nem ugyanúgy írták át ugyanakkor a cikknek a két rekordjában. A brit és amerikai helyesírási különbségeket jól kezeli az algoritmus, ahogyan a központosítás kis különbségeit is, de nem minden alapszó szerepel a valódi duplikátumok megállapításához. Az értékelő számára a duplikátumokat felderítő keresésre feltehetőleg az a leghatékonyabb módszer, ha néhány kiválasztott folyóirat alapján végzi a tesztelést.

Minden egyes duplikátum vizuális ellenőrzése mellett más módszer is van a duplikátumokra vonatkozó jelentések megbízhatóságának ellenőrzésére. Például abban az esetben, amikor az *ASLIB Proceedings* a tesztelt folyóirat, az INSPEC, a Library and Information Science Abstracts és a Social SciSearch adja az ISA legjobb kontroll csoportját. Ha igaz az a feltételezés, hogy a DIALOG szoftvere hibásan azonosítja a duplikátumokat bármelyik korábban említett okból, akkor ennek minden olyan adatbázisban meg kell mutatkoznia, amely ugyanazt a folyóiratot feldolgozza. A 4.11. ábra világosan megmutatja, hogy nem ez a helyzet. A kontroll csoportban csak a Social SciSearchben van duplikátum, ott is csupán egyetlen pár. Az ISA 224 rekordjából 156 duplikátum pár része.

The screenshot shows the DIALOG WEB interface with a search window and a duplicate check window. The search window displays results for the query 'JN=ASLIB PR? AND PY=1990:1992' across four sets (S1). The duplicate check window shows the results of a duplicate check performed on these sets, identifying duplicates within each set and across sets.

Set	Term Searched	Items	File
S1	JN=ASLIB PR? AND PY=1990:1992	78	2
S1	JN=ASLIB PR? AND PY=1990:1992	145	7
S1	JN=ASLIB PR? AND PY=1990:1992	121	61
S1	JN=ASLIB PR? AND PY=1990:1992	224	202

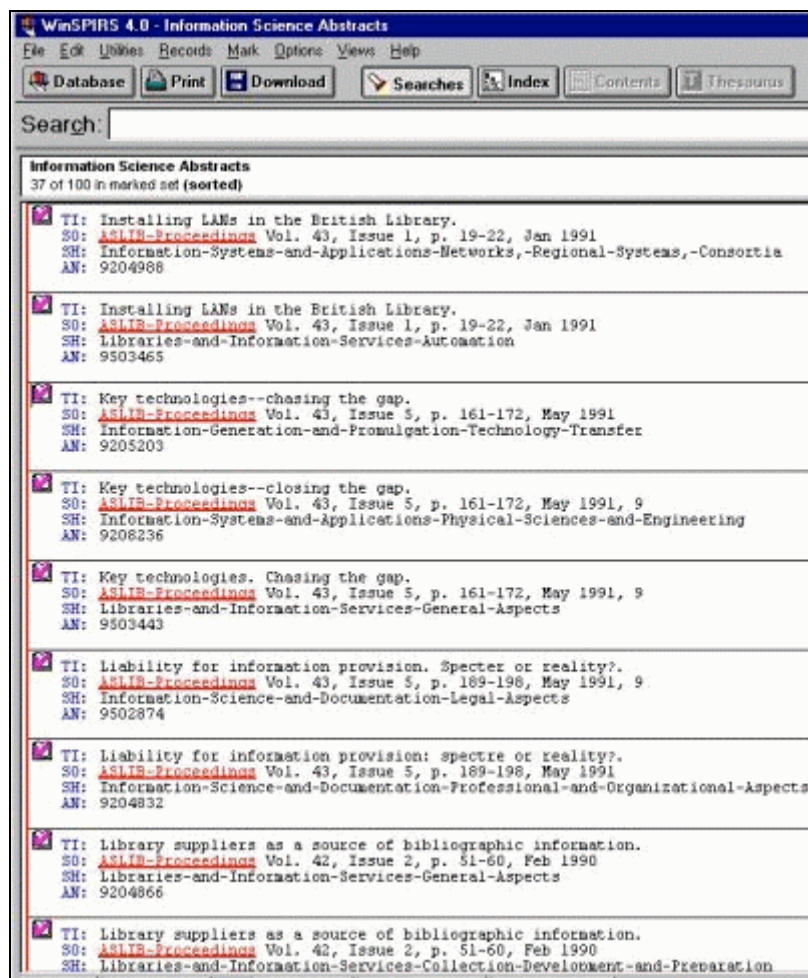
DIALOG Command: `s jn=aslib pr? and py=1990:1992` Submit 202

S1	JN=ASLIB PR? AND PY=1990:1992	568
S2	IDO S1 FROM 2 (duplicates only)	0
S3	IDO S1 FROM 7 (duplicates only)	2
S4	IDO S1 FROM 61 (duplicates only)	0
S5	IDO S1 FROM 202 (duplicates only)	156

4.11. ábra: Duplum-eredmények ellenőrzése kontroll-csoportokkal.

Az ISA WinSPIRS változatából vett példák annak tömörebb outputjával (4.12. ábra) egyértelművé teszik, hogy ezek valóban olyan rekordok, amelyeket kétszer vagy háromszor vittek be az ISA-ba, így például a Key Technologiesről szóló cikk esetében, ahol elírt alcímeket találhatunk. Az adatok bevitele évének elemzése rekordok ezreiben jelezte, hogy a duplikátumok nagy százalékát vitték be 1987 után, pedig ez volt az az év, amelyről a szerkesztő azt állította, hogy a duplikátumok hozzáadása megszűnt. A tipikus minta erre a folyóiratra és a legtöbb duplikátumra az, hogy a duplikátumok az első rekordok után három évvel kerültek be az adatbázisba. Ezek nélkül a duplikátumok nélkül az adatbázis nem mutatott volna következetesen és kerekén 800 rekordot a legtöbb aktualizáláskor 1995-ben.

Az új fájl producer elsőbbséget adott a duplikátumok és a háromszoros rekordok kiszűrésének, eltávolításának, de még mindig elég maradt 2000 végére, hogy az esetet az ISA DIALOG -os online változatából illusztrálni lehessen. (4.13. ábra)



4.12. ábra: Részlet az ISA duplikátumainak listájáról



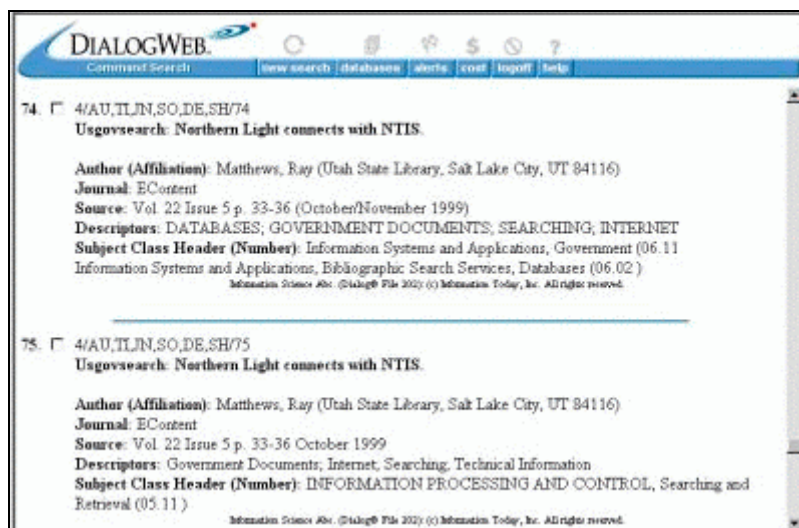
4.13. ábra: Még mindig maradtak duplikátumok az ISA-ban.

Más online és CD-ROM keresőprogramok nem teszik lehetővé a duplikátumok felderítését (kivéve az Ovidot, amely 1999-ben egészítette ki szoftverét ezzel a lehetőséggel), de azok, akik letöltenek rekordot egy bibliográfiai információt kezelő szoftverbe (például a Reference Managerbe vagy a ProCite-ba) használhatják ezeknek a programoknak a testre szabható duplikátum felderítő tulajdonságát. Az ISA webes változatának most megszűnt, használó-orientált Search-by-Search verziójában a duplikált és háromszoros rekordok megjelenítése és

kinyomtatása nagy gondot jelenthetett az alkalmi használóknak, akik tipikusan saját zsebből fizetnek. Nem csoda, hogy ez a szolgáltatás rövid idő alatt megszűnt.

Az adatbázisok CD-ROM-változatában a duplikátumok nem jelentenek zsebbevágó kiadásokat, de plusz munkát igényel a duplikátumok eltávolítása az eredmények végső listájáról. Természetesen jelzéssel is szolgálnak az adatbázis minőségéről. 15 évvel az adatbázis készítésének megkezdése után, 1998-ban a Documentation Abstracts, Inc. felbontotta szerződését az IFI/Plenummal, és az adatbázis előállítását az IFI/Plenumtól az Information Todayhoz került át.

Sajnos még mindig kerülnek be új duplikátumok a rendszerbe, ahogy azt a 4.14. ábra mutatja. Ez a helyzet megkérdőjelezi a duplikátumok felderítésének hatékonyságát. A duplikátumok terhet jelentenek a használók számára mind időben, mind pénzben kifejezve. Vegyük észre, hogy nem csak a deskriptorok, hanem még a széles tárgyi csoportok is különböznek a duplikátumok párjaiban. A nyomtatott változatban a rekordok listáját a széles tárgyi csoportok szerint rendezik, a duplikátumok nem jönnek elő ugyanabban a kategóriában, így a használóknak nincs déja vu érzése. Az online és a CD-ROM-változatok használóinak biztosan van.



4.14. ábra: Újabb hozzáadott duplikátumok.

Vannak további fontosabb szempontok, amelyek hatással lehetnek arra, hogy valaki hogyan értékeli egy adatbázis megfelelőségét, mint például a feldolgozott dokumentumok típusa és nyelve, a szereplő cikkek típusai, a magfolyóiratok felvételének mértéke, a rekordok időbeli eloszlásának aránya. Ezek a szempontok önmagukban is érdekesek, s a következő részekben tárgyaljuk őket.

Időbeliség

Az adatbázisok jelentős mértékben különböznek feltárt anyaguk retrospektivitásában. Önmagában semmi rossz nincs abban, ha egy adatbázis rövid időszak anyagát tárja fel, a nagyon hosszú időtartam sem feltétlenül erény, ez a használók preferenciáin és az adatbázis jellegén múlik. Számos online rendszerben alapbeállításban a kereshető időtartam a kurrens év és az azt megelőző két év. Ez azt sugallja, hogy a legtöbb használót a két vagy három évtől nem régebbi cikkek érdeklik. 1998-ban a UMI (most Bell & Howell) bevezette a Newsstand adatbázisát, az induló év 1998 volt. Bár ez szokatlanul rövid retrospektív feltártság a referáló és a teljes szövegű adatbázisok között, az adatbázis célja - ahogy arra neve is céloz - az, hogy elektronikus újságosstand legyen. Ez sokkal retrospektívabb, mint a valódi újságosstandok, ahol csak a folyóiratok, magazinok és napilapok kurrens száma található meg. A *Newsweek* adatbázis változatának megindulása 2000 februárjában bármilyen visszamenőleges anyag (back file) nélkül azonban mindenesetre túl extrém volt egy egyetlen forrást tartalmazó adatbázis számára.

A Ziff-Davis cég *Computer Select* CD-ROM adatbázisában hasonlóan rövid időtartam - egy év - anyaga volt, jó okkal. A *PC Magazine Plus*, egy másik CD-ROM termék hasonló elveket követ, s minden új kiadásában az utolsó 12 hónap számait tartalmazza, levágva a 13. hónap anyagát. Ezt a szokatlan feltártság-típus *gördülő feltártság* néven is ismert. A számítógépes szakirodalom avul el a legrövidebb idő alatt, különösen azok a cikkek, amelyek új számítógépes hardverről vagy szoftverről számolnak be vagy azt értékelik. A tegnapi hírek már történelemnek számítanak a számítógépes technikában.

Ez nem érvényes ugyanúgy a számítógép-tudományi kutatásokról beszámoló cikkekre, bár figyelemre méltó, hogy az Institute for Scientific Information (ISI) Journal Citation Reports-ában a természettudományi részhalmazon belül a számítógép-tudományi folyóiratoknak van a legrövidebb felezési ideje. A hivatkozási felezési idő azt jelenti, hogy a megjelenés idejétől számítva visszamenőleg hány év alatt jön össze a hivatkozások 50%-a egy adott folyóiratban. Ezek nagyon hasznos adatok annak megítéléséhez, hogy egy adatbázis retrospektivitása mennyire felel meg egy adott diszciplínának.

Bár az irodalomkeresésben a retrospektivitás igénye egyéni preferenciák szerint változik, s ugyanazon egyén esetében is különböző témákhoz kapcsolódóan, az avulás egy adott szakterületen jó mérőszám, amihez képest meg lehet ítélni az adatbázisban feldolgozott időtartamot (az adatbázis retrospektivitását).

A könyvtár- és információtudomány területén a hivatkozási felezési időt az ISI által figyelt 56 folyóirat közül 21-re számolják ki, csak azokra, amelyek 100 vagy több hivatkozást kaptak. A hivatkozási felezési idő több mint 10 év volt két folyóirat esetében, 1,2 és 8,8 év között volt a többinél. (Az ISI nem számolja ki a pontos hivatkozási felezési időt azokban az esetekben, amikor az több mint 10 év.) A felezési időnek ezek az adatai megfelelő perspektívába állítják a könyvtár- és információtudományi adatbázisok és a jelentős könyvtár- és információ-tudományi anyagot tartalmazó adatbázisok retrospektív feltártságát, amit a 4.16. ábra mutat. Ez azt sugallja, hogy a 15-20 éves időtartam több mint megfelelő lenne a legtöbb keresés esetén. Ha a keresőt a széles körben használt szakmai folyóiratok érdeklik a könyvtár- és információtudomány területén a tudományos folyóiratok helyett, még rövidebb időtartam is elég lehet. Ez megfelelő megvilágításba helyezi az adatbázisok retrospektív feltártságának fontosságát.

Journal Citation Reports - 1997 Social Sciences Edition

File Edit View Calculations Options Help

Journal Rankings

Mark	Rank	Journal Abbreviation	ISSN	1997 Total Cites	Impact Factor	Immediacy Index	1997 Articles	Cited Half-life
	1	ACAD PSYCHIATR	1042-9670	77	0.420	0.045	22	
	2	ACTA PSYCHIAT SCAND	0001-690X	6245	1.588	0.136	169	8.5
	3	AM J GERIAT PSYCHIAT	1064-7481	274	0.670	0.432	37	3.0
	4	AM J ORTHOPSYCHIAT	0002-9432	1862	1.718	0.689	61	> 10.0
	5	AM J PSYCHIAT	0002-953X	22478	6.501	0.943	283	7.3
	6	ANXIETY	1070-9797	113	1.512			2.7
	7	ANXIETY STRESS COPIN	1061-5806	81	0.583	0.000	18	
	8	ARCH GEN PSYCHIAT	0003-990X	23238	10.751	1.798	104	9.4
	9	ARCH PSYCHIAT NURS	0883-9417	207	0.439	0.000	38	5.4
	10	AUST NZ J PSYCHIAT	0004-8674	839	0.574	0.340	100	7.0
	11	B AM ACAD PSYCH LAW	0091-634X	408	0.469		0	7.8
	12	B MENNINGER CLIN	0025-9284	272	0.321	0.077	26	8.7
	13	BEHAV MED	0896-4289	219	0.622	0.105	19	6.6
	14	BRIT J MED PSYCHOL	0007-1129	981	0.667	0.167	30	> 10.0
	15	BRIT J PSYCHIAT	0007-1250	13851	3.265	0.744	195	8.3
	16	CAN J PSYCHIAT	0706-7437	1511	1.151	0.520	102	7.1

Sort: Journal Abbreviation Filter: PSYCHIATRY 80 Journals 0 marked

4.17. ábra: Pszichiátriai folyóiratok hivatkozási felezési ideje.

Az ember azt gondolhatná, az, hogy a PsycINFO adatbázisban feldolgozott anyag 1966-ig megy vissza, több mint elegendő a hivatkozási felezési idő fényében. Érdekes és szokatlan fejlemény az, hogy az Amerikai Pszichológiai Társaság (American Psychological Association) jelentős számú rekorddal, 300000 dokumentum anyagával egészítette ki az adatbázist, gondosan és szisztematikusan kiválasztott 1966 előtti publikációkkal.

A spektrum másik végén találjuk a humán tudományok diszciplínáit, ahol az elmúlt századokban megjelent könyveket, értekezéseket és cikkeket gyakran használják a mai kutatók. Sok társadalomtudós számára a negyedszázados anyag kurrensnek számít. Ebből a szempontból az ISI Arts and Humanities Search (20 év) és a Wilson cég Humanities Abstracts (16 év) retrospektivitásának különbsége jelentős lehet.

Az adatbázis szintjén vizsgált időtartamon túl figyelni kell egy szakterület magfolyóiratai és egyéb fontos folyóiratai feltártságának egyedi különbségeire is. Ezek közé tartozik a feltárás késői kezdése, korai abbahagyása, az egyenetlen feltártság és a hézagok a feltártságban. Ezekhez a szimptómákhoz kapcsolódik a feltártság sekélyisége és lassúsága, de ezeket a feltártság mélységével kapcsolatos részben tárgyaljuk, a frissesség és az aktualizálás mintáit pedig később ebben a fejezetben. Az időszak feltártsága problémáit az ISA adatbázis alapján illusztráljuk, amely a sok folyóirat különleges egyedi sajátosságainak a legszélesebb választékát kínálja.

Egy adatbázis induló dátuma általában a legkorábbi évet jelzi, amelyből a publikációk egy részét felveszi a szolgáltatás, és semmiképp sugallhatja azt, hogy akkortól minden forrást feldolgoznak. Ez nyilvánvaló olyan esetekben, amikor egy cím később kezdett megjelenni az adatbázis induló időszakánál. Például az *Internet Reference Services* című folyóirat lehetséges feltártságának legkorábbi dátuma 1997 - amikor a lap elindult - függetlenül attól, hogy mi az adatbázis indulási ideje. Sok adatbázis azonban nem kezd el egy folyóiratot azonnal feldolgozni. Amikor egy folyóiratot kiválasztanak, lehetséges, hogy csak attól az évtől kezdve tárják fel, vagy retrospektív feltárással az indulás évétől. Ez az utóbbi jobb megközelítés ésszerű határok között, feltételezve, hogy a szakterület igazi magfolyóiratai közül semmi sem hiányzik az adatbázisból.

A címek késői felvétele a legkönnyebben úgy állapítható meg, ha összehasonlítunk versengő adatbázisokat, ahogy azt a 4.18. ábra illusztrálja. Egy adatbázis által feltárt időszak jellegzetességeinek megállapításához a leghatékonyabb módszer az, hogy megjelenítjük a kiadási év indexét. A folyóiratok szintjén a feltárt időszak gyorsan megállapítható a DIALOG RANK parancsával. Ez kibontja egy, a felhasználó által meghatározott mezőnek az értékeit, mint például a kiadási évet ebben az esetben, s rangsorolja őket az előfordulás értékei szerint vagy betűrendben. A rangsorolható eredményhalmazok méretének van határa. Jelenleg ez 10000 rekord.

Miután kiválasztottuk a legfontosabb folyóiratokat és mindegyikükre vonatkozóan megalkottuk a halmazt, ezeket rangsoroltatni kell, megjelenítve a rekordok számát minden egyes évre kronológiai sorrendben (4.19. ábra). Ami szembetűnő az ISA adatbázisban a kiváló brit könyvtár-gépesítési folyóirat, a *Program* esetében, az nem a feltárt időszak, hanem a feltártság mértéke, vagy inkább a mélység hiánya. Nehéz eldönteni, vajon hézagnak tekintsük-e az évek kihagyását (mint 1972, 1975-1976, 1978-1979, s majdnem egy évtized 1988 és 1996 között), a feltárás korai abbahagyásának, időleges felfüggesztésének vagy csak egy szánalmas feltáratlanságnak. Az 1977-ből származó egyetlen tétel még életjelnek sem elég.

RQ 1960			
	Database starts	RQ records from	Dataness in years
LISA	1969	1973	4
Social SciSearch	1972	1975	3
ISA	1966	1970	4
LibLit	1984	1984	0

4.18. ábra: Magfolyóiratok későn kezdett feltárása

Az adatbázis indulási ideje, az RQ-ból származó rekordok, Késői indulás években kifejezve

Egyes adatfájl előállítók olyan folyóiratlistát szolgáltatnak, amely világosan jelzi, melyik évtől tárnak fel egy adott folyóiratot s melyik évben hagyták abba a feltárást. (A feltárás abbahagyása sok okkal igazolható. A természetes az, amikor egy folyóirat megszűnik vagy megváltoztatja címét; azaz a feltártság az alatt a cím alatt szűnik meg.) Az EBSCO, a Bell & Howell és az H. W. Wilson cégek a feltárás kezdő és befejező évét is jelzik, a folyóiratok listáját elérhetővé teszik a weben. Egyes folyóiratlisták mintái elérhetőek a <http://www2.hawaii.edu/~jacso/extra/savvy/journalbase/journalbase.html> címen.

Egyesek lehetővé teszik a felhasználó számára, hogy a folyóiratlista outputját személyre szabja. A Bell & Howell például felkínálja a választást, hogy a következő mezők közül melyek szerepeljenek: Cím, a folyóirat kódja, ISSN, a referálás kezdetének dátuma, a teljes szöveg közzélése kezdetének dátuma és a képi formában való közzelés dátuma. A LISA publikált egy folyóiratlistát a weben 2000-ben, de ez már akkor is elavult volt. Nem tartalmazza azokat a folyóiratokat, amelyeknek a címe megváltozott 1999-ben, mint például a *Database EContentre* vagy a *Library Software Review Library Computingra*. Csak a régi címek alatt sorolják fel őket (*Database* és *Library Software Review*).

A LISA folyóiratlistájának tartalma a lehető legegyszerűbb: csupán a folyóirat címe szerepel. Az ISA még kevésbé informatív a folyóiratairól. Az új szerkesztői gárda azonnal leállította a magfolyóiratok listájának közreadását. Kétségtelen, hogy kellemetlen ténynek bizonyult, hogy a korábbi fájlkészítők hanyagul kezelték az időszaki kiadványokat azzal, hogy kihagyták magfolyóiratok teljes számait, köteteit vagy több kötetét, amelyeket pedig teljes egészükben indexelniük kellett volna (Jacsó 1997e). Ehelyett egy szerkesztőségi cikk felsorolt új folyóiratokat, amelyeket a jövőben fel akartak tárni.

(<http://www2.hawaii.edu/~jacso/extra/savvy/journalbase/journalbase.html>)

RANKING SET BY PY (JOURNAL: PROGRAM)		

RANK: S1/1-230 Field: PY= File(s): 202		
(Rank fields found in 230 records -- 20 u		
RANK No.	Items	Term
-----	-----	----
15	6	1966
11	8	1967
9	13	1968
13	7	1969
10	9	1970
16	6	1971
8	15	1973
18	3	1974
20	1	1977
12	8	1980
5	18	1981
14	7	1982
4	19	1983
1	25	1984
19	3	1985
6	18	1986
17	5	1987
2	23	1997
3	20	1998
7	16	1999
---end of results---		

4.19. ábra: Egy folyóiratból vett tételek rangsorolása a kiadási év szerint, időrendben.

A felsorolt címek némelyike megszűnt jóval a bejelentés előtt: másokat nem tártak fel a következő évben, míg továbbiak csupán egy vagy két rekordot produkáltak 1998-ban, 1999-ben és 2000-ben, beleértve két olyat, amelyeket a szerkesztők a használóknak különösen a figyelmébe ajánlottak (4.20. ábra)

Az adatbázisok kiadói (szemben a fájlok előállítóival) általában nem teszik elérhetővé az egyes időszakok feltártságával kapcsolatos információkat, bár ezt könnyű lenne megtenni az adatbázis online vagy CD-ROM változatának súgóiban.

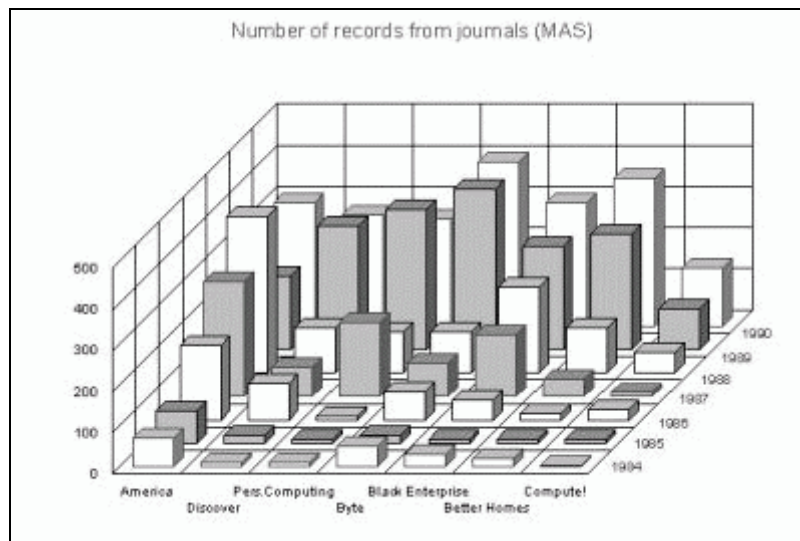
A fájlok előállítói által szolgáltatott folyóiratlistákról a feltárás kezdetét nem lehet készpénznek venni. Egyes fájl előállítók olyan folyóiratokat mondanak feltártnak, amelyeket csupán az adatbázis néhány kezdeti évében vettek fel. A Magazine Article Summaries (MAS), az EBSCO első indexelő és referáló adatbázisa azt állította, hogy a legtöbb címet 1984-től feltárja. Valójában csak néhány rekorddal hintették meg az adatbázist (még a nagyközönségnek

szóló, kétségtelen magfolyóiratok esetében is) 1984-től (és gyakran későbbtől) a folyóiratok számaiból (4. 21a ábra).

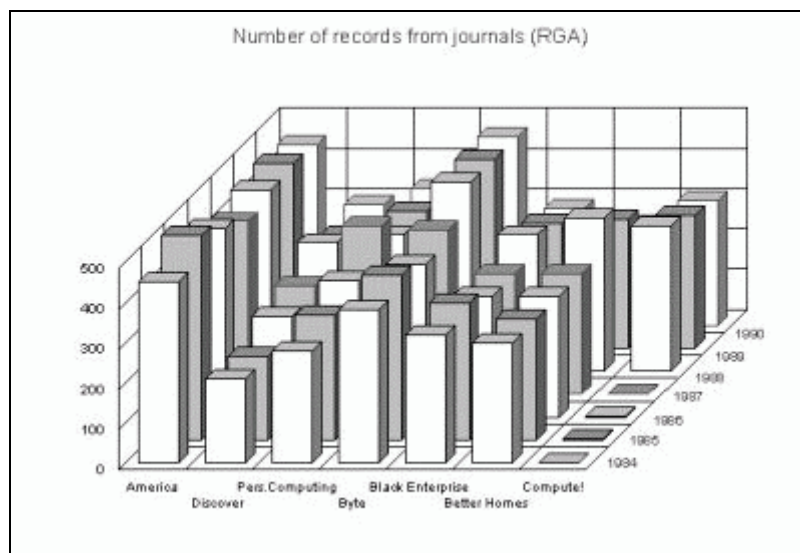
1998 items	1999 items	2000 items	Journals	1998 items	1999 items	2000 items	Journals
5	0	0	AI Magazine	5	0	0	Booklist
16	0	0	Bottom Line: Managing Library Finances	18	16	0	Bulletin of the Japan Special Libr. Association
1	0	0	Byte	2	0	0	Computer Life
10	4	0	Computerworld	1	0	0	DBMS
0	0	0	Digital Publishing Strategies	0	0	0	Fee for Service
1	1	1	Forbes	2	0	0	Fortune
1	0	0	Home Office Computing	4	2	0	Imaging
134	114	0	Information Today	2	1	0	InfoWorld
0	0	0	Inside the Internet	3	2	0	International Journal of Information Sciences and Decision Making
44	15	0	Internet World (was Webweek)	8	3	0	Intranet Professional
2	0	0	LAN Times	18	24	0	Link Up
5	6	0	Multimedia Schools	3	1	0	Network
2	0	0	New Media (NewMedia)	29	0	0	PC Magazine
9	6	0	PC Week	2	2	0	PC World
0	0	0	Proceedings of the Geoscience Information Society	27	0	0	Proceedings of the Off-Campus Library Service Conference
6	0	0	School Libraries Worldwide	6	0	0	Software Magazine
1	0	0	WebWeek (now Internet World)	2	0	0	Windows Magazine

4.20. ábra: Új folyóiratok, amelyek feltárását ígérte az ISA és tényleges feltártságuk

Egy gyors pillantás alapján is nyilvánvaló, hogy éles ellentétben van ezzel az H. W. Wilson cég adatbázisaiban a folyóiratok jellegzetesen alapos és kiegyensúlyozott feltártsága, ahogy azt a 4.21.b ábra mutatja. (Egy hirtelen hullámvázst a feltártságban (a feltárt tételek számában) megindokolhatja az, ha egy folyóirat megjelenési gyakorisága növekszik vagy más okból nő az egy adott évben publikált cikkek száma. Ez volt a helyzet, amikor a *CD-ROM Professional* havi megjelenésű magazinná vált, míg azelőtt csak évi hat száma jelent meg.



4. 21a. ábra: Az adatbázis meghintése néhány tétellel fontos forrásokból a MAS adatbázisban.



4. 21b. ábra: Fontos források jól kiegyensúlyozott feltártsága a Readers' Guide Abstracts adatbázisban.

Mélység

Az adatbázisban szereplő folyóiratok feltártsága mélységének vizsgálata és megértése alapvető akkor, amikor az adatbázisnak a minőségét értékeljük. Pusztán az, hogy egyes folyóiratok jelen vannak az adatbázisban, nem garantálja, hogy megfelelő mértékben fel is tárják azokat. Az 1960-as években és az 1970-es évek elején a folyóiratok feltártságára vonatkozó vizsgálatok gyakran korlátozódtak annak meghatározására, hogy folyóiratok csoportjai szerepelnek-e egy vagy több nyomtatott indexelő és referáló forrásban, s milyen átfedés van köztük a folyóiratok szintjén. Ezek a vizsgálatok általában az indexelő és referáló források kumulált indexein alapultak, de még így is fárasztó, unalmas munkát jelentettek. Gilchrist (1966), Goldstein (1973) és Edwards (1976) alapos vizsgálatai a folyóiratcímek szintjén való összehasonlításra korlátozódtak, s nem tárgyalták a folyóiratok feltártságának mélységét. Így is előkészítették az utat a későbbi vizsgálatokhoz, amelyek összehasonlították a kiválasztott folyóiratok feltártságának mélységét az adatbázisokban.

Fontos tudni, hogy nem csak az alkalmi használók nem veszik észre, hogy fontos folyóiratok milyen sok cikke hiányzik egy adatbázisból, hanem kutatók sem, akik bibliometriai és tudományometriai kutatásokat végeznek. Ezek a kutatások a kiadás mintáit és a hivatkozási szokásokat vizsgálják, hogy rangsorolják a szerzőket, intézményeket, folyóiratokat, sőt országokat is az adatbázisokban végzett keresések alapján. Kutatási frontokat és trendeket vizsgálnak olyan diszciplínákban, ahol a folyóirat-publikációk a domináns források a termékenység jelzéséhez. A publikációs produktivitás az alapvető szempont előléptetések, támogatások és meghatározott állások esetében. Súlyosan torzíthatják a bibliometriai és tudományometriai vizsgálatok eredményeit a hézagok és a feltártság jelentős hullámsága, valamint a folyóiratok anyagának sekély, következtelen, lassú feltárása vagy a feltárás idő előtt való abbahagyása, s félrevezethetik azokat a használókat, akik azt gondolják, hogy amit találtak egy adatbázisban, az minden, amit a témáról az adatbázisban feltártnak mondott nagy presztízsű folyóiratokban írtak.

A hozzáértő használók azonban tudják, hogy két vagy három adatbázisban kell keresniük ahhoz, hogy megfelelően részletes (alapos) keresési eredményekhez jussanak. Azonban még tapasztalt használók is áldozatául eshetnek az adatbázisok feltártságáról szóló valótlan állításoknak, a klisészerűen megírt reklámanyagokban szereplő, mindenre kiterjedő feltárással szóló ígéreteknek. Ahogy azt a 3. fejezetben láttuk a tárgyi feltártsággal kapcsolatban, a legabszurdabb PR állítások könnyen ellenőrizhetők a címekben szereplő szavak alapján végzett kereséssel több adatbázisban. A folyóiratok feltártságának mélységét, kimerítő voltát nehezebb igazolni.

A folyóiratcímek rövidítésének, központosításának és helyesírási változatainak nagy sokasága egyes adatbázisokon belül és különböző adatbázisok között defenzív keresési stratégiákat követel meg, amely megjósolja, felkutatja a lehetséges variációkat, s igazodik hozzájuk. Amikor a folyóiratcímek mezőjét csak mondatonként és nem szavanként indexelik (mint a legtöbb DIALOG adatbázisban), a keresőkérdés megalkotása elég nehéz feladat. Ha a folyóiratcím mezőt szavanként és kifejezésként is indexelik, a keresőkérdés sokkal rugalmasabban megalkotható, mint például a legtöbb adatbázis Ovid és OCLC változataiban.

A nyomtatott útmutatók és a folyóiratok listái minden egyes folyóiraatra vonatkozóan tartalmazhatják az első és az utolsó feldolgozott évet. Ezek hasznosak lehetnek az ezeket a listákat vizsgáló használók orientálására, de azt nem garantálják, hogy a jelzett években valóban alaposan feltárták ezeket a lapokat. Ahogy az előző fejezetben láthattuk, az újonnan feldolgozott folyóiratok büszke bejelentése sem garancia. Az is elég gyakori, hogy a jelzett időszakból az első évben a folyóirat feltártsága elég sekély. Ez igaz lehet az egész adatbázisra, ahogy azt az a példa bizonyítja, amelyben hét folyóirat első néhány évének a Readers' Guide Abstractsben és a Magazine Article Summariesben való feltártságát hasonlítottuk össze. A feltártság ábrázolása 4. 21a és a 4. 21b ábrákon magáért beszél.

A folyóiratoknak magfolyóiratként való besorolását nem lehet készpénznek venni minden adatbázisban. Az ilyen meghatározás a legkimerítőbben (általában borítótól borítóig) feltárt folyóiratokra érvényes. Meghatározott dokumentumtípusok kihagyása egy magfolyóirat egy számából (mint a nekrológok és a hírek) elfogadható lehet (különösen, ha ezt egyértelművé teszik a dokumentációban vagy a sűgőben, s ha ezt a szabályt következetesen alkalmazzák). A nagy cikkek kihagyása azonban nem fogadható el. Állítólagos magfolyóiratok teljes számainak vagy kötetinek kihagyása egyenlő azzal, amit más szakmákban műhibának nevezünk. A könyvtár- és információtudományi anyagot feltáró adatbázisok között csak az ISA és az INSPEC az, amely egyes folyóiratokat magfolyóiratként határoz meg, s az ISA abbahagyta ezt a gyakorlatot, ahogy azt a korábbiakban már megtárgyaltuk. Meghatározásuk szerint az ISI adatbázisokban minden folyóirat magfolyóirat külön jelzés nélkül, mivel az ISI adatbázisokban minden folyóiratot borítótól borítóig feldolgozottak mondanak.

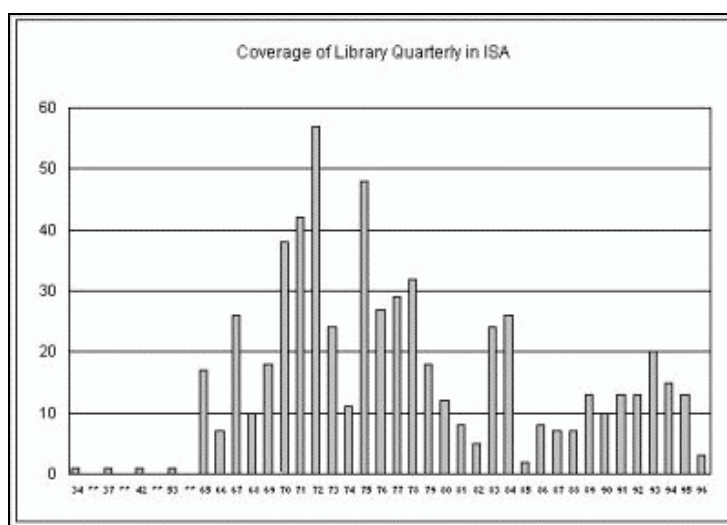
A magfolyóiratokhoz való hozzájutás nehézségeire utaló kifogások (Allcock 1997) nem tűnnek jogosnak, figyelembe véve, hogy az ISA összes magfolyóirata az USA-ból, Nagy-Britanniából és Kanadából származik. Ugyanez vonatkozik a folyóiratok magas árára is. Ez a dolog velejárója, s nem igazolja teljes kötetek kihagyását (Jacsó 1997e). Az ilyen állítások nem kelhetnek szimpátiát olyan használókban, akik magas árakat fizetnek az adatbázisokért és akik tudják, hogy a legtöbb állítólagos magfolyóirat éves előfizetési ára egy könyvtár- és információtudományi adatbázisban 100 dollár alatt van.

Egyes folyóiratok egyetlen adatbázisban való feltártságán túltekintve jobb perspektívát nyerhetünk, ha van valamilyen szint, amihez a feltártság mélységét viszonyíthatjuk.

Az adatbázisban feltárt anyag időbelisége nem választható el a feltárás mélységétől. A mélység nem könnyen meghatározható, mert mozgó célpont. A feltárás mélysége - azaz a rekordok mennyisége egy folyóiratról az adatbázisban való feltártsága idején - évről évre változik,

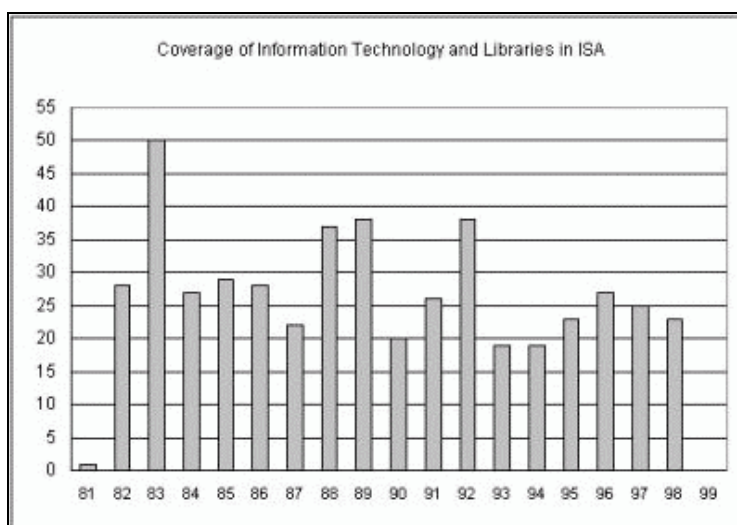
függetlenül a folyóiratban évente megjelenő cikkek mennyiségének lehetséges változásaitól. Ezt mutatta a korábbi, 4.19. ábra, a *Program* című, nagyhírű könyvtár-automatizálással foglalkozó brit folyóirat százalmas feltártsági mintája az ISA adatbázisban.

Rendkívül egyenetlen, hullámvasútszerű feltárás nagyon lehangoló olyan folyóiratok esetében, amelyek a szakterületük élvonalbeli kiadványai, és elfogadhatatlan valódi magfolyóiratok esetében. A *Library Quarterly* feltétlenül megérdemelné, hogy egy kötetének minden cikkét felvegyék egy adatbázisba, de az ISA esetében szemmel láthatóan nem ez a helyzet (4.22. ábra). Ugyanez érvényes egy másik nagyra tartott folyóiratra, a *Library Trends*-re. Anélkül, hogy megnéznénk feltártságát más adatbázisokban, teljesen nyilvánvaló, hogy a kiszámíthatatlan feltártságnak semmi köze ahhoz, hogy épp mennyi cikk jelent meg ebben a két folyóiratban.



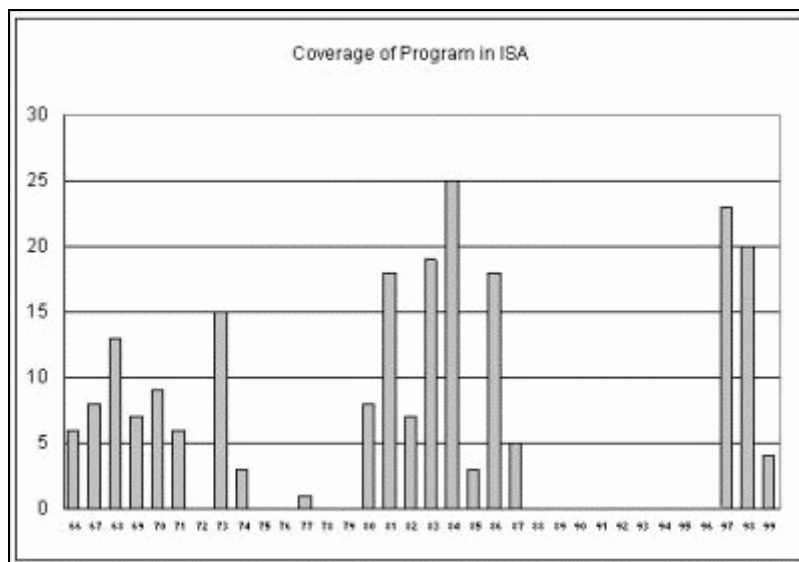
4.22. ábra: A *Library Quarterly* hullámvasútszerű feltártsága az ISA-ban.

Amikor hasonló minta érvényesül az ISA-ban olyan magfolyóiratok esetében, mint a *Government Information Quarterly*, *Journal of Documentation*, *RQ* (most *Reference & User Services Quarterly*) vagy az *Information Technology and Libraries* (4.23. ábra), akkor az ember nehezen találja elhíhetőknek, hogy „a magfolyóiratokat a maguk teljességében feldolgozzuk”, ahogy azt az ISA Használói útmutatója állítja.

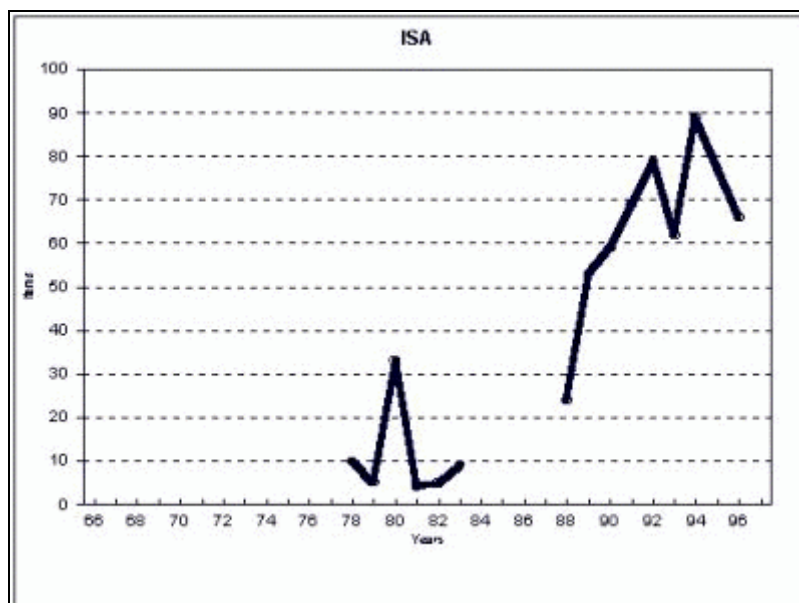


4.23. ábra: Az *Information Technology and Libraries* hullámvasútszerű feltártsága az ISA-ban.

A hullámvasút lefelé menő részének különleges esete az a hézag, ami akkor történik, amikor egyetlen rekord sincs egy egész kötetből vagy több kötetből. Bár az ember együtt érezhet az adatbázis szerkesztőjének azzal a problémájával, amit az ISA korábbi adatbázis szerkesztője osztott meg velünk (Allcock, 1997), hogy alkalmasszerűen hiányzik egy vagy két lapszám, az ISA adatbázis esetében ez visszatérő probléma, amire nincs példa egyetlen más adatbázisnál sem. Ez nagy probléma nagypresztízsű folyóiratok esetében még akkor is, ha nem tekintik őket magfolyóiratnak az adatfájl készítője, ahogy ez a *Program* című folyóirat esetében (4.24. ábra) előfordult, s még nagyobb probléma az adatbázis magfolyóiratai esetében, mint a *Scientometrics* (4.25. ábra).



4.24. ábra: Hézagok a nagy presztízsű *Program* feltártságában az ISA-ban.

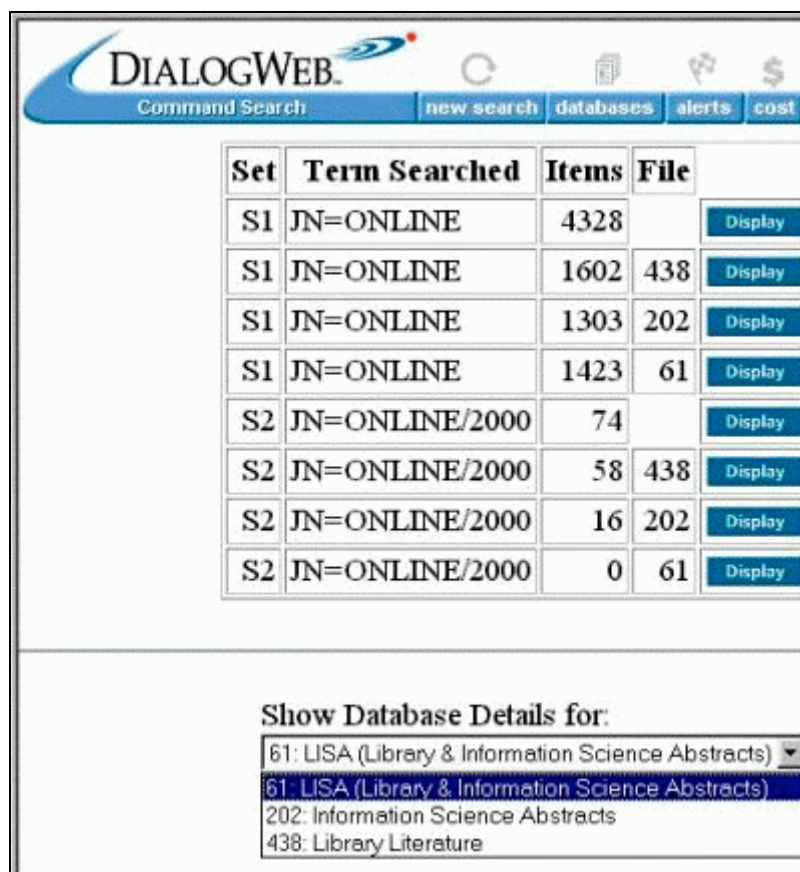


4.25. ábra: Hézagok a *Scientometrics* című magfolyóirat feltártságában az ISA-ban.

A hézag (a feldolgozás időleges felfüggesztése) a feldolgozás korai megszüntetésévé változhat egy fontos folyóirat esetében. Néha nehéz megállapítani, hogy egy folyóirat feltárása milyen állapotban van. A feltárás néhány éves teljes felfüggesztése után, néhány, a klinikai

halál állapotában levő cím (a fájl készítőjének perspektívájából) némi életjelt mutat. Ez volt a helyzet az *RQ*, az egyik legmagasabban rangsorolt könyvtári folyóirat esetében, amely az ISA nyomtatott kiadásában a magfolyóiratok listáján maradt 1994-ben és 1995-ben, annak ellenére, hogy ezekből az évekből egyetlen rekord sem került be az adatbázisba az *RQ*-ból. 1996-ban újra felbukkan az ISA-ban, de csupán négy rekorddal. Ez elég sajátos kezelése egy magfolyóiratnak. Egy csavar a kiváló folyóirat feltártságában az volt, hogy 1997 közepén a korábbi cím, *RQ* alatt valóban megszüntették kiadását. 1998 végéig új címe, *Reference & User Services Quarterly* alatt sem választották ki feldolgozásra az ISA-ban.

Nehéz megérteni, hogy egy fájl előállítója miért hagyja abba egy olyan folyóirat feldolgozását, amelyet kulcsfontosságúnak tekintenek szakterületén. Ez történt a LISA adatbázissal, amikor leállította az *Online* magazin feldolgozását. 2000 végén még mindig nem voltak rekordok a folyóirat 1999-es és 2000-es számaiból (4.26. ábra). Nem lehet eldönteni, hogy ez feltárás szüneteltetése vagy megszüntetése. Bármelyikről van szó, baklövés. Ugyanez történhet az *EContent*tel is a LISA-ban. Míg az 1999-es számokból 36 rekord került be az adatbázisba, a 2000-esekből egyetlen sem. Miután ezeket az eredményeket bemutatta egy konferencián, a szerzőnek azt a választ kapta, hogy a LISA erőfeszítéseket tesz a hiányok pótlására, s folytatják ennek a két fontos folyóiratnak a feldolgozását. Valóban, mikorra ez a könyv nyomdába ment, a LISA 2001. június 28-i aktualizálása során kiegészítették ezekkel a hiányzó rekordokkal. Jobb későn, mint soha.



The screenshot shows the DIALOGWEB search results page. At the top, there is a navigation bar with 'Command Search', 'new search', 'databases', 'alerts', and 'cost'. Below this is a table with the following data:

Set	Term Searched	Items	File	
S1	JN=ONLINE	4328		Display
S1	JN=ONLINE	1602	438	Display
S1	JN=ONLINE	1303	202	Display
S1	JN=ONLINE	1423	61	Display
S2	JN=ONLINE/2000	74		Display
S2	JN=ONLINE/2000	58	438	Display
S2	JN=ONLINE/2000	16	202	Display
S2	JN=ONLINE/2000	0	61	Display

Below the table, there is a section titled 'Show Database Details for:' with a dropdown menu showing the following options:

- 61: LISA (Library & Information Science Abstracts)
- 61: LISA (Library & Information Science Abstracts)
- 202: Information Science Abstracts
- 438: Library Literature

4.26. ábra: Az *Online* című folyóirat feldolgozásának felfüggesztése vagy megszüntetése a LISA adatbázisban.

Az ISA adatbázis új előállítója megpróbálja befoltozni azokat hézagokat, amelyekkel először egy publikálatlan dokumentumban szembesítette az ISA igazgatótanácsát ennek a könyvnek a szerzője, később egy cikkben publikálta erre vonatkozó eredményeit a *Library and Infor-*

mation Science Research című folyóiratban (Jacsó 1998a). Ezekre a visszamenőleges teljessé tevő próbálkozásokra nagyon nagy szükség volt, de hatással volt fontos folyóiratok új számainak feldolgozására is, ahogy azt ennek a fejezetnek a frissességre vonatkozó részében megtárgyaltuk. A hiányok pótlását 1990-től kezdve tervezték (ésszerű cél), de számos olyan cím esetén nem érték el, amelyek alapvető fontosságúak az adatbázis szakterületén.

A címjegyzék adatbázisokban való retrospektív feltárás összetettebb kérdés. Bizonyos címjegyzékekben, mint például cégek címjegyzékeiben és ki kicsoda típusú adatbázisokban a retrospektív feltárás nem olyan fontos, mert a használókat leginkább a friss cégszámok érdeklik, mint például a jelenlegi név, cím, e-mail cím és faxszám, mivel általában egy rekordot készítenek minden kiadás számára és a korábbi kiadások rekordjait is megtartják. Ugyanez érvényes filmes és zenei címjegyzékekre is. Ilyen esetekben fontos tudni, időben mennyire megy vissza az adatbázis. A Kongresszusi Könyvtár REMARC adatbázisa az egyik olyan könyvkatalógus, amely a leginkább retrospektív, és különösen fontos olyan kutatók számára, akik fel akarnak kutatni egy, a 15. században megjelent könyvet. Ezt nem lehet elvégezni a *Books in Print*, *British Books in Print*, Amazon.com, Barnes & Noble vagy akár a *Books Out of Print* adatbázisban.

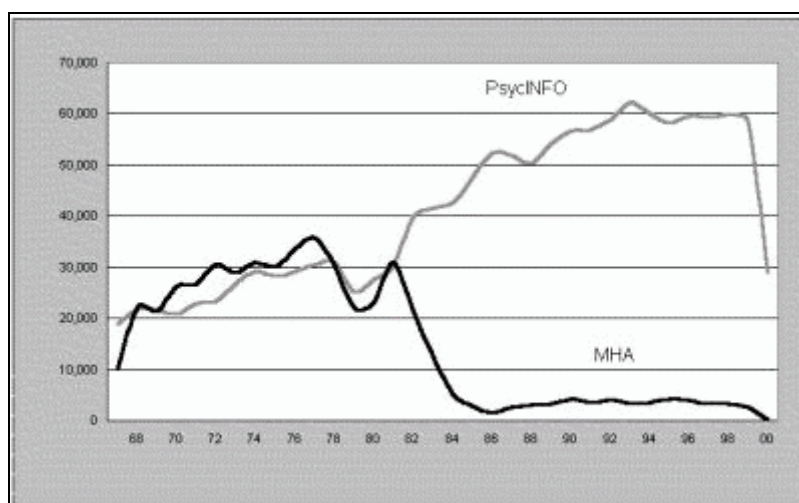
A bibliográfiai és a teljes szövegű adatbázisokhoz hasonlóan a címjegyzékek esetében sem szabad elfogadni azt, amit az adatbázis kezdő évének mondanak. Gyakran csak néhány rekord van ebből az időszakból. Másrészt azonban mind a bibliográfiai, mind a címjegyzék adatbázisok gyakran sokkal jobban visszamennek időben, jelentős számú rekorddal az előtről, mint ami az adatbázis címkéjén szerepel vagy amit a reklámanyagok állítanak. Az ISA például elég sok 1966 előtti cikket tartalmaz, pedig azt határozták meg indulási évként. Az Ovid honlapján 1979-et jelöli meg az AGRICOLA indulási évként, de a kiadási év indexe azt mutatja, hogy jelentős anyag található benne az 1970-es évek közepéről (4.27. ábra).

Year of Publication Index Display	
Perform Search Back in Index Main Search Page Forward in Index	
Choose from among the following index entries:	
Select	Year of Publication # of Citations
<input checked="" type="checkbox"/>	1975.yr. 7042
<input checked="" type="checkbox"/>	1976.yr. 8367
<input checked="" type="checkbox"/>	1977.yr. 17963
<input checked="" type="checkbox"/>	1978.yr. 86354
<input type="checkbox"/>	1979.yr. 176817
<input type="checkbox"/>	1977.yr. 67
<input type="checkbox"/>	197u.yr. 607
<input type="checkbox"/>	1980.yr. 173997
<input type="checkbox"/>	1981.yr. 159271

4.27. ábra: Az AGRICOLA korábban kezdte a jelentős mértékű feltárást, mint ahogy azt állították.

Ha ábrázoljuk az egyes időszakokból feltárt anyagot, rendkívül lényeges különbségeket mutathatunk ki az adatbázisok között. Míg a PsycINFO jelentős növekedést mutat a feltárt anyagban az évek során, ami tökéletes összhangban van a szakterület publikációinak mennyiségi növekedésével, addig a Mental Health Abstracts példátlan csökkenést mutat azóta, hogy 1983-ban az IFI/Plenum átvette (4.28. ábra). Az éves növekedés nagyságrendekkel csökkent. A két adatbázis, amely fej-fej mellett haladt az 1980-as évek elején, azóta ellentétes irányba halad. Zuhanórepülése, a feltárt források, a folyóiratbázis és a feltárás mélységének drámai csökkenése miatt, ahogy azt az 5. fejezetben megtárgyaljuk, az MHA-t nagyon kis mértékben tudják hasznosítani a pszichiátriát és pszichológiát hallgató egyetemisták és a gyakorló szakemberek.

Az MHA hasonló problémákat mutat a feltárt folyóiratok szempontjából, mint amit az előzőkben az ISA-ról elmondtunk. Bár sohasem különböztette meg magfolyóiratokat a rendszeresen és alkalomszerűen feltárt folyóiratoktól, sok alapvető folyóirat feltárásának megszüntetése olyan területekről, amelyeken az MHA különösen hasznosnak tekinthető, mint a pszichofarmakológia és a pszichiátriai kezelés, aláásta ezt az adatbázist (bár néhány előfizető egyetem nem vette észre és továbbra is előfizet az MHA adatbázisra).



4.28. ábra: A PsycINFO és az MHA adatbázisban feldolgozott anyag mennyisége évenkénti bontásban

Az adatbázisok által feldolgozott időszak meghatározása könnyű. A legtöbb rendszerben a kiadási év indexe megjeleníthető és letölthető egy fájlba, amelyből egy táblázat generálható. Ha ez az opció nem elérhető, a kiadási évre vonatkozó kereséssorozatot lehet elvégezni, az eredményként megkapott keresési eredmények elmenthetők egy fájlba, s onnan egy táblázatba importálhatók. Bár az adatgyűjtés gyors és könnyű, mindig gondos megközelítést igényel. Ahogyan azt a 9., a rekord teljességéről szóló fejezetben megtárgyaljuk, sok adatbázisban építésük első néhány évében nem vették fel a rekordokba a kiadási évet. Másokban a kiadási év mező néhány egyéb adatot is tartalmaz. Ha ez jelentős mértékben történik így, akkor az torzíthatja az egy adatbázisban feltárt időszakról kapott képet.

A hibás adatokat nehéz észrevenni, kivéve azt, ha azok első pillantásra nyilvánvalóan hibásak. Mielőtt elhinnék, hogy az AGRICOLA-ban rekordok ezrei találhatók olyan dokumentumokról, amelyek Gutenberg születése előtt jelentek meg (4.29. ábra), gondoljanak arra, hogy ezek sajtóhibák (például 1078 1978 helyett) vagy más hibás számok, amelyeket begépeltek a kiadási év mezőjébe. Mielőtt valami összeesküvést sejtjenek a 2000-en túli kiadási évek esetében, jobb ha tudjuk, hogy valószínűleg ezek is sajtóhibák. Az ilyen kiadási év nem volt teljesen lehetetlen 1999-ben egyes adatbázisokban, de csak olyanokban, mint az EVENTSLINE,

amelyben 2005-ig vannak rekordok tervezett eseményekre, vagy a *Books in Print*ben, amelyben van egy részhalmaz, amely a jövőben megjelenő címekre vonatkozó rekordokat tartalmazza.

Set	Term Searched	Items
S1	PY<1400	1887
S2	PY="1078"	18

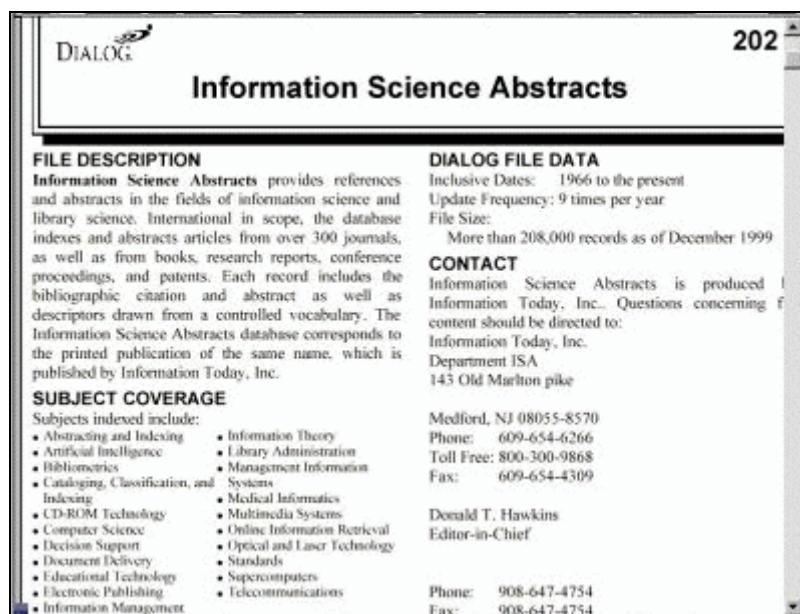
4.29. ábra: Lehetetlen kiadási évek az AGRICOLA adatbázisban.

Hasonló módon abszurdak a PY=200 értékek az Information Science Abstracts adatbázisban, ezeket az első 2000-es aktualizálás idején bekerült rekordok kapták. Azon túl, hogy ez elég rossz kezdés volt az új évezredben (ha önök azt fogadják el, hogy az 2000-ben kezdődött), ez megfosztotta tőlük azokat a használókat, akik keresésüket a kurrens évre szűkítették, hogy potenciálisan releváns és friss rekordokhoz jussanak. Bár ez a hibás kiadási év csak 71 rekordban jelenik meg, ezek 100%-át jelentik azoknak a tételeknek, amelyeket a 2000-es év első aktualizálásakor folyóiratok 2000-es számaiból adtak hozzá az adatbázishoz. Ironikus látni a tévedést egy olyan cikk rekordjában, amely a Y2K bug túlélését említi - kétségkívül egy kicsit túl korán (4.30. ábra). Az év folyamán később az ISA korrigálta ezeknek a rekordoknak a kiadási évét.

<p>Record 1 of 2 in ISA 1966-2000/03</p> <p>TI: On the Net in 2000.</p> <p>AU: Notess, -Greg-R</p> <p>SO: Online Vol. 24, Issue 1, p. 71-73, January 200.</p> <p>PY: 200</p> <p>AB: As the world plunges forward into 2000, with the dreaded Y2K transition behind it, the Internet marches onward, growing, mutating, and affecting all sorts of interactions between businesses, consumers, family, and friends. The pace of Internet change was dizzying in 1999, and there is great excitement for the future of intranets, extranets, and cost-cutting through Web-based consumer service. Forecasts some changes that may be in store for 2000 and beyond by overviewing some of the major trends of 1999 with respect to electronic commerce, vertical portals, shopping search engines, domain name registration, and mergers and acquisitions in the information industry.</p> <p>AN: 9904021</p>
--

4.30. ábra: A Y2K bug az ISA egyik rekordjában

Amikor egy fájl előállítását egy új cég veszi át, mindig érdemes újra meglátogatni és -értékelni egy adatbázis minőségét, benne a feltártság mélységét. Ez történt 1998-ban, amikor az Information Today megszerezte az ISA adatbázist és alkalmazta a korábbi technikai tanácsadót. Bár az új producer megszüntette a magfolyóiratok listáját, az adatbázis szakterületi feltártságával kapcsolatos, revideált megállapítások (4.31. ábra) tartalmaznak néhány kiindulópontot ahhoz, hogy ellenőrizni lehessen a feltártság mélységéhez kapcsolódó, nagyon szükséges változásokat azokban a témákban, amelyeket az adatbázis elsődleges területeinek tekintenek. A 30 hónap, ami eltelt az ISA átvételétől ennek a könyvnek a befejezéséig, elég időt adott ahhoz, hogy lássuk, vajon a pozitív változások megvalósultak-e.



4.31. ábra: Az ISA adatbázis tárgyi feltártságára vonatkozó, revideált állítás.

Sajnos a feltártság mélységének problémái az adatbázis (állításuk szerint) elsődleges területein (ahogyan a frissesség hiányának problémái is) az ISA adatbázis új vezetősége idején is megmaradtak. A kiadási évre vonatkozó gyors keresés is arról árulkodik, hogy az adatbázisba kerülő rekordok száma drasztikusan csökkent az elmúlt három évben az új vezetőség alatt (4.32. ábra).

A tényleges számok ellentmondanak az 1998 októberi szerkesztőségi cikk ígéretes megállapításainak, amely szerint, hogy „az ISA következő néhány száma több cikket fog tartalmazni, mint általában (szokott)”. Éppen az ellenkezője történt. Ez a trend nagyon valószínűtlenné teszi, hogy a feltártság mélysége az adatbázis elsődlegesnek minősített szakterületein fejlődni fog.

A feldolgozott folyóiratok száma nem változott jelentős mértékben 2000-ig (4.34. ábra), ami azt sejteti, hogy ha sokkal kevesebb rekord került be az adatbázisba a legújabb 30 hónapban, ez csak úgy lehetséges, ha a folyóiratok (és a témák) feldolgozottsága még sekélyebb volt. 1999 óta konferencia-kiadványok címeit a folyóiratcím mezőben szerepeltetik, és a konferencia előadások dokumentumtípusként az „article” (cikk) megjelölést kapták (egy újabb dokumentumtípus, azután, hogy a konferencia előadások besorolása a „monographic” és a „monographic chapter” dokumentumtípusok között váltakozott). Ez növeli mind a feldolgozott „folyóiratok”, mind a feldolgozott „cikkek” számát ebben a két évben, így a kép valamivel rosszabb, mint amilyennek tűnik.

DIALOGWEB		
Command Search		
		new search databases
E1	8352	PY=1985
E2	9419	PY=1986
E3	9404	*PY=1987
E4	9623	PY=1988
E5	8625	PY=1989
E6	8421	PY=1990
E7	7529	PY=1991
E8	8348	PY=1992
E9	7856	PY=1993
E10	8064	PY=1994
E11	8921	PY=1995
E12	9327	PY=1996
E13	6449	PY=1997
E14	4655	PY=1998
E15	3618	PY=1999
E16	71	PY=2000
E17	1102	PY=2000

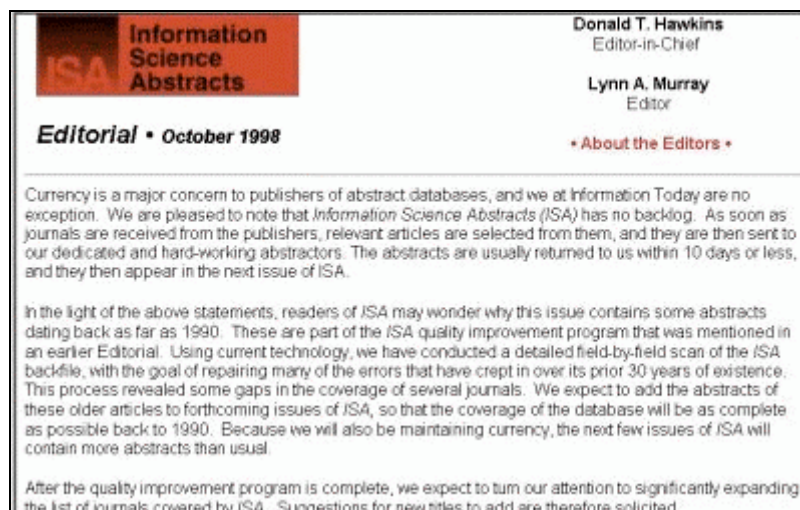
4.32. ábra.: Az ISA adatbázishoz adott rekordok számának gyors csökkenése a legutóbbi években.

Szerkesztőségi cikk. 1988 október

A frissesség fontos kérdés egy referáló adatbázis kiadója számára, mi sem vagyunk kivételek ez alól itt az Information Today-nél. Örömmel közöljük, hogy az *Information Science Abstracts*nek (ISA) nincs lemaradása. Mihelyt a folyóiratok megérkeznek a kiadótól, a releváns cikkeket kiválasztjuk belőlük, s azután elküldjük őket elkötelezett és szorgalmas referálóinknak. A referátumok általában 10 nap múlva (vagy még hamarabb) visszaérkeznek hozzánk, s azután megjelennek az ISA következő számában.

A fenti állítások fényében az ISA olvasói csodálkozhatnak, miért tartalmaz ez a szám néhány olyan referátumot, amelyek egészen 1990-ig mennek vissza. Ez része az ISA minőségfejlesztő programjának, amit egy korábbi szerkesztőségi cikkben már megemlítettünk. A legújabb technika felhasználásával részletesen, mezőről mezőre áttekintettük az ISA régi anyagot tartalmazó fájlját abból a célból, hogy kijavítunk sok olyan hibát, amely a 30 éves létezés során becsúszott. Ez a folyamat több hiányt tárt fel sok folyóirat feltártságában. Várakozásaink szerint ezekhez a cikkekhez referátumokat kapcsolunk az ISA későbbi számaiban, így az adatbázis feldolgozottsága a lehető legteljesebb lesz 1990-ig visszamenően. Mivel a frissességet is fenn fogjuk tartani, az ISA következő néhány száma több rekordot fog tartalmazni, mint általában.

Miután a minőségfejlesztési programot befejeztük, figyelmünket várhatólag az ISA-ban feldolgozott folyóiratok számának jelentős mértékű növelése felé fogjuk fordítani. Ezért nyomatékosan kérjük javaslataikat az új címekre.



4.33. ábra: Néhány ígéretes állítás egy korai ISA szerkesztőségi cikkből.

Az adatbázis szerkesztői által írt, a DIALOG által publikált esszé (<http://library.Dialog.com/products/f202.html>) fényében, amely azt állította, hogy nincs semmi lemaradása az adatbázisnak, a 2000 szeptemberi tény, hogy csak 114 primer forrás volt (beszámítva a konferenciaköteteket), azt sugallja, hogy nem csak a rekordok száma csökkent, hanem a 2000-ben feldolgozott források száma is (4.34. ábra). (Mivel az adatbázis jelenleg nem érhető el a DIALOG-nál, ezért természetesen ez a dokumentum sem található a jelzett címen.

Ugyanaz a 4.33. ábrában bemutatott szerkesztőségi cikk felvázolta tervüket „az ISA által feldolgozott folyóiratok listájának jelentős mértékű növeléséről” is. Erre szemmel láthatóan nem került sor. Bár a 4.32. ábra számai valamennyire felduzzasztottak a folyóiratcímek következtlen, pontatlan, ebből következően különböző írásmódja miatt, ez egyformán érvényes a régi és az új vezetés időszakára, s nem torzítja el a képet.

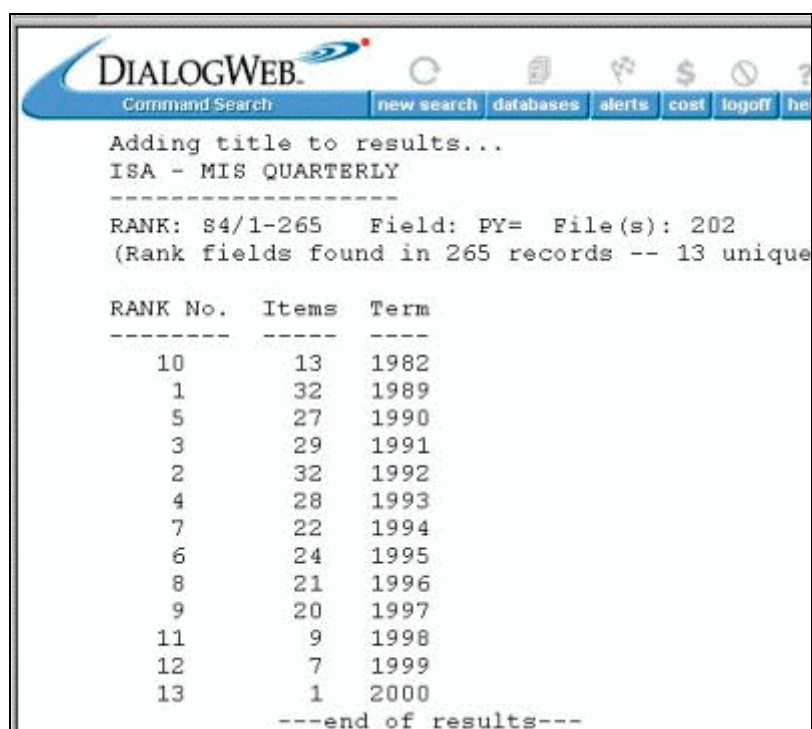
A feltártság mélysége az egyes folyóiratok szintjén is gyatrább lett. A legutolsó három év mintái ugyanazt a sekély vagy hullámvasútszerű feltártságot és hézagokat mutatják, mint a melyik az adatbázis 1998 előtti szegmensét is jellemezte. Bár a magfolyóiratok listáját többé nem teszik közzé, a feltártság mélységének megállapítására mintaként kiválasztott folyóiratok reprezentálják a legfontosabb folyóiratokat, amelyek állítólag az ISA fókuszában vannak.

Például a *MIS Quarterly* (a teljesen kiírt, az ISA által szintén használt *Management Information Systems Quarterly* cím alatt is kerestük) meredeken csúszik lefelé (4.35. ábra), annak ellenére, hogy a könyvtár- és információtudományi kategóriában a Journal Citation Reports a legmagasabbra rangsorolta 1999-ben. Ugyanez a csökkenő trend igaz a *Library Administration and Management* című folyóiratra is, amely az ISA bluesheetje (adatbázis-leírása) által elsődlegesnek jelzett szakterülethez kapcsolódik, s a Journal Citation Reports is magasan rangsorolta (4.36. ábra).

Publication year	# of "journals"	# of "journal" records	Total # of records	Depth factor
1995	224	6,421	8,921	28.67
1996	234	6,838	9,327	29.22
1997	228	5,653	6,449	24.79
1998	278	4,645	4,655	16.71
1999	234	3,618	3,618	15.46
2000	117	1,173	1,173	10.03

Kiadási év / A folyóiratok száma / A folyóiratokból származó rekordok száma / A rekordok teljes száma / Mélységtényező

4.34. ábra: Csökkenő mélységtényező 1995 és 2000 szeptembere között




The screenshot shows the DIALOGWEB Command Search interface. The search results for 'ISA - MIS QUARTERLY' are displayed, showing a rank of 84/1-265 and 202 files. The results are sorted by rank, with the top 13 unique terms listed below.

RANK No.	Items	Term
10	13	1982
1	32	1989
5	27	1990
3	29	1991
2	32	1992
4	28	1993
7	22	1994
6	24	1995
8	21	1996
9	20	1997
11	9	1998
12	7	1999
13	1	2000


---end of results---

4.35. ábra: Az *MIS Quarterly* feltártsága a lefelé vezető lejtőn van.

		
Command Search new search databases alerts cost logoff		
ISA - LIBRARY ADMINISTRATION & MANAGEMENT		
RANK: S17/1-248 Field: PY= File(s): 202 (Rank fields found in 248 records -- 13 unique)		
RANK No.	Items	Term
9	12	1988
10	12	1989
11	10	1990
5	22	1991
8	17	1992
7	20	1993
4	24	1994
1	38	1995
2	36	1996
3	26	1997
6	21	1998
12	5	1999
13	5	2000
---end of results---		

4.36. ábra: A *Library Administration and Management* csökkenő feltártsága az ISA-ban.

Az ISA adatbázis-leírásának a tárgyi feltártságot bemutató részénél első témaként sorolják fel a referálás és indexelés témáját, de az ennek szentelt *NFAIS Newsletter* 1999-ben teljes mértékben ignorálták, aztán 2000-ben újra visszavették, ahogy azt a 4.37. ábra mutatja.

Set	Term Searched	Items
S1	JN="NFAIS NEWSLETTER"	199
S2	S1/2000	33
S3	S1/1999	0
S4	S1/1998	17
S5	S1/1997	34
S6	S1/1996	36
Show Database Details for: 202: Information Science Abstracts  Bluesheet		

4.37. ábra: Az *NFAIS Newsletter* teljes 1999-es anyaga feltáratlan az ISA-ban.

Fontos folyóiratok gyenge feltártságán túl van egy kérdés azokkal a folyóiratokkal kapcsolatban is, amelyeket egyáltalán nem vesznek fel, bár nyilvánvalóan a legjobb források közé tartoznak azon a szakterületen, amelyet az adatbázis feldolgozottnak mond. Ezeket a kérdéseket a folyóiratbázissal foglalkozó 5. fejezetben tárgyaljuk meg.

Frissesség

Az online adatbázisok egyik leggyakrabban említett előnye nyomtatott megfelelőikhez képest gyorsaságuk. Az adatbázisok frissessége arra utal, hogy a primer dokumentumok megjelenése után milyen gyorsan válik elérhetővé egy rekord CD-ROM vagy online adatbázisban. Ennek a mérőszámnak az értékeléséről nagy számban jelentek meg jelentések 1960-as évek óta a könyvtár- és információtudomány nyomtatott indexelő és referáló szolgáltatásairól Gilchrist (1966), Gilchrist és Presanis (1971), Dansey (1973), Edwards (1976) és Turtle és Robinson (1974) tanulmányaiban. Bottle és Efthimiadis (1984), Ernest, Lange és Herring (1988) és Jacsó (1992a) kiterjesztette az időbeli elmaradás (késés) vizsgálatát az elektronikus termékekre, s a könyvtár- és információtudományon túlra. Jacsó (1992a) különféle technikákat mutatott be az időbeli elmaradás széles skálán való mérésére. Ezeknek a technikáknak némelyikét használta Lawrence és Lenti (1995), amikor az International Aerospace Abstracts adatbázis frissességét tesztelték, összevetve néhány hozzá hasonló adatbázissal.

A hasonló anyagot feldolgozó adatbázisok összehasonlítása különösen informatív. Hightower és Schwarzwald (1991) 24, az anyagtudományt feldolgozó adatbázis frissességét mérte. Az időbeli különbségek megdöbbentőek voltak. Míg néhány adatbázisnak sikerült a rekordok csaknem 67 százalékát felvenni az adatbázisba ugyanabban az évben, amikor az eredeti dokumentum megjelent, a Soviet Science & Technology adatbázisban egyetlen ilyen rekord sem volt. Ez megmagyarázhatja azt, hogy a „szputnyik nyomás” nélkül az NTIS is miért csak elszomorítón alacsony mértékben, 5%-ban tette a rekordokat elérhetővé abban az évben, amikor az eredeti dokumentum megjelent.

1998-ra ez az arány javult: a rekordok 18%-a jelent meg ugyanabban az évben. 1999-ben ez tovább nőtt, 24%-ra, de a mennyiség rovására. Amíg 1998-ban összesen 72143 rekordot adtak az adatbázishoz, addig 1999-ben csak 45430-at.

Soremark (1990) a MEDLINE-ba és az EMBASE-be is bekerült rekordokkal kapcsolatban azt tapasztalta, hogy az előbbiben átlagosan 2-4 hónapos volt a lemaradás. 1999-re a MEDLINE időszerűsége messze felülmúlta az EMBASE-ét - legalábbis a PubMed webhelyen, amely bevezette a PreMEDLINE rekordokat. Ezek a rekordok, amelyeket naponta adnak hozzá az adatbázishoz, alapvető bibliográfiai adatokat és referátumokat tartalmaznak, de MeSH deszkriptorokat nem. Amikor ezeket a rekordokat kiegészítik, kész MEDLINE rekordokká válnak. Erre az aktualizálásra naponta kerül sor a PubMed rendszerben. Figyelembe véve, hogy ez egy ingyenes adatbázis, nagy kincs azok számára, akik gyorsan tudomást akarnak szerezni a megjelenő publikációkról. A frissesség másik tesztje során, 1999. december 23-án *medical* keresőszó és *2000 mint megjelenési év* (PY=2000) 186 rekordot adott az EMBASE-ben a DIALOG rendszerben és 443-at a PubMedben.

Jaguszewski és Kemp (1995) négy témafigyelő szolgáltatás frissességét hasonlította össze a kémia és a matematika területén. Úgy találták, hogy az Uncover volt a legfrissebb, a továbbiak sorrendje Inside Information, ContentsFirst és Current Contents on Diskette volt, bár a kémia területén a ContentsFirst frissebb volt az Inside Informationnél.

Azok a webes adatbázisok, amelyeket közvetlenül a fájlok előállítói szolgáltatnak, drámai módon növelik az adatbázisok naprakészségét, mivel a közvetítő, az adatbázis kiadója kiiktatódik. A mostani korai szakaszban a tartalomszolgáltató által közvetlenül a weben való megjelentetés az adatbázisok közreadásának egy újabb csatornája a már létezők mellett. Hosszútávon ez a trend fenyegetheti az online szolgáltatásokat, amelyek mások tartalmát publikálják. Mivel a weben keresztül mindenki elérhető, harmadik fél tartalmát közvetítő online kiadóknak keményen kell dolgozniuk, hogy igazolják a plusz költségeket és a közvetítésből adódó elkerülhetetlen késést. Ez különösen érvényes a kis online szolgáltatásokra.

Az egy helyen való vásárlás, a több adatbázisban való keresés, a duplikátumok kiszűrése és a kifinomult és hatékony keresőgépek vonzása jó ok marad arra, hogy az adatbázisokat „szupermarketekben” is kínálják a tartalomszolgáltatók által való publikálás mellett. A használóknak azonban össze kell hasonlítaniuk a különböző hostoknál elérhető adatok naprakészségét.

Az ilyen összehasonlítások esetén figyelni kell arra, hogy minden hostnál ugyanaz-e a forrásfájl. Egyes hostok például nem használják a MEDLINE-nak a nem angol nyelvű részalmazát, amely drámai módon növelheti a naprakészséget, figyelembe véve, hogy milyen soká lehet hozzájutni az idegen nyelvű primer dokumentumokhoz.

Ha különböző adatbázisok frissességét összehasonlítják, különösen figyelni kell arra, hogy az almát almával hasonlítsuk össze. A hetente aktualizált adatbázisok nyilvánvalóan frissebbek, mint a havonta aktualizáltak, amelyek ugyanakkor a negyedévente frissítetteket múlják felül. Ezek a tényezők igazolhatják azonban, hogy miért az egyik adatbázist használják, s nem egy másik vele összehasonlíthatót. (Az, ha a tesztet azon a napon végzik el, amikor az egyik adatbázist épp aktualizálták, míg a másikat épp aktualizálás előtt áll, igazságtalanul torzíthatja az eredményeket.

A napilap adatbázisok összehasonlításakor észre kell venni, hogy a teljes szövegű változat valószínűleg közvetlenül bekerül az online szolgáltatásba, mielőtt a nyomtatott változatokat elkezdenék kinyomtatni. Azok az adatbázisok, amelyek jelentős mértékben adnak index-kifejezéseket az újságcikkek rekordjaihoz, mint például a National Newspaper Index, jóval lassabbak, mert az indexelés folyamata időbe telik. Azoknál az adatbázisoknál, amelyek referátumokat is szolgáltatnak (mint például a Bell & Howell cég Newspaper Abstracts Daily című szolgáltatása), valószínűleg nagyobb a lemaradás. Nem véletlen, hogy Husszein, Jordánia királya temetését követő napon a legtöbb rekordja a teljes szövegű digitális napilapoknak volt, amelyeknek anyagát közvetlenül az adatbázis-változatba töltik, és a legtöbb indexelő és referáló adatbázis nem is tudott a király elhunytáról.

A hírügynökségeket tekintik a legfrissebbnek az adatbázisok közül, de egy 1999. július 15-én elvégzett teszt azt mutatta, hogy nem minden hírügynökség volt percrekész, és a Canada Newswire adatbázis - amely napi többszöri frissítést ígér - 10 napos késésben volt a DIALOG-nál (4.38. ábra). Meg kell jegyezni, hogy az adatbázis-szolgáltatók az információt egy meghatározott moratórium (általában néhány óra) lejártá után kapják meg. 10 napos késés messze kívül esik minden moratóriumon. Az AP News, U. S. Newswire, AFP English Wire és az AFP French Wire egynapos lemaradásban volt, a PR Newswire két nappal maradt el. Csak a Japan Economic Newswire és a Canada Newswire volt jelentős mértékben lemaradva.

A napilapok definíciójuk szerint naponta jelennek meg, de ez nem jelenti azt, hogy szó szerint naprakészek egy harmadik fél online szolgáltatójánál. Egy gyors pillantás a DIALOG adatbázisainak Papers (Napilapok) kategóriájának szalagcímeire, bannereire, s máris benyomást szerezhetünk arról, mennyire frissek az adatbázisok. Sajnos a bannerek nem mindig tükrözik megbízhatóan az aktualizálás dátumát. Pedig úgy kellene lennie, mert az aktualizálás dátuma automatizált folyamat részeként adódik hozzá a bannerhez. Egy adatbázis aktualizálásának legmegbízhatóbb ellenőrzése az, ha olyan keresést folytatunk le, amelyben aznapi adatokra kérdezzünk (vagy talán még az előző egy vagy két napra is).

Az 1999. december 23-án elvégzett ilyen keresés megmutatta, hogy a legtöbb napilap adatbázis valóban naprakész volt, s csak néhány késett egy vagy több napot. A *San Jose Mercury* tűnt a leginkább késésben levőnek. A banner a legutóbbi aktualizálás dátumaként 1999 december 18-át jelölte meg. Az aktualizálás mezőjében az UD=19991223, UD=1991222 és az UD=19991221 keresés azt igazolta, hogy azt utoljára 1999 december 21-én aktualizálták (4.39. ábra) - nem túl jó egy, a Szilícium-völgyből származó napilap számára, de nem olyan rossz, mint amit a banner sugall. Tapasztalataim szerint a bannerek dátumai körülbelül az

esetek 80 százalékában pontosak a DIALOG-nál az összes olyan adatbázisra vonatkozóan, amely tartalmaz könyvtár- és információtudományi anyagot tartalmaz.

Az Ovid rendszer használói már belépéskor látják, mennyire frissek az adatbázisok, mert az adatbázisok listája a dátum bannerrel együtt automatikusan megjelenik (4.40. ábra). (Az Ovid kérésre e-mail üzeneteket is küld az adatbázisok frissítéséről.) A SilverPlatter és az H. W. Wilson cégeknek hasonló nyitó képernyőjük van, amely információt ad a használóknak az adatbázisok frissítésének helyzetéről (4.41. és 4.42. ábra). A DIALOG-ban az adatbázisok és a bannerek kilistázása nem automatikus, a használónak kell kezdeményeznie az adatbázisok kiválasztásával és a SHOW FILES paranccsal (4.43. ábra).

DIALINDEX®	
File	Name
-----	-----
258:	AP News Jul_1984-1999/Jul 14
261:	UPI News_1999-1999/Jul 15
605:	U.S. Newswire_1999-1999/Jul 14
606:	Africa News_1999-1999/Jul 15
607:	ITAR/TASS News_1999-1999/Jul 15
609:	Bridge World Markets News_1989-1999/Jul 15
610:	Business Wire_1999-1999/Jul 15
612:	Japan Economic Newswire(TM)_1984-1999/Jul 09
613:	PR Newswire_1999-1999/Jul 13
614:	AFP English Wire_1991--1999/Jul 14
615:	AFP Intl French Wire_1991--1999/Jul 14
616:	Canada NewsWire_1999-1999/Jul 05
618:	Xinhua News_1999-1999/Jul 15
649:	NEWSWIRE ASAP(TM)_1999/Jul 15

4.38. ábra: A hírügynökségi adatbázisok listája az aktualizálás bannerével

Dialog Response	
File 146:	WASHINGTON POST ONLINE 1983-1999/DEC 23 (c) 1999 WASHINGTON POST
File 471:	NEW YORK TIMES FULLTEXT-90 DAY 1999/DEC 22 (c) 1999 THE NEW YORK TIMES
File 489:	THE NEWS-SENTINEL 1991-1999/Dec 22 (c) 1999 FT. WAYNE NEWSPAPERS, INC
File 490:	TALLAHASSEE DEMOCRAT 1993- 1999/Dec 22 (c) 1999 TALLAHASSEE DEMOCRAT
File 492:	ARIZONA REPUB/PHOENIX GAZ 1986-1999/Dec 22 (c) 1999 PHOENIX NEWSPAPERS
File 494:	ST LOUISPOST-DISPATCH 1988-1999/Dec 21 (c) 1999 ST LOUIS POST-DISPATCH
File 496:	THE SACRAMENTO BEE 1988-1999/Dec 21 (c) 1999 SACRAMENTO BEE
File 498:	DETROIT FREE PRESS 1987-1999/Dec 21 (c) 1999 DETROIT FREE PRESS INC.
File 630:	LOS ANGELES TIMES 1993-1999/DEC 22 (c) 1999 LOS ANGELES TIMES
File 631:	BOSTON GLOBE 1980-1999/Dec 22 (c) 1999 BOSTON GLOBE
File 633:	PHIL.INQUIRER 1983-1999/Dec 21 (c) 1999 PHILADELPHIA NEWSPAPERS INC
File 634:	San Jose Mercury Jun 1985-1999/Dec 18 (c) 1999 San Jose Mercury News
File 638:	NEWSDAY/NEW YORK NEWSDAY 1987-1999/Dec 22 (c) 1999 NEWSDAY INC.

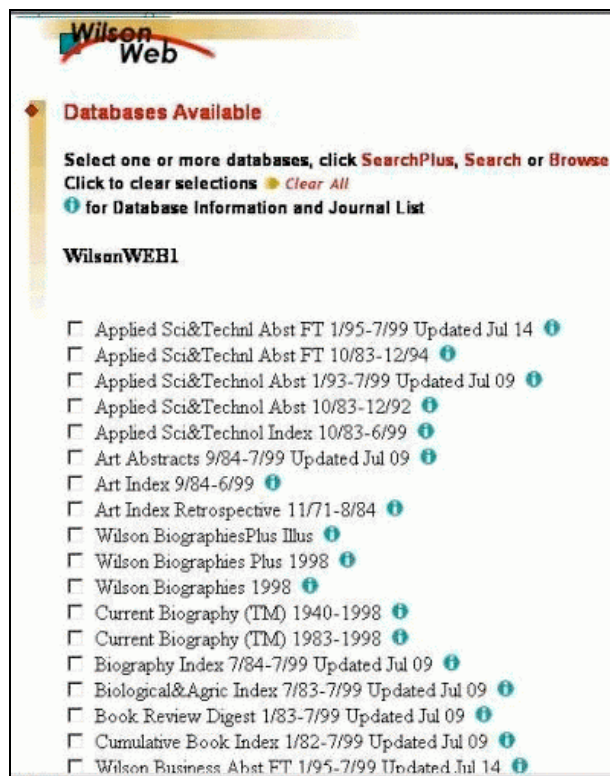
4.39. ábra: Az aktuális nap feldolgozottságának ellenőrzése a napilap-adatbázisokban.

	Wilson Art Abstracts	1983 to May 1999
	Wilson Applied Science and Technology	1983 to May 1999
	Wilson Biological & Agricultural Index	1983 to June 1999
	Wilson Business Abstracts	1983 to June 1999
	Wilson General Science Abstracts	1984 to May 1999
	Wilson Humanities Abstracts	1984 to June 1999
	Wilson Index to Legal Periodicals	1983 to June 1999
	Wilson Library Literature	1984 to June 1999
	Wilson Readers Guide Abstracts	1983 to June 1999
	Wilson Social Sciences Abstracts	1984 to June 1999

4.40. ábra: Részlet az Ovid adatbázis listájáról a dátumok bannerével.



4.41. ábra. Részlet a SilverPlatter adatbázis listájáról a dátumok bannerével.



4.42. ábra. Részlet az H. W. Wilson cég adatbázis listájáról a dátumok bannerével.

A korábbi képernyőképek azt is illusztrálják, milyen különbségek vannak az adatbázisok frissítésének lemaradásában a különböző hostoknál. Nem meglepő, hogy a fájl előállítójának változata volt a legfrissebb az H. W. Wilson cég Biological & Agricultural Index, Art Abstracts, Applied Science & Technology, továbbá a Business Abstracts esetében. A Business Abstracts, Biological & Agricultural Index és a Humanities Abstracts esetében az Ovid és a DIALOG azonos volt, a Readers' Guide Abstracts esetében az Ovid jobb volt a DIALOG-nál, míg a DIALOG volt jobb az összes többi fájlnál az 1999 júniusi frissítéseknél.

DIALINDEX®	
File	Name
98:	General Sci Abs/Full-Text_1984-1999/Jun
99:	Wilson Appl. Sci & Tech Abs_1983-1999/Jun
141:	Readers Guide_1983-1999/May
142:	Social Sciences Abstracts_1983-1999/Jun
143:	Biol. & Agric. Index_1983-1999/Jun
435:	Art Abstracts_1984-1999/Jun
436:	Humanities Abs Full Text_1984-1999/Jun
437:	Education Abstracts_1983-1999/Jun
438:	Library Literature_1984-1999/Jun
553:	Wilson Bus. Abs. FullText_1982-1999/Jun

4.43. ábra: Részlet a DIALOG adatbázisainak listájáról a dátumok bannerével.

Egy adatbázis frissítésének következetessége természetesen alapkövetelmény a rekordok megfelelő időben való elérhetőségéhez. Ha megnézzük az aktualizálás indexét, ez némileg fényt derít az alkalmazott mintára, de nem mondja el a teljes történetet. Az (update, aktualizálás) mező értéke jelzi az adatbázishoz adott rekordok halmaza esetében az évet és hetet, hónapot vagy negyedévet, vagy csak a tervezett aktualizálási dátumot. Néha ez egybeesik a tényleges aktualizálási idővel, néha nem. Mindazonáltal az UD index váratlan hézagokat mutathat.

Ez a helyzet az ISA adatbázis esetében, amely, bár évi 11 aktualizálást ígért, arra csak 8 alkalommal került sor 1999-ben (4.44. ábra). 2000 elején az ISA bejelentette, hogy az adatbázis aktualizálásainak számát évi 11-ről 9-re csökkentik; körülbelül 6 hét különbség lesz köztük. Ez vágyálomnak bizonyult. Az adatbázis aktualizálása még rendszertelenebb, ötlet-szerűbb volt 2000-ben, mint azelőtt. 2000 novemberéig az ISA-t hatszor aktualizálták, s nem egyenletes megoszlásban.

A LISA (4.45. ábra) és a Library Literature (4.46. ábra) adatbázisokat 12 alkalommal aktualizálták 1999-ben, s a frissítések nagyon rendszeresek voltak mindkét adatbázis esetében. 1998-ban azonban a LISA is komoly hézagokat mutatott. Az UD mezők szerint nem került sor aktualizálásra januárban, februárban, áprilisban, májusban, júniusban és novemberben, a márciusi aktualizálás is minimális volt (4.45. ábra). A Library Literature-t mind 1998-ban, mind 1999-ben az előre tervezett ütemben aktualizálták (4.46. ábra).

Set	Term Searched	Items
S1	UD=199901	609
S2	UD=199902	469
S3	UD=199903	433
S4	UD=199904	495
S5	UD=199905	0
S6	UD=199906	452
S7	UD=199907	0
S8	UD=199908	422
S9	UD=199909	0
S10	UD=199910	385
S11	UD=199911	917
S12	UD=199912	0

Show Database Details for:

202: Information Science Abstracts [Bluesheet](#)

4.44. ábra: Az ISA aktualizálásának mintája 1999-ben.

Set	Term Searched	Items
S1	UD=199901	1328
S2	UD=199902	826
S3	UD=199903	1201
S4	UD=199904	888
S5	UD=199905	1051
S6	UD=199906	1092
S7	UD=199907	1318
S8	UD=199908	1163
S9	UD=199909	1201
S10	UD=199910	997
S11	UD=199911	1708
S12	UD=199912	1152

Show Database Details for:

61: LISA (Library & Information Science Abstracts) [Bli](#)

4.45. ábra: A LISA aktualizálásának mintája 1999-ben.


Set	Term Searched	Items
S1	UD=199901	1088
S2	UD=199902	1442
S3	UD=199903	2268
S4	UD=199904	3076
S5	UD=199905	1448
S6	UD=199906	1632
S7	UD=199907	1444
S8	UD=199908	1678
S9	UD=199909	1851
S10	UD=199910	1215
S11	UD=199911	1616
S12	UD=199912	1311

Show Database Details for:
438: Library Literature
Bluesheet
Rates

4.46. ábra: A Library Literature aktualizálásának mintája 1999-ben.

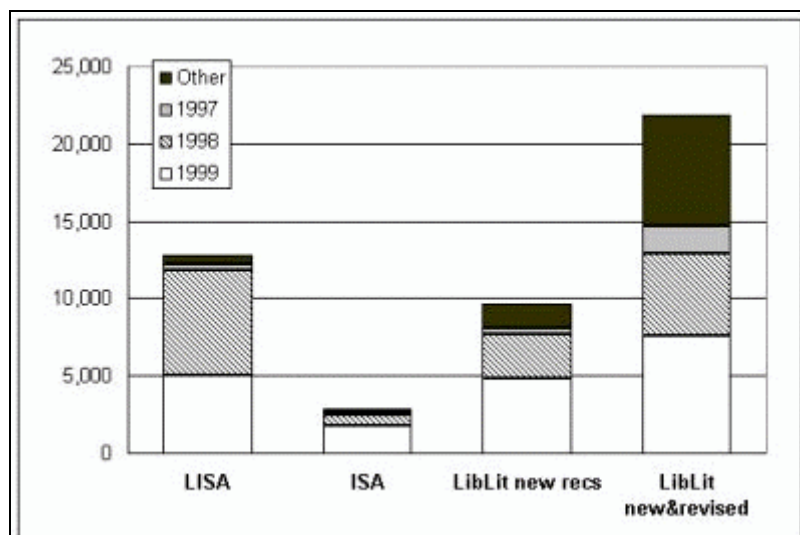
Azok számára, akik adatbázisaikat rendszeresen aktualizálták, az UD index megbízható jelzője volt az aktualizálás mintájának. Egyes adatbázis-készítők azonban kihagynak egy vagy több hónapot az aktualizálásból, aztán pedig egyszerre szolgáltatják több hónap anyagát, hogy utolérjék magukat. Az adatbázishoz adott rekordok esetében az UD mezőket ahhoz a hónaphoz igazítják, amikor az aktualizálást el kellett volna végezni. Ezek után a használó nem tudja megmondani, hogy az aktualizálás mintája valóban olyan volt-e, amilyennek tűnik. Ez ahhoz hasonló, mint amikor egy bébiszitter a tervezett öt vagy hat etetést elmulasztja, végül az összes ételt egyszerre lenyomja a bébi torkán este 10-kor, mielőtt a szülők hazaérnek. Így a bébi megkapott minden ételt, ami járt neki arra a napra, vagy nem?

A *Database* hasábjain Allcock (1997) és Jacsó (1997c) között lefolyt vita után a DIALOG számos adatbázisában bevezette a ZD mezőt. Ezt automatikusan generálják, amikor az adatbázist ténylegesen aktualizálják. Tükrözi a pontos dátumot és az adatbázishoz adott rekordok számát. Az UD és a ZD indexek tételeinek összehasonlítása hajszálpontosan megmutatja a különbségeket (4.47. ábra). Az ISA esetében a januárban esedékes aktualizálásra április 9-én került sor, a februárira május 20-án, a márciusra június 7-én és így tovább. A májusi aktualizálást kihagyták, és minden bizonnyal a júliusi volt az eleve kihagyásra szánt hónap. Ezt használhatták volna fel arra, hogy valamennyire behozzák a lemaradást és növeljék az új rekordok mennyiségét, amely alig volt több az előző évek 50 százalékánál.

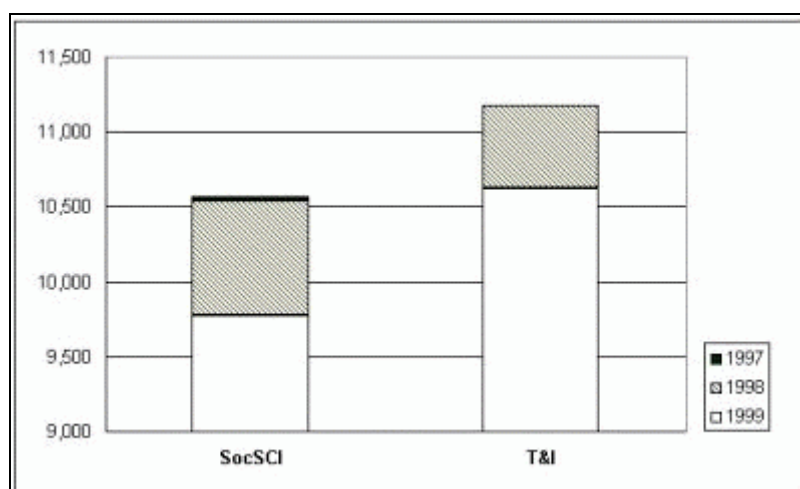
 new search databases alerts		
Database Details		
Set	Term Searched	Items
S1	UD=199901	609
S2	ZD=19990409	609
S3	UD=199902	469
S4	ZD=19990520	469
S5	UD=199903	433
S6	ZD=19990607	433
S7	UD=199904	495
S8	ZD=19990713	495
S9	UD=199905	0
S10	UD=199906	452
S11	ZD=19991022	452
Show Database Details for: 202: Information Science Abstracts Bluesheet		

4.47. ábra: AZ ISA UD és ZD indexei értékeinek összehasonlítása a tényleges aktualizálási minták megállapítása érdekében

Még egy „sikítóan tökéletes” aktualizálási minta sem garantálja a kurrens információt. Lehetséges, hogy egyes adatbázisokban egy adott év második vagy harmadik aktualizálása sem tartalmaz egyetlen rekordot sem abból az évből, még ha az aktualizálásra havonként is kerül sor. Másik oldalról azonban más havonta aktualizált adatbázisok esetében már az első aktualizálás is tartalmazhat cikkeket, amelyeket abban az évben adtak ki. A 4.48. és 4.49. ábra egyrészt a három könyvtár- és információtudományi adatbázis, másrészt a Social SciSearch és a Trade & Industry adatbázisok könyvtár- és információtudományi részhalmaza 1999-es teljes aktualizálási korpuszának összetételét illusztrálja. A diagramok megmutatják, hogy a rekordok mekkora százaléka vonatkozott az adott év dokumentumaira, az előző évekre, a két évvel azelőttiekre és a régebbiekre. Az ISA-hoz hozzáadott rekordok csekély mennyisége különös figyelmet érdemel, mivel az körülbelül 30%-a az előző évek aktualizálási mennyiségének.



4.48. ábra: A rekordok megjelenési év szerinti megoszlása a könyvtár- és információtudományi adatbázisok 1999-es aktualizálásaiban.



4.49. ábra: A rekordok megjelenési év szerinti megoszlása a könyvtár- és információtudományi anyagot tartalmazó adatbázisok 1999-es aktualizálásaiban.

Az időbeli elmaradás mérésének másik módja megfelelő nagy mintán az, hogy kiválasztjuk egy adott évben publikált rekordok ezreit vagy akár tízezreit, s meghatározzuk, milyen az aktualizálási évek megoszlása azokra a rekordokra vonatkozóan - azaz, mennyit adtak hozzá közülük az adatbázishoz a megjelenés évében, egy évvel később, két év múlva, három év múlva vagy akár több évvel később. A 4.50. és 4.51. ábrák ennek a technikának az eredményeit illusztrálják az ISA és a LibLit összehasonlításával. A LISA nem volt használható erre a tesztre, mert 1997 novembere és 1998 júniusa között számos aktualizálás anyagát vonták össze egyetlen nagy tétellé (UD=199711-199806) az UD indexben, ez 7905 rekordot tartalmaz. Ez azt jelenti, hogy nem volt mód arra, hogy meghatározzuk az 1997 novembere és 1998 júniusa közti aktualizálás különbségeit, vagy bármely dátumét a kettő között.

Mindig ellenőrizni kell az ilyen egyedi sajátosságokat, mivel komolyan torzíthatják az eredményeket. A LibLit adatbázisra vonatkozóan csak azokat a rekordokat elemeztük, amelyekben megvolt a ST=NEW RECORD mező értéke, mivel az H. W. Wilson cég jelentős

rekordellenőrző műveletet hajtott végre (de nem a kurrens tételek rovására). Néhány kisméretű korrekciós futtatásnak nincs hatása az eredményekre.

DISTRIBUTION OF UPDATE YEARS FOR PY=1990 PUBLICATIONS IN ISA			

RANK: S1/1-8421 Field: BN= File(s): 202			
(Rank fields found in 8421 records -- 10 unique terms)			
RANK No.	Items	%Ranked	Term
-----	-----	-----	----
2	3132	37.2%	90
1	4111	48.8%	91
3	716	08.5%	92
4	301	03.6%	93
9	13	00.2%	94
6	38	00.5%	95
8	19	00.2%	96
10	2	00.0%	97
5	69	00.8%	98
7	20	00.2%	99
---end of results---			

4.50. ábra: Tesztkérdés a publikáció évéhez képesti aktualizálás késés meghatározására az ISA-ban

DISTRIBUTION OF UPDATE YEARS FOR PY=1990 PUBLICATIONS IN LIBLIT			

RANK: S1/1-15301 Field: AA= File(s): 438			
(Rank fields found in 15301 records -- 5 unique terms)			
RANK No.	Items	%Ranked	Term
-----	-----	-----	----
1	7199	47.0%	BLIB90
2	6727	44.0%	BLIB91
3	1034	06.8%	BLIB92
4	328	02.1%	BLIB93
5	13	00.1%	BLIB94
---end of results---			

4.51. ábra: Az aktualizálás késése a publikáció évéhez képest a LIBLIT adatbázisban.

Az igaz, hogy a következő évben való aktualizálás egészen más dolog egy olyan publikáció esetében, amely decemberben jelent meg s a következő januárban került be az adatbázisba, mint amelyik 1997 februárjában jelent meg és a következő év novemberében került be az adatbázisba. Ha sok rekorddal dolgozunk, akkor az ilyen szélsőségek kiegyenlítődnek.

Az ilyen technikák vonzereje az, hogy nagyon könnyű finomítani a halmazt és megismételni a keresést. Csak néhány percbe telik, ha ellenőrizni akarjuk például azt, hogy van-e különbség a késés mintájában a konferencia-előadások és a folyóiratcikkek között, vagy az angol nyelvű és az idegen nyelvű rekordok között.

Érdeemes ilyen megkülönböztetéseket tenni, mert egy nemzetközi adatbázis, mint amilyen a LISA, sok folyóiratot dolgoz fel fejlődő és kevésbé fejlett országokból. Ezekben az esetekben a késések nagy része a forráspublikációk késői megjelenésének tulajdonítható (például a januári szám áprilisban jelenik meg). A folyóiratok postai küldéséből adódó késések is jelentősek lehetnek. Természetesen az USA-ban megjelenő kiadványok, különösen a tudományos folyóiratok, köztük a könyvtár- és információtudományiak, szintén gyakran késnek. Néha a késés olyan nagy, hogy a kiadó két számot összevon (spórolva a postaköltségen, de nagy gondot okozva az időszaki kiadványok számbavételében).

Arra is figyelni kell, hogy ne számítsuk be a korrekciós rekordokat, amelyek az eredeti rekordnál hónapokkal később jelenhetnek meg egy adatbázisban. Ez akkor lehetséges, ha a korrekciós rekordokat valamilyen speciális szimbólummal jelölik meg az aktualizálás dátuma után, mint az UD=9902C esetben, vagy speciális mezőt használnak, ahogy az H. W. Wilson cég teszi az ST=NEW RECORD (új rekord) és az ST=REVISION OR CORRECTION RECORD (revideált vagy korrigált rekord) indextételekkel.

Még egy másik lehetőség a késés megállapítására a folyóiratok szintjén az, hogy annak az adatelemnek a számait használjuk, amely a rekordok létrehozásának és bevitelének évét azonosítja. Az ilyen adatelemek nem elérhetőek minden adatbázisban, néhányban azonban igen. Az Information Science Abstractsban például a BN adatelem első két számjegye az adatbázis nyomtatott változatának évét mutatja, így nagyon könnyű rekonstruálni, hogy milyen késéssel vittek be az adatbázisba rekordokat olyan folyóiratokból, amelyeket az IFI/Plenum magfolyóiratoknak mondott.

Az *RQ* című folyóiratból származó eredmények halmazának az aktualizálás éve szerinti rangsorolása az aktualizálás mintáinak kimutatása érdekében (4.52. ábra) nem kelti azt a benyomást, hogy a lap a magfolyóiratoknak kijáró kezelésben részesült volna. Csak 1990-től dolgozzák fel (ebből az évből csupán hat rekord került be belőle az adatbázisba jókora késéssel), azután semmi 1994-ben, s újra csupán 4 rekord 1996-ban. Az ISA adatbázis fő jellegzetességeinek értékelésekor Jacsó (1998a) az *RQ*-t használta olyan folyóirat példaként, amelynek teljesen dilettáns kezelése különös egy magfolyóiratnak tekintett tételnek is. Az *RQ*-ból (és néhány más magfolyóiratból) származó rekordok 1997-ben, 1998-ban és 1999-ben kerültek be - megkésve - az ISA-ba, így most egyes évek már jól fel vannak dolgozva (4.53. ábra).

Ez a késői hozzáadás magyarázza meg, miért került be olyan sok új rekord az *RQ*-ból ezekben az években, annak ellenére, hogy a folyóirat 1997 közepén beszüntette megjelenését ezen a címen. A hézagok sietős betömődése azonban nem növelte az adatbázis frissességét. Sok folyóiratból nem kerültek be új rekordok. A rekordok éves mennyisége rendkívüli mértékben csökkent, s 1999-ben minden idők legalacsonyabb értékét érte el.

Az *RQ* aktualizálási évének és kiadási évének mintája a LibLit adatbázisban sokkal kedvezőbb frissességet mutat (4.54. ábra). Ideális esetben az aktualizálás évét és a kiadás évét reprezentáló oszlopoknak csaknem azonosaknak kellene lenniük minden egyes évben. Ugyanaz az aszinkronitás, amit az *RQ*-val kapcsolatban láttunk, jellemző sok más „magfolyóiratra” is az ISA-ban, köztük a *CD-ROM Professionalre*, amely ugyancsak más címen jelenik meg 1996 utántól. Ahogy azt a 4.55. ábra mutatja, több rekord került be az ISA-ba 1997-ben és 1998-ban (a folyóirat ilyen címen való megszűnése utáni két évben), mint a megelőző hat évben összesen. Bár ez kétségkívül késői ébredés, jobb későn, mint soha.

Dialog Response		
Adding title to results...		
RQ UPDATES IN ISA FOR 1990-1997 BY UD YEAR NUMBER		

RANK: S6/1-284 Field: BN= File(s): 202		
(Rank fields found in 284 records -- 9 unique terms)		
RANK No.	Items	Term
-----	-----	-----
8	6	90
5	30	91
4	31	92
7	23	93
3	33	95
9	4	96
1	78	97
2	53	98
6	26	99
---end of results---		

4.52. ábra: Az *RQ* aktualizálási mintája az ISA-ban

DialogWeb Command Mode	Dialog Response	http://www.dialogweb.com/cgi/twclient
Adding title to results...		
RECORDS OF RQ BY PUBLICATION YEAR		

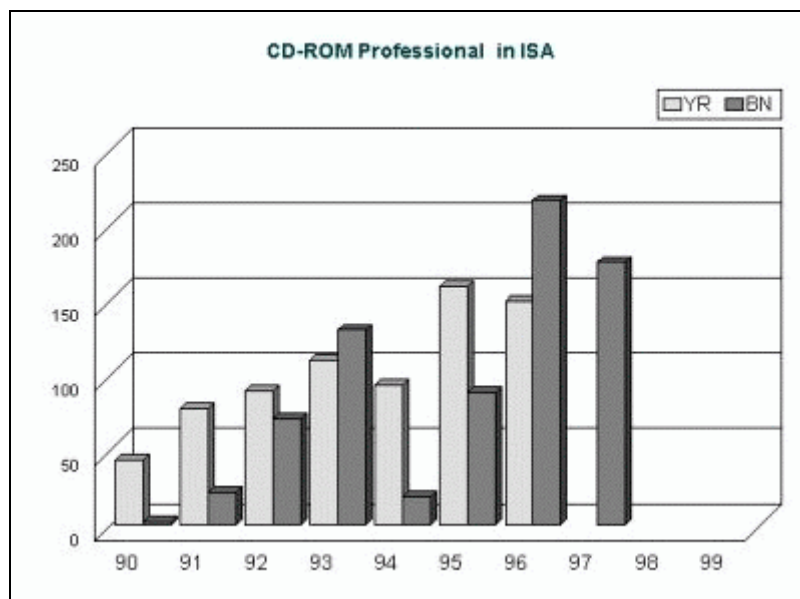
RANK: S1/1-710 Field: PT= File(s): 202		
(Rank fields found in 710 records -- 28 unique terms)		
RANK No.	Items	Term
-----	-----	-----
27	7	1970
24	13	1971
16	24	1972
22	19	1973
26	12	1974
23	15	1975
17	21	1976
28	5	1977
25	13	1978
18	21	1979
10	27	1980
12	26	1981
6	35	1982
20	20	1983
11	27	1984
15	25	1985
13	26	1986
8	30	1987
7	32	1988
9	28	1989
14	26	1990
19	21	1991
1	56	1992
3	40	1993
5	36	1994
4	37	1995
2	48	1996
21	20	1997
---end of results---		
P = next page Pn = Jump to page n		
P- = previous page M = More Options Exit = Leave RANK		
To view records from RANK, enter VIEW followed by RANK number, format, and item(s) to display, e.g., VIEW 2/9/ALL.		
Enter desired option(s) or enter RANK number(s) to save terms.		

© 1998 The Dialog Corporation plc		

4.53. ábra: Az *RQ* kiadási év mintája az ISA-ban

rqinliblit.gif [1:1] (Background)			Dialog R		
Adding title to results. RQ RECORDS BY PUBLICATIO			Adding title to results... RQ RECORDS ADDED TO LIBLIT - RA		
RANK: S2/1-1354 Field: (Rank fields found in 13			RANK: S2/1-1354 Field: AA= F (Rank fields found in 1354 reco		
RANK No.	Items	Term	RANK No.	Items	Term
13	56	1984	1	146	BLIB85
1	126	1985	4	118	BLIB86
3	114	1986	9	89	BLIB87
5	112	1987	2	145	BLIB88
4	113	1988	5	114	BLIB89
2	118	1989	6	113	BLIB90
6	112	1990	7	100	BLIB91
9	100	1991	11	80	BLIB92
7	110	1992	3	144	BLIB93
8	107	1993	8	94	BLIB94
10	88	1994	13	51	BLIB95
11	85	1995	10	85	BLIB96
12	79	1996	12	75	BLIB97
14	34	1997			
---end of res			---end of results---		

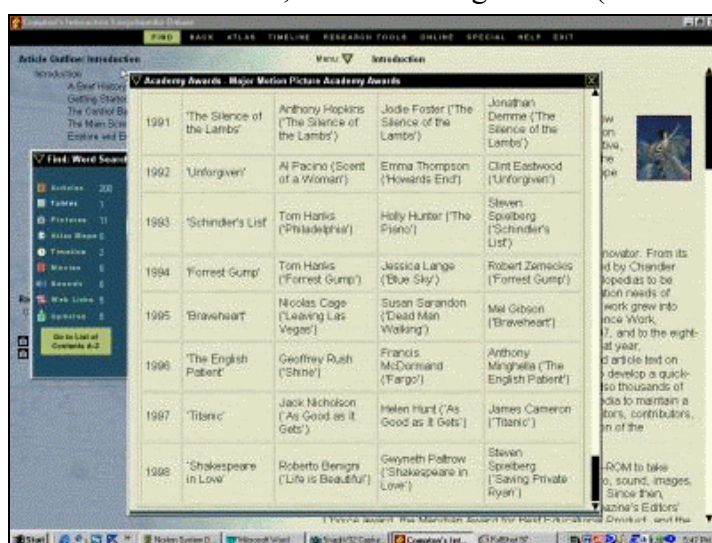
4.54. ábra: Az RQ tényleges aktualizálási éve és kiadási éve mintája a LibLit adatbázisban.



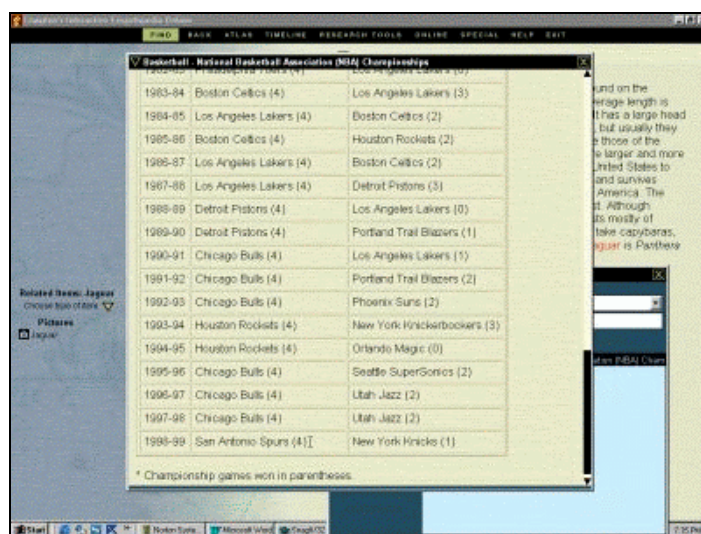
4.55. ábra: A CD-ROM Professional tényleges aktualizálási éve és kiadási éve mintája az ISA-ban.

Az adatbázisok naprakészsége kritikusabb kérdés, mint retrospektivitása, mivel a legtöbb használó számára a friss információ fontosabb, mint a történeti jellegű. A használók gyakran hajlandók is többet fizetni a kurrens információért. Ezt világosan illusztrálják az átíró, a műsorok írott szövegét nyújtó szolgáltatások árai, sokkal többet kérnek a friss TV- vagy rádióműsorok szövegéért, mint az egyhetes vagy egyhónapos anyagokért. Néha egy adatbázisnak mondjuk az 1966 és 1982 közötti retrospektív részhalmaza sokkal kevesebbe kerül, mint az 1983-1999-es szegmentum.

A címtár, enciklopédia vagy biográfiai adatbázisok esetében a frissesség ellenőrzésének legjobb módja az, ha megnézzük, hogy egy személlyel, céggel, országgal vagy folyóirattal kapcsolatos, valóban kurrens információk megtalálhatóak-e az adatbázisban. Például a legfőbb CD-ROM-os enciklopédiák esetében, amelyeket 1999 nyarának végén jelentettek meg, a frissesség tesztelésének egyik kézenfekvő módja, ha megnézzük, szerepeltek-e bennük az 1998-as Oscar-díjas filmek (1999 márciusában közzéttek), az 1999-es Pulitzer-díj nyertesei a dráma és költészet kategóriájában (1999. április közepe) és az NBA 1999-es bajnoka (az 1999. június 25-i mérkőzésen döntött el, ki lesz). A három nagyközönségnek szóló enciklopédia közül egyikben sem voltak benne az 1999-es Pulitzer-díjak. A *Compton's* bizonyult a legfrissebbnek a másik két teszt alapján (4.56. és 4.57. ábra.) A *Microsoft Encarta* volt a második legjobb, ahol a *Compton's*hoz hasonlóan a friss adatok vonzóan megtervezett táblázatokba kerültek be (4.58. ábra). Meglepetésre a különben nagyon jó *Grolier Encyclopedia* volt messze a legrosszabb mindhárom tesztben. Az Oscar-díj címszónál 1990-től máig ígéri a listát, ám az utolsó lista az 1997-es (1998 márciusában bejelentett) győztesek névsora. A rosszul megtervezett táblázat más kérdés, de az is lehangoló volt (4.59. ábra)



4.56. ábra: Az 1999 nyarán megjelent *Compton's* 2000-ben benne voltak az 1999 márciusában bejelentett 1998-as Oscar-díjas filmek.



4.57. ábra: Az 1999 nyarán megjelent *Compton's* 2000-ben benne volt az NBA 1999-es NBA bajnok.

Microsoft Encarta Encyclopedia Deluxe 2000

Best Picture and Director

Table Section 1 of 17

YEAR	BEST PICTURE	BEST DIRECTOR
1984	Amadeus	Miloš Forman, Amadeus
1985	Out of Africa	Sydney Pollack, Out of Africa
1986	Platoon	Oliver Stone, Platoon
1987	The Last Emperor	Bernardo Bertolucci, The Last Emperor
1988	Rain Man	Barry Levinson, Rain Man
1989	Driving Miss Daisy	Oliver Stone, Born on the Fourth of July
1990	Dances With Wolves	Kevin Costner, Dances With Wolves
1991	The Silence of the Lambs	Jonathan Demme, The Silence of the Lambs
1992	Unforgiven	Clint Eastwood, Unforgiven
1993	Schindler's List	Steven Spielberg, Schindler's List
1994	Forrest Gump	Robert Zemeckis, Forrest Gump
1995	Braveheart	Mel Gibson, Braveheart
1996	The English Patient	Anthony Minghella, The English Patient
1997	Titanic	James Cameron, Titanic
1998	Shakespeare in Love	Steven Spielberg, Saving Private Ryan

Academy Award Winners

4.58. ábra: A *Microsoft Encarta 2000*-ben benne voltak az 1998-as Oscar-díjas filmek

Year 2000 Grolier Multimedia Encyclopedia

Articles Media Atlas Timelines Research Starters Interactivities

Text Table Related Articles

Academy Awards (Major), 1990-Present

Actor: Tom Hanks, Forrest Gump
 Actress: Jessica Lange, Blue Sky
 Sup. Actor: Martin Landau, Ed Wood
 Sup. Actress: Dianne Wiest, Bullets over Broadway
 1996
 Film: Braveheart
 Director: Mel Gibson, Braveheart
 Actor: Nicholas Cage, Leaving Las Vegas
 Actress: Susan Sarandon, Dead Man Walking
 Sup. Actor: Kevin Spacey, The Usual Suspects
 Sup. Actress: Mira Sorvino, Mighty Aphrodite
 1996
 Film: The English Patient
 Director: Anthony Minghella, The English Patient
 Actor: Geoffrey Rush, Shine
 Actress: Frances McDormand, Fargo
 Sup. Actor: Cuba Gooding Jr., Jerry Maguire
 Sup. Actress: Juliette Binoche, The English Patient
 1997
 Film: Titanic
 Director: James Cameron, Titanic
 Actor: Jack Nicholson, As Good As It Gets
 Actress: Helen Hunt, As Good As It Gets
 Sup. Actor: Robin Williams, Good Will Hunting
 Sup. Actress: Kim Basinger, L. A. Confidential

*Winners are selected in the spring of the following year. For example, The English Patient was awarded an Oscar on Mar. 24, 1997, as the best motion picture of 1996.

4.59. ábra: A *Grolier 2000*-ben csak 1997-ig voltak benne az Oscar-díjas filmek.

Az online adatbázisok és online enciklopédiák előnyben vannak, mert folyamatosan és elég gyakran aktualizálhatják őket. A Jordánia királyára, Husszeinre vonatkozó tételeket három hónappal halála után néztem meg, majd minden azt követő hónapban, ez volt az egyik teszt, amit a frissesség ellenőrzésére használtam. Különösen kiábrándító volt, hogy az A&E Biography adatbázis (<http://www.biography.com>) még 10 hónappal halála után sem aktualizálta a tételt (4.60. ábra), s nem volt tisztában hat 1999-ben elhunyt Nobel-díjas közül ötnek a halálával.



4.60. ábra: Husszein király elavult életrajza 1999 végén.

Persze minden relatív. A *Columbia Encyclopedia* harmadik kiadásának az Electric Library-ben található változatában Husszein király uralkodik, Frank Sinatra énekel, Sir Georg (nem George) Solti vezényel 2000 szeptemberében. Az ötödik kiadás, amelyet 2000 márciusában váltott fel a hatodik kiadás, szintén elérhető ingyen az Information Please weboldalon keresztül (<http://www.infoplease.com>), itt friss információ volt található minden teszthez kapcsolódóan. A legtöbb használó csak a „Columbia Encyclopedia” logót látja, s feltehetőleg nem tud a két verzió frissességbeli (és egyéb) különbségeiről. Sok használó örökké peches maradt volna, mert begépelve magát az *encyclopedia* szót URL-ként (akár a *www* előtag és a *com* szuffix utótag nélkül) az enciklopédia Electric Library-beli változatához jutottak el, amelyet szerencsére szintén frissítettek 2001-ben és amely a *Columbia Encyclopedia* 6. kiadásának változtatás, rövidítés nélküli szövegét kínálja.

5. Az adatbázisok által feltárt források

Az egy adatbázisban feltárt források utalnak a fájl előállítója által feldolgozott primer források típusára, műfajára, nyelvére és földrajzi származására. A feltárt folyóiratok köre különösen fontos, mert minden területnek megvannak a magfolyóiratai, amelyek a legnagyobb példányszámban jelennek meg, a legnagyobb presztízzsel rendelkeznek a szakmában, s amelyeket leginkább elismernek az egyetemi körök. A „legnagyobb példányszám” szempontját természetesen óvatosan kell megközelíteni, mert a bulvárlapok a legnagyobb példányszámú kiadványok közé tartoznak, de nem túl relevánsak semmilyen szakma számára, kivéve talán a szórakoztatóipart. Ezen kívül van néhány olyan túlértékelt lektorált folyóirat, amely 300 vagy 400 példányban jelenik meg, s mivel gyakran hiányzik belőlük a releváns tartalom, még sokkal alacsonyabb lenne a példányszámuk, ha nem egyesületi tagság fejében küldenék el, hanem csak előfizetéses úton terjesztenék őket.

A források típusai

A feldolgozott források típusai fontos megkülönböztető tényezők lehetnek azonos szakterületet feldolgozó adatbázisok között, vagy „ugyanannak” az adatbázisnak a különböző változatai között. Érdekes módon ez a fontos megkülönböztető tényező rejtve maradhat az adatbázis tipikus használója előtt, ha az adatbázis készítője nem hangsúlyozza ezt. A dokumentumokat feldolgozó (azaz indexelő, referáló és teljes szövegű) adatbázisok forrásainak körét általában könnyű meghatározni. Az útmutató adatbázisok forrásainak értékelése problematikusabb, mert nagyobb a változatosság és a többértelműség abból a szempontból, hogy mit minősítsünk forrásdokumentumnak, témának vagy tárgynak.

Útmutató adatbázisok

Kevés használó veszi észre, hogy a Bowker cég adatbázisának a DIALOG-nál elérhető változata nem csak a *Books in Print* köteteiből származó rekordokat tartalmaz (a *Forthcoming Books* anyagával együtt), hanem a *Books Out of Print* rekordjait is. A Bowker a könyvek két kategóriájának címeit külön nyomtatott kötetekben jelenteti meg, különböző címek alatt. A könyvek státusza szerint elkülönített CD-ROM-okat is megjelentetnek. A legtöbb online és CD-ROM-kiadó csak a kapható és a jövőben megjelenő (in-print és forthcoming) státuszú könyvek rekordjait ajánlja fel. A DIALOG változat státusz indexének böngészése és keresése világossá teszi, hogy 1999 közepén csaknem 900000 rekord volt ebben az adatbázisban, amely igazán megérdemelné a BIP/BOP becenevet (5.1. ábra)

Ezen felül van még több mint 220000 rekord kritikákkal a könyvkritikák jól ismert és hosszú idő óta elismert forrásaiból (5.2. ábra). A CD-ROM-kiadványok között van egy még teljesebb változat, a *Books in Print with Book Reviews*; ezért a DIALOG változatnak ez a plusz tulajdonsága megérdemelné ennek említését is a címben, mint például „BIP/BOP with Reviews” (BIP/BOP könyvkritikákkal). Ez előnyös tulajdonság lehetne az Amazon.commal, a Barnes & Noble-lal és a Bordersszel folytatott kegyetlen versenyben. Furcsa módon, annak ellenére, hogy ez a két fontos és alapvető forrás is szerepel benne, a DIALOG változat olcsóbb, mint a BIP néhány másik változata, amelyben nincsenek kritikák és már nem kapható könyvek. (A *Books in Print*nek volt egy ingyenes változata, amelyet a Bowker minden regisztrált használó számára elérhetővé tett, de ez 1999 végén pénzes szolgáltatás lett).

Search History
Database Details

Set	Term Searched	Items	
S1	ST=ACTIVE RECORD	1402763	Display
S2	ST=OUT OF PRINT	890309	Display

Show Database Details for:
470: Books in Print® [Shardsheet](#) [Rates](#)

5.1. ábra: Már nem kapható könyvek rekordjai a Books in Print DIALOG változatában

Search History
Database Details

Set	Term Searched	Items	
S1	UD=?/REVIEW	223428	Display

Show Database Details for:
470: Books in Print® [Shardsheet](#) [Rates](#)

5.2. ábra: Rekordok könyvkritikával a DIALOG Books in Print adatbázisában

Az Ulrich's International Periodicals Directory és az EBSCO The Serials Directory (TSD) adatbázisai közötti, a földrajzi feltártság szempontjából megfigyelhető lényeges különbségen túl a feldolgozott időszaki kiadványok típusai is fontos megkülönböztető tényezők. Úgy tűnik, hogy az Ulrich's sokkal több házi kiadványt és belső orgánumot dolgoz fel, mint a TSD. Ez nem feltétlenül fontos a tipikus használó számára, de az lehet az időszaki kiadványokkal foglalkozó könyvtárosnak, aki az utolsó reménye lehet a keresőnek egy furcsa periodikum után való vadászatban. Az időszaki kiadvány státusza szűrési szempont a PubListben, amely az Ulrich's nagyon fontos és ingyenes részhalmaza. A PubList csak az aktív címeket tartalmazza. Ez a korlátozás nem túl komoly hátrány a legtöbb használó számára, de nem említik meg az alapos online súgó fájlban, amely csupán a 150000 időszaki kiadványra és 8000 újságra utal.

A PubList által feldolgozott időszaki kiadványok státuszára vonatkozó információ hiánya zavaró lehet olyankor, amikor egy folyóirat épp nem rég változtatta meg a címét (azaz kiadása megszűnt a korábbi cím alatt). Például a *Database* magazin címe *EContent* lett. Az a használó, aki nincs tudatában ennek a címváltozásnak, meg lehet lepve, hogy nem találja ezt a folyóiratot a cím szerinti kereséskor, mert a régi címet nem indexelik a címindexben. Míg a PubList keze valamennyire meg van kötve az Ulrich's miatt, ahhoz joguk van, hogy eldöntsék, hogyan indexelik az adatfájlt és milyen információkat szolgáltatnak a gyakran feltett kérdések (FAQ) fájljában. (Vannak más különbségek is a rekord tartalmában, ahogy azt a 6. fejezetben megtárgyaljuk.) Anélkül, hogy blaszfémiának hangzana, érdemes megemlíteni, hogy a Merriam Webster Biographical Dictionary-nek a Zane Publishing által megjelentetett, emlékeztetősen gyenge változatában nem utalnak a speciális „státusz” szűrőre. Csak már nem élő híres emberek szerepelnek benne. A zavart még növeli az, hogy sok híres ember, aki még élt a nyomtatott változat több mint egy évtizeddel ezelőtti megjelenésekor, időközben meg-

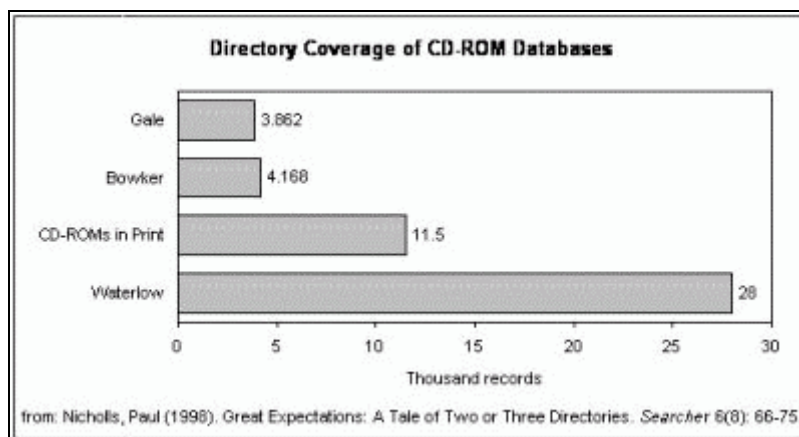
halt, így a használó számára valószínűleg kideríthetetlen marad, milyen „forrásokat” is dolgoz fel ez az adatbázis.

A cégek címjegyzékeinek adatbázisaiban az, hogy egy céget a tőzsdére vittek-e vagy sem, gyakran fontos feltétel, és ez egyes cégadatbázisok feltárt forrásait elfogadhatatlanná teheti. A cégek címjegyzékét tartalmazó adatbázisok másik lényeges különbségének az az oka, hogy a bekerüléshez a cégeknek meg kell felelniük a nettó forgalom vagy az alkalmazottak száma (vagy mindkettő) tekintetében egy minimális szintnek. A feltárt forrásoknak ezek a különbségei magyarázzák azt meg, hogy a Moody's 5000 Plus miért tartalmaz körülbelül 40%-kal kevesebb cégről rekordot, mint a Standard & Poor's, s kevesebb mint felét annak, ami a Disclosure-ban található (körülbelül 12000).

Újra csak azt lehet mondani, hogy olyan sok kereső számára, aki a legnagyobb cégeket keresi egy bizonyos iparágon belül, a feltárt források különbségei nem lényegesek, mivel a legfontosabb szereplőkről az összes adatbázis tartalmaz rekordokat. Másrészt azonban ezeknek az adatbázisoknak a rendkívüli szelektivitása a feltárt forrásokban nyilvánvalóvá válik, ha azt nézzük, hogy az Egyesült Államokban mennyi cég van (a becslések szerint 15 millió körül). Még a legnagyobb cég címjegyzék adatbázisok, az American Business Information Directory és a Dun's Market Identifiers is csak kevesebb mint 70%-ra vonatkozóan tartalmaz információkat (Lavin 1998).

A filmes címjegyzékek által feldolgozott források különbségei nehezzé teszik összehasonlításukat. A *Cinemanía*, amelynek a CD-ROM-változata 1997-ben, majd az online 1998-ban szűnt meg, csak nagyjátékfilmeket vett fel, míg a *Bowker's Complete Video Directory* oktatófilmek ezreire vonatkozó rekordokat is tartalmaz.

Az adatbázisok címjegyzékeinek összehasonlításakor Nicholls (1998) kimutatta, hogy sem a Gale Directory of Databases, sem a Bowker-Saur World Databases adatbázisa nem tárja fel megfelelő mértékben a CD-ROM adatbázisokat. Ez nyilvánvalóvá válik, amikor összehasonlítjuk a Gale saját CD-ROMs in Print adatbázisát és a Waterlow New Media Information Multimedia and CD-ROM Directoryját (5.3. ábra) A médium típusán kívül az adatbázis műfaja is szempont a Gale Directory of Databases és a World Databases címjegyzékekben feltárt forrásokban.

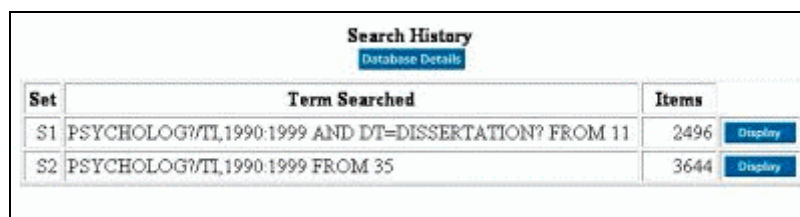


5.3. ábra: CD-ROM adatbázisok feltártsága négy címjegyzékben

Dokumentumokat feldolgozó adatbázisok

Viszonylag könnyű a dokumentumokat feldolgozó adatbázisok forrásainak körét meghatározni: a legtipikusabbak a folyóiratok, magazinok, újságok, monografikus művek, könyvfejezetek, konferencia előadások és a disszertációk. A szakterületek magfolyóiratainak feltárásán kívül nincsenek kötelező szabályok vagy egyetemes elvárások a feltárt forrásokkal kapcsolatban. Ezeket az egyedi használók és a használói közösség általános preferenciái határozzák meg.

Még ugyanannak a keresőnek a preferenciái is változók lehetnek a keresés céljától függően. Egy doktori disszertáció tervezetének készítésekor a jelölt valószínűleg jobban kedveli azokat az adatbázisokat, amelyekben a lehető legkiterjedtebb a feldolgozott dokumentumok köre, doktori disszertációkat és magiszteri szakdolgozatokat is tartalmaznak. Sok egyéb ok mellett ez zárna ki a pszichológia területén a Mental Health Abstracts használatát, amelyben nem találni disszertációkat. Ezzel ellentétben a PsycINFO jelentős mértékben dolgoz fel ilyen dokumentumokat, de ez még mindig nem jelent teljességet. (A *PsycLIT*, a CD-ROM-változat egyáltalán nem tartalmaz disszertációkra vonatkozó rekordokat.) Weston és Lauerdale (1988) azt állapította meg, hogy a PsycINFO nem vette fel a Dissertation Abstractsben megtalálható pszichológiai disszertációk egyharmadát. Egy 1999 közepén végzett egyszerű keresés megerősítette ezt az eredményt (5.4. ábra). Az egyszerűség és összehasonlíthatóság kedvéért a keresést a *psycholog* szótőre korlátoztuk a címezőben, mert nem lehetett tudni, hogy vajon csak azokat a disszertációkat veszi-e át a PsycINFO a Dissertation Abstractsből, amelyeket a pszichológia alá soroltak be, vagy a kapcsolódó területekről származókat is. A keresés az 1990 és 1999 közötti évekre vonatkozott.



Search History			
Database Details			
Set	Term Searched	Items	
S1	PSYCHOLOG?/TI,1990:1999 AND DT=DISSERTATION? FROM 11	2496	Display
S2	PSYCHOLOG?/TI,1990:1999 FROM 35	3644	Display

5.4. ábra: A PsycINFO és a Dissertation Abstracts által feldolgozott anyag összehasonlítása a címező végzett keresés alapján.

A Dissertation Abstracts kiterjedt jellege relatívnak bizonyult, amikor az OCLC WorldCat adatbázisával hasonlították össze. Perry és Salisbury (1995) vizsgálata szerint ez utóbbi összességében kétszer annyi disszertációt tartalmazott. Ez azzal a ténnyel van összefüggésben, hogy az OCLC adatbázisában rekordok tízezrei találhatók a tagkönyvtárakból alapidipomás szakdolgozatokról. Ugyanakkor azonban lehetséges, hogy nem találhatók meg olyan neves iskolákban készült magiszteri szakdolgozatok és doktori disszertációk, amelyeknek a könyvtárai nem OCLC tagok. Egy megbízható keresőnek figyelembe kell vennie ezeket a tényeket, amikor a számokat nézi.

Azt a pusztá állítást, hogy egy adatbázis feldolgoz disszertációkat, nem szabad minden további nélkül elfogadni. A LibLit, amely a legkövetkezetesebben dolgozza fel a forrásdokumentumokat és minden tekintetben teljesíti azt, amit ígér, nem igazolja azt az állítást, hogy feldolgoz könyvtár- és információtudományi disszertációkat. Csak néhány tucatnyi ilyen rekordjuk van. Az Information Science Abstracts ugyanazt állította, azonban ők sem valószínűsítették ezt meg következetesen. 1985-ből 47 ilyen rekord került be az adatbázisba; 1986-ból 475 és 1987-ből 245, ezt követően abbahagyták feltárásukat.

A hírlevelek fontosak lehetnek sok. kereső számára. A NewsNet kizárólag erre a dokumentumtípusra építő adatbázisa sikeres üzletnek bizonyult 1997-ig, amikor a webre alapozott hírlevél szolgáltatások arra készítették a Newsnetet, hogy vonuljon ki erről a piacról.

A konferencia-előadások alapvetőek azok számára, akik a természet- vagy társadalomtudományok legújabb fejleményeivel lépést akarnak tartani. A legtöbb tudományos folyóiratnak olyan lemaradásai vannak a kapott kéziratok közreadásában, hogy sok kutató számára a konferenciák számítanak a legjobb helynek az aktuális témákban elért eredményeik közzétételére. Egy egyszerű keresés alapján úgy tűnik, hogy a LISA és az ISA sokkal nagyobb mértékben - ha nem is következetesen - dolgoz fel konferenciaköteteket, mint a LibLit (5.6. ábra). Hangsúlyozni kell, hogy a LISA-ban és az ISA-ban nem lehet keresni a „conference proceedings” (konferenciakötet) vagy a „conference paper” (konferencia-előadás) dokumentumtípus szerint, míg a LibLitben a „speech” (beszéd) dokumentumtípus jelentése többértelmű lehet. Az MHA itt azok számára sem megfelelő, akik konferencia-előadásokról igényelnek információt, míg a PsycINFO szerény mértékben tárja fel ezt a dokumentumtípust.

Dialog Response		
DIALOG RANK Results (Alphabetized Results)		

RANK: S1/1-767 Field: PY= File(s): 202		
(Rank fields found in 767 records -- 3 unique terms) Page 1 of 1		
RANK No.	Items	Term
-----	-----	-----
3	47	1985
1	475	1986
2	245	1987
---end of results---		

5.5. ábra: Disszertációk rekordjai az ISA-ban

Search History				
Database Details				
Set	Term Searched	Items	File	
S1	SO=(PROCEEDINGS OR PAPERS OR MEETING OR CONFERENCE)	29158		Display
S1	SO=(PROCEEDINGS OR PAPERS OR MEETING OR CONFERENCE)	1246	438	Display
S1	SO=(PROCEEDINGS OR PAPERS OR MEETING OR CONFERENCE)	16496	202	Display
S1	SO=(PROCEEDINGS OR PAPERS OR MEETING OR CONFERENCE)	11416	61	Display
S2	DT=SPEECH	8660		Display
S2	DT=SPEECH	8660	438	Display
S2	DT=SPEECH	0	202	Display
S2	DT=SPEECH	0	61	Display
S3	S1 OR S2	37347		Display
S3	S1 OR S2	9435	438	Display
S3	S1 OR S2	16496	202	Display
S3	S1 OR S2	11416	61	Display

5.6. ábra: Konferencia-előadások az ISA, LibLit és LISA adatbázisokban

A kutatási jelentések fontosak lehet(né)nek sok kereséskor, de gyakran nehézkes hozzáférhetőségük kevésbé vonzóvá teheti őket. A National Technical Information Services (NTIS) épített olyan adatbázist, amit eredetileg a kutatási jelentésekre alapozott volt. Az ERIC adatbázisnak van egy olyan részhalmaza, a RIE (Resources in Education) gyűjtemény, amely csaknem egymillió kutatási jelentést, konferencia előadást és oktatási anyagot tartalmaz. A LISA az 1990-es évek elején adta hozzá az adatbázishoz kutatási részhalmozát, amely folyamatban levő kutatási projektekhez tartozó rekordokat tartalmaz, köztük kutatási jelentésekről szólókat.

A humán tudományok területének kutatói, egyetemi és főiskolai hallgatói és oktatói arról ismertek, hogy sokkal jobban kedvelik a könyveket, mint a természettudósok. A társadalomtudósok ebből a szempontból középen állnak, s természetesen az egyéni preferenciák is különbözőek lehetnek. A LibLit (és az H. W. Wilson adatbázis család) nagyon sok könyvkritikát tartalmaz, a LISA semmit, az ISA gyakorlatilag semmit. A könyvkritikákra (és az egyéb szemlecikkekre) vonatkozó rövid bibliográfiai hivatkozások a Wilson adatbázisokban kevésbé informatívak, mint amelyek olyan adatbázisokban találhatók, amelyek gyakran tartalmazznak arra vonatkozó jelzést, hogy a kritika pozitív, negatív vagy semleges. Ezt az információt gyakran nehéz megállapítani a kritika alapján, így nem mindig található meg az adatbázisok rekordjaiban, de a fájlkészítők dicséretet érdemelnek azért, hogy megpróbálják megadni. A „referátumok” azonban egy másik történetet jelentenek, ahogy azt a 11. fejezetben megtárgyaljuk.

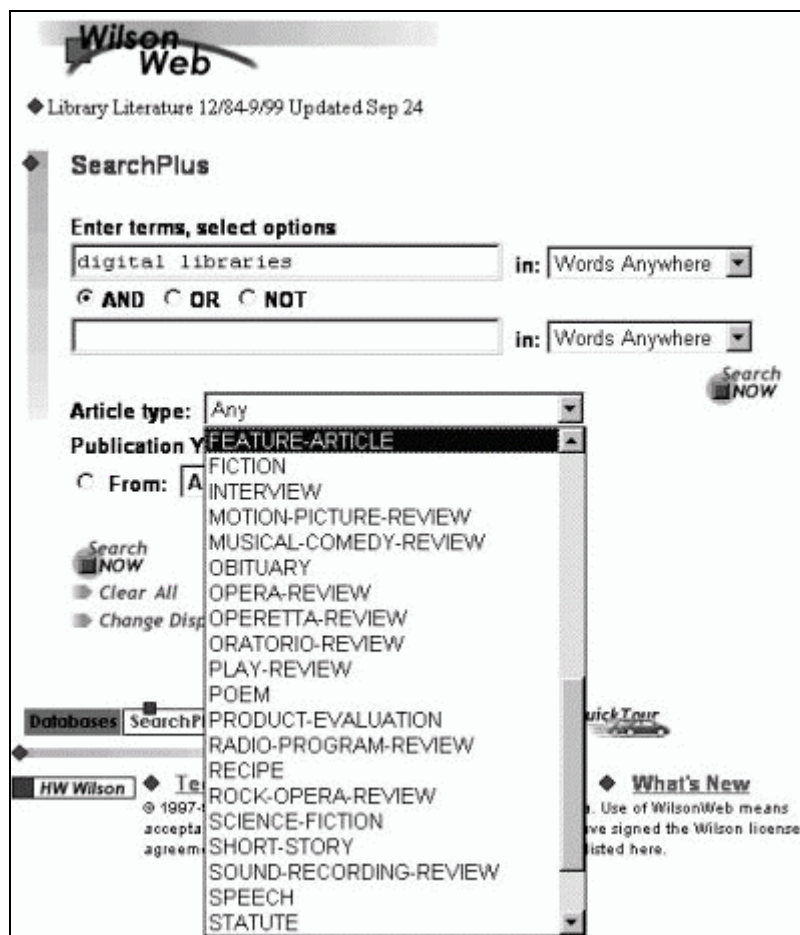
Szorosan kapcsolódik ide egy másik dokumentumtípus: a könyvek rekordjainak analitikus szintje, amelyeket egy könyv egyes fejezeteiről készítenek. Ez feltehetőleg az egyik legfontosabb új fejlemény mind a PsycINFO, mind a PsycLIT adatbázisban. Értéküket még növeli az, hogy a legtöbb online és CD-ROM-változatban a monografikus és az analitikus rekordokat hiperhivatkozások kapcsolják össze. Ez szoftver kérdése, de tipikus esete annak, amikor a szoftver a legtöbbet hozza ki az adatok tartalmából. Azért érdemel külön említést, mert az adatbázisok online és CD-ROM-változata nagyon gyakran teljes mértékben különbözik egymástól ebből a szempontból, meghatározott dokumentumtípusok felvételében vagy kihagyásában.

A Sociofile, a Sociological Abstracts CD-ROM-változata például 1997-ig nem tartalmazott könyvkritikákról rekordokat. A Biological Abstracts CD-ROM-változata sem a könyvkritikákról, sem a konferenciakötetéről nem tartalmaz rekordokat, a konferencia-előadásokról nem is beszélve. Az online változat sok használója számára pontosan ezeknek a információforrásoknak a felvétele jelenti a lényeges különbséget, s hogy ennek eredményeképp több egyedi rekordot kapnak, amikor több adatbázisban végeznek keresést (Snow 1998).

A szerkesztőségi cikkekre vonatkozó rekordok jelenléte egy adatbázisban nem feltétlenül számít előnynek. Ez a szerkesztőségi cikk típusától függ. Tudományos folyóiratokban és a nagyközönségnek szóló, legjobb minőségű periodikumokban, mint a *Newsweek*, *Time* és a *U.S. News & World Report*, a szerkesztőségi cikkek gyakran fontosabbak, mint egyes cikkek. A szaklapokban a szerkesztőségi cikkek általában csak a cikkek mini bevezetői, nem sokban különböznek azoktól az udvarias bevezetőktől, amelyekben a konferenciákon a levezető elnökök bemutatják a következő előadót. Vannak persze kivételek. Például a *Database*-ben (most *EContent*) és az *Online*-ban a szerkesztőségi cikkek tipikusan saját jogú mini-cikkek.

Az adatbázisok minőségének hiányára vonatkozó legkíméletlenebb kritikák szerkesztőségi cikként vagy a vendégszerkesztő bevezetőjeként jelennek meg. Bizonyos mértékig ugyanez érvényes a szerkesztőhöz írt levelekre is (amelyeket a legtöbb adatbázis nem vesz fel). Ezek a korábban megjelent cikkekhez kapcsolódóan fontos (azokat támogató vagy vitató) információkat tartalmazhatnak, s ezért azokhoz lehetne és kellene kapcsolni őket. Sok közülük természetesen csak általános dicséreteket tartalmaznak a szerkesztőkről vagy a szerzőkről, valódi tartalom nélkül. Ezek megfelelnek a levegőbe dobott csókoknak, s nem érdemlik meg, hogy külön rekordot kapjanak egy adatbázisban. Az H. W. Wilson gyakorlata különösen jó ennek a dilemmának megoldásához: a szerkesztőkhöz írt leveleket akkor veszik fel, ha jelentős személyektől származnak. Az ISA-ban nem szerepel ilyen dokumentumtípus, s a gordiuszi csomót úgy oldják meg, hogy az „article” (cikk) dokumentumtípust rendelik hozzá a szerkesztőhöz írt levelek rekordjaihoz.

A dokumentumtípusok indexének böngészése nem mindig tárja fel ezeknek a dokumentumtípusoknak a jelenlétét. Lehetséges, hogy egy általános dokumentumtípus, az „article” (cikk) alá sorolják be őket. Más adatbázisokban van egy külön index, amelyet gyakran „article type”-nak (cikktípus) neveznek, amely lehetővé teszi a megkülönböztetést. (Ez igaz az összes H. W. Wilson és a legtöbb Gale/IAC adatbázisra). A LISA adatbázis úgy kerüli el az esetleges problémákat, hogy egyszerűen nem tartalmaz dokumentumtípusra vagy cikktípusra vonatkozó információkat (kivéve a kutatási jelentéseket). Néha a dokumentumtípusok elnevezései furesák, vagy a kódolás, vagy a terminológia miatt, amely a szoftver problémája lehet. A LibLit esetében az H. W. Wilson saját változatában (5.7.a ábra), s ugyanúgy a SilverPlatter (5.7.b ábra) és az Ovid (5.7.c ábra) változatban a cikkek típusát kiírják, s nem valamilyen rejtélyes kódot használnak rájuk. Ez sokkal jobb megoldás, mint amit a LibLit DOS-os változatában használnak, valamint az ERIC adatbázis jelenlegi DIALOG-os változatában, amely még mindig numerikus kódokat használ a dokumentumtípusokra.



5. 7a. ábra: A LibLit H. W. Wilson változata

Publication Type Index Display

Perform Search

Back in Index Main Search Page Forward in Index

Choose from among the following index entries:

Select	Publication Type	# of Citations
<input type="checkbox"/>	analytic.pt.	13599
<input type="checkbox"/>	article.pt.	141926
<input type="checkbox"/>	autobiography.pt.	74
<input type="checkbox"/>	bibliography.pt.	8063
<input type="checkbox"/>	biography.pt.	1534
<input type="checkbox"/>	book excerpt.pt.	353
<input type="checkbox"/>	book review.pt.	32234
<input type="checkbox"/>	exhibition.pt.	817
<input type="checkbox"/>	feature article.pt.	117007
<input type="checkbox"/>	fiction.pt.	4
<input type="checkbox"/>	interview.pt.	1699
<input type="checkbox"/>	monograph.pt.	7881
<input type="checkbox"/>	obituary.pt.	1674
<input type="checkbox"/>	product review.pt.	1937
<input type="checkbox"/>	speech.pt.	8480
<input type="checkbox"/>	symposium.pt.	260

5. 7b. ábra: A LibLit SilverPlatter változata

WebSPIRS - Netscape

Limit Your Search

Set Limits Cancel

Logout

HELP

Databases

Searches

Suggest

Index

Thesaurus

Database Information

Show/Hide Records

About WebSPIRS

PY-Publication Year: ☒ Any Year ☐ 1997 to 1999

☐ From To

PD-Physical Description: (any value) BETWEEN BIBL CHART CHARTS

To select multiple, hold down the shift, ctrl, or command key and click.

LA-Language Of Work: (any value) AFRIKAANS CHINESE DANISH DUTCH

To select multiple, hold down the shift, ctrl, or command key and click.

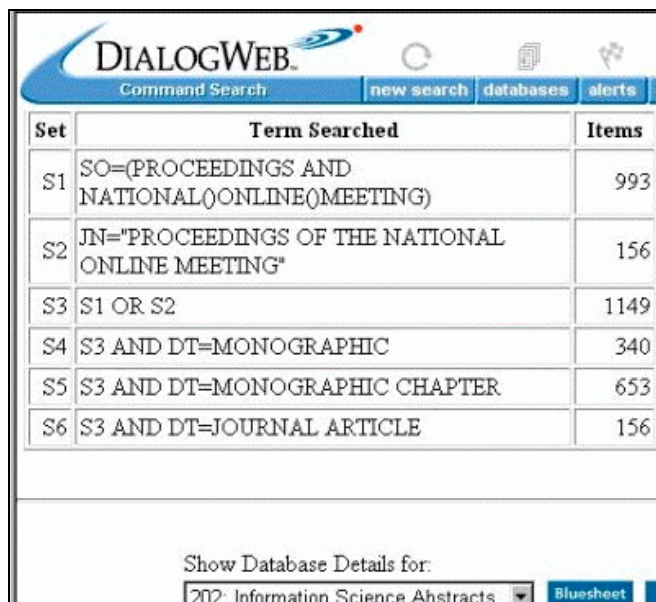
DT-Document Type: ARTICLE AUTOBIOGRAPHY BIBLIOGRAPHY BIOGRAPHY BOOK

Document: Done

5. 7c. ábra: A LibLit Ovid változata

Ezen kívül lehetséges az, hogy a fájl készítője nem alkalmazta következetesen a dokumentumtípusok kategóriáit. Ez a helyzet az ISA esetében, ahol a konferencia-előadásokat „monographic” (monografikus mű) és „monographic chapter” (monografikus mű fejezete, könyvfejezet) formában azonosítják. Ez nem csak grammatikai szempontból alkalmatlan, hanem a

dokumentumtípus indexet is megbízhatatlanná teszi a keresés szempontjából, mivel mindkét fogalmat használják az egyedi konferencia-előadásokra (ahogy az az 5.8. ábrán szereplő keresési eredményekből látható). Egy kötetre vonatkozó rekord lehet dokumentumtípusként „monograph” vagy egy előadás „monographic chapter”, de ha nevén nevezzük a gyereket, s egy konferencia-előadást konferencia-előadásnak nevezünk, az segíthet elkerülni a zavart. Ahogy korábban említettük, az ISA dokumentumtípus sagájának legutolsó csavarja az volt, hogy 1999-ben és 2000-ben az összes, az adatbázisba újonnan felvett rekordot a dokumentumtípus mezőben „article” (cikk) formában határoztak meg, közte a konferencia-előadásokat is. További gond még a használó számára, hogy ezeket mind a folyóiratcím (JN, Journal Name), mind a forrás (SO, Source) mezőben keresni kell. Az elsőt kifejezésként indexelik, a másodikat szavanként. Ebből következően nem lehet a keresést egyetlen paranccsal elvégezni.



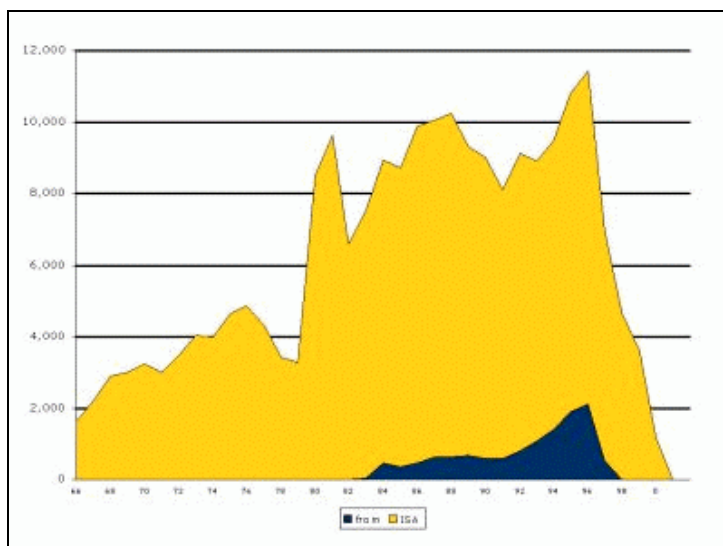
Set	Term Searched	Items
S1	SO=(PROCEEDINGS AND NATIONAL()ONLINE()MEETING)	993
S2	JN="PROCEEDINGS OF THE NATIONAL ONLINE MEETING"	156
S3	S1 OR S2	1149
S4	S3 AND DT=MONOGRAPHIC	340
S5	S3 AND DT=MONOGRAPHIC CHAPTER	653
S6	S3 AND DT=JOURNAL ARTICLE	156

Show Database Details for:
202: Information Science Abstracts Bluesheet

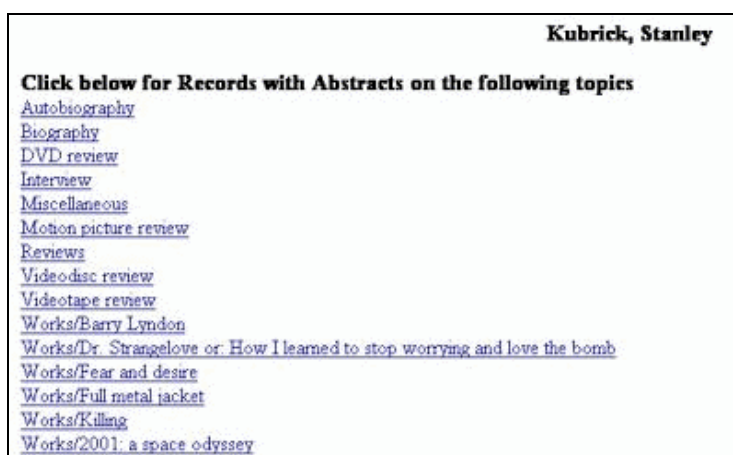
5.8. ábra: A konferencia-előadásokat következtetlenül azonosítják az ISA adatbázisban.

A használók szerethetik vagy nem szerethetik meghatározott dokumentumtípusok felvételét vagy kizárását. Addig, amíg a gyakorlat egyértelmű és következetes, tudják, mit kapnak. Bár a szabadalmi rekordok felvétele az ISA adatbázisba önmagában nem volt feltétlenül jó vagy rossz döntés, mértékük éves növekedése gyanús volt, különösen azért, mert az adatbázishoz évente hozzáadott rekordok teljes száma elég állandó maradt (5.9. ábra). Az a tény, hogy a fájl készítője, az IFI/Plenum ezeket a rekordokat saját szabadalmi adatbázisából importálta, nem erősítette az indokot. Miután Allcock (1997) és Jacsó (1997c) a *Database* hasábjain folytatott levélváltása során ezt megtárgyalta, több szabadalmi rekord nem került be az adatbázisba.

Az H. W. Wilson cég adatbázisai szokatlanul nagy arányban tartalmazzák különböző típusú kritikák rekordjait. Ez különösen hasznos forrás lehetne, ha a kritikák rekordjai tartalmaznának referátumot, de erre soha nem kerül sor. A nagyon specifikus tárgyszavak és altárgyszavak kombinációi segíthetnek ezen bizonyos mértékig, de más szempontból akadályt is jelenthetnek. Ez történik a Biographies Plus Illustrated adatbázis esetében. Ezt vonzó módon összekapcsolják az H. W. Wilson cég referáló/indexelő és teljes szövegű adatbázisaival. A használatnak azonban minden egyes kritikátípusra rá kell kattintania csak azért, hogy eljussanak a rekordokhoz (5.10. ábra). A használókat valószínűleg jobban érdekli magának egy adott műnek a kritikája, mint azok, amelyek egy meghatározott médiumon találhatóak meg (lézerlemez, DVD, videokazetta), s a rekordokat össze lehetne hozni a műnek a címe alatt, ahelyett, hogy szétaprózzák őket.



5.9. ábra: A szabadalmi rekordok száma az ISA adatbázisban



5.10. ábra: Túl sok kritikátípus található a Biographies Plus Illustrated adatbázisban

Folyóiratbázis

A folyóiratok különleges kezelést érdemelnek, amikor az adatbázisok által feltárt forrásokat értékeljük. A folyóiratok képezik az indexelő, referáló, teljes szövegű és képi adatbázisok forrásainak leghagyományosabb és legalapvetőbb típusát. A folyóiratok fontosságát tekintve nem csoda, hogy feltártságuk kapja a legnagyobb figyelmet a dokumentumokat feldolgozó adatbázisok értékelése során. A folyóiratbázis gyakran az első szempont adatbázisok összehasonlításakor. A feldolgozott folyóiratok száma állandó összetevője még a legrövidebb adatbázis leírásoknak is.

A pusztán szám önmagában nem elegendő az értékeléshez. A feltártság mértéke lehet széles, de sekély, a cikkek feltárása lehet túlzottan válogató, és néhány alapvető folyóirat hiánya (alapvető a potenciális használói közösség vagy annak hangadó képviselői szerint) kizárhat egy vizsgált adatbázist abból, hogy komolyan számításba vegyék. Másrészt azonban egy szemmel láthatóan csekély folyóiratbázissal rendelkező adatbázis épp megfeleltethet sok kutató számára. Az Ovid Mental Health Collection című, teljes szövegű, adatbázisa a példa erre. Ebben csupán 10 periodikus kiadvány szerepel. Nyolc olyan van köztük, amelyet az Institute

for Scientific Information figyel. A nyolc közül hatot impakt faktora alapján az 1., 2., 4., 6., 7. és 8. helyre rangsorolt 80 pszichiátriai folyóirat közül a Social Science Journal Citations Reports adatbázis. Hét folyóirat tartozik a kombinált pszichológiai és pszichiátriai kategória 489 folyóiratot tartalmazó listáján az első 20 közé. Ugyanez a 7 cím a teljes társadalomtudományi kategória 1672 folyóiratot tartalmazó listáján az első 36 között szerepel (5.11. ábra). Ez a rendkívüli folyóiratbázis a legjobb példája annak, hogy a kicsi a szép.

Címek a Mental Health Collection-ben	Rangsor az összes		
	Társada- lomtudo- mányi	Pszicho- lógiai és pszichi- átriai	Pszí- chi- átriai
	folyóirat között		
American Journal of Psychiatry	5	4	2
Annual Review of Psychology	13	7	NA
Archives of General Psychiatry	1	1	1
British Journal of Psychiatry	28	14	6
Current Opinion in Psychiatry	841	315	62
Journal of Clinical Psychopharmacology			
Journal of the American Academy of Child and Adolescent Psychiatry	16	9	4
Psychological Medicine	36	20	8
Psychosomatic Medicine	32	17	7
Year Book of Psychiatry and Applied Mental Health			
A folyóiratok teljes száma az adott kategóriában	1672	489	80

Titles in Mental Health Collection	All SocSci rank	All Psyc. rank	All Psychiatry rank
American Journal of Psychiatry	5	4	2
Annual Review of Psychology	13	7	n/a
Archives of General Psychiatry	1	1	1
British Journal of Psychiatry	28	14	6
Current Opinion in Psychiatry	841	315	62
Journal of Clinical Psychopharmacology			
Journal of the American Academy of Child and Adolescent Psychiatry	16	9	4
Psychological Medicine	36	20	8
Psychosomatic Medicine	32	17	7
Year Book of Psychiatry and Applied Mental Health			
Total number of journals in this category	1672	489	80

**5.11. ábra: Az Ovid Mental Health Collection
adatbázisában szereplő folyóiratok helyezése impakt
faktoruk szerint**

Előfordulhat, hogy a döntéshozók vagy a potenciális használók csak azért szeretnek jobban egy adatbázist, mint egy másikat, mert egy vagy két kedvenc folyóiratukat feldolgozzák. Ilyen esetekben jó ötlet azt ajánlani a fájl előállítóinak, hogy egészítsék ki az adatbázist néhány, a könyvtár által kedvelt címmel, mivel az ilyen kéréseket könnyű teljesíteni. Egy Hernon és Metoyer-Duran által végzett felmérés (1992), amelyet öt felsőoktatási és szakkönyvtár 43 könyvtárossával végeztek, azt mutatta ki, hogy az H. W. Wilson cég LibLit adatbázisa nem

tartalmazott néhány olyan folyóiratot, amely fontos volt ezeknek a könyvtáraknak a számára. Tudva azt, hogy az H. W. Wilson cég milyen figyelmet szentel ügyfeleinek, s milyen szisztematikusan értékeli folyóiratbázisát, ezt a tapasztalt hiányosságot a fájl előállítói biztosan pótolták volna, ha az igényelt címeket meghatározták volna a könyvtárosok.

Bár különböző diszciplínákban nem létezik a magfolyóiratok olyan formális listája, amely kielégítené minden nyilvános, iskolai, szak- és felsőoktatási könyvtár könyvtárosait és dolgozóit, vagy akár teljes használói közönségét, számos hasznos forrás van, amely közvetetten felhasználható annak megítélésére, hogy egy adott adatbázis folyóiratbázisa mennyire megfelelő. Ez sokkal könnyebb egy szakkönyvtár esetében, amely egy adott diszciplínához tartozó adatbázist keres, mint egy nagy egyetemi könyvtár esetében, amelynek nagyon összetett használói közönségének kiszolgálása érdekében szüksége van mind szakterületi, mind multidiszciplináris adatbázisokra.

Mindazonáltal a könyvtár folyóirat-állománya jó kiindulópontként szolgálhat. Ez azon a feltételezésen alapul, hogy a könyvtárnak jól meghatározott gyűjteményfejlesztési szabályzata van, a kiválasztás és a selejtezés szempontjaival, amelyek a folyóiratok előfizetésére is érvényesek. A pénzügyi korlátok biztosan megakadályoznak egy könyvtárat abban, hogy minden kíváncsi periodikumra előfizessen, de az adatbázisok ideális együttese várhatóan feldolgozza a könyvtár által előfizetett folyóiratok többségét. További címeket is örömmel fogadhatnak, különösen cikkek teljes szövegű vagy lehetőleg képi változatából, olyan folyóiratokból, amelyek az adatbázisban megtalálhatók. A teljes szöveg rendelkezésre bocsátása plusz díjakat is jelenthet, ahogy azt a 12., az árakkal kapcsolatos megfontolásokat tartalmazó fejezetben megtárgyaljuk.

A hagyományos gyűjteményfejlesztési útmutatók, mint *Katz Guide to Magazines* című műve vagy a folyóiratok Brandon-Hill listája orvosi könyvtárak számára, szintén kiváló források adatbázisok folyóiratbázisának értékeléséhez. Az utóbbi ingyenesen elérhető a <http://www.nlm.nih.gov/psr/outreach/branchill.html> címen. (2004-ben abbahagyták (erről tájékoztatás a <http://nlm.gov/libinfo/ejournals/branchill.html> címen olvasható), folytatása a Doody's Core Titles int he Health Sciences (<http://www.doody.com/dct/>). A *ProQuest Medical Library* CD-ROM-os változatát az teszi kiváló forrássá sok fejlődő országban, hogy a Brandon-Hill listán szereplő címek többségét tartalmazza.

Az Institute for Scientific Information *Science Citation Index*, *Social Sciences Citation Index*, *Arts & Humanities Citation Index* és a Journal Citation Reports (JCR) adatbázisok számára feldolgozott folyóiratainak halmaza és részhalmazai jelentik sok használó számára ténylegesen azon folyóiratok irányadó körét, amelyet tartalmaznia kell egy felsőoktatási kutatásra szánt adatbázisnak. Még akkor is, ha egy adott diszciplínára vonatkozóan a halmaz nem is kezelhető némi fenntartások nélkül, ez nagyon jó kiindulópont egy adatbázis folyóiratbázisának értékeléséhez. Például a könyvtár- és információtudományi folyóiratok választéka elég korlátozott a JCR-ban (65 cím körül mozog), de a folyóiratok körülbelül 90%-át valószínűleg széles körben elfogadják a könyvtár- és információtudomány alapvető forrásaként. A lista bővíthető olyan címekkel, amelyek nem szerepelnek a JCR-ban, de amelyeket széles körben használnak és nagyra értékelnek a szakemberek szakterületükön, ilyen például a *Cataloging & Classification Quarterly*, *Computers in Libraries* és a *Searcher* a könyvtár- és információtudományi (Library and Information Science, LIS) kategóriában.

Az, hogy egy adatbázis nem tartalmaz néhány olyan címet, amely szerepel az ISI által feldolgozott folyóiratok között, nem feltétlenül rossz jel. Az ISI-nál szerepel néhány olyan folyóirat, amely sokkal inkább a politikai korrektség, mint a tudományos érdem miatt kapott privilegizált helyet. Az orosz *Nauchno Tekhnicheskaja Informatsiya* esetében úgy tűnhetett,

hogy van némi befolyása a szakmára az ISI mérőszámai alapján, de ennek cikkeire az orosz és kelet-európai kutatók nem annyira tudományos relevanciájuk miatt hivatkoztak, hanem azért, hogy megfeleljenek az elvárásoknak és „szolgálják a hazát”. A *Zeitschrift für Bibliotheks-wesen und Bibliographie* (ZBB) releváns lehet néhány német nyelvű ország tudósai számára, de szerepeltetése biztosan nem alapvető fontosságú a könyvtár- és információtudományi adatbázisok legtöbb használója számára. Nem véletlen, hogy az 1997-ben kapott 17 hivatkozás közül 15 magából a ZBB-ből származik. Az, hogy még mindig szerepelnek az ISI listáján, nem igazán a folyóiratok érdeme. Valószínűleg ottmaradtak, mint a vonatra várók a restiben. Az ISI listáján való megjelenésük oka valószínűleg sokkal inkább tehetetlenség, mint érdem. Az ISI hitelét erősíti, hogy 1996 óta az orosz sorozat két folyóirata közül egyik sincs már a könyvtár- és információtudományi folyóiratok között. Feltételezem, hogy hamarosan a ZBB-t is le fogják venni a listáról.

Az MHA adatbázis súlyos hiányossága, hogy a Journal Citation Reports pszichológiai és pszichiátriai részének 20 legmagasabb impakt faktorú folyóirata közül egyik sem szerepel benne. Az impakt faktor méri az arányt a folyóirat által az előző két évben kapott hivatkozások száma és a folyóiratban ugyanabban az időszakban megjelent olyan publikációk száma között, amelyekre lehetett (volna) hivatkozni. Annak a 20 vezető pszichiátriai és pszichológiai folyóiratnak a teljes hiánya, amelyek abszolút számban a legtöbb hivatkozást kapták (5.13. ábra), az MHA hiányossága bénító. Nem meglepő, hogy a 20 legproduktívabb pszichiátriai és pszichológiai folyóirat közül (5.14. ábra) egyik sem szerepel az MHA adatbázis által 1997-ben feldolgozott 227 folyóirat csoportjában (5.15. ábra). (Technikai okok miatt csak az első 16 folyóirat mutatható meg egy képernyőképen a következő ábrák mindegyikén. Az MHA-ból származó képernyőkép [5.15. ábra] csak az első 13, 1997-ből legtöbb tételt adó folyóiratot mutatja.)

Rank	Journal Abbreviation	ISSN	Total Cites	Impact Factor	Index	Articles	Cited Half-Life
1	ARCH GEN PSYCHIAT	0003-908X	23230	10.751	1.730	104	9.4
2	BEHAV BRAIN SCI	0140-525X	2642	8.118	0.667	6	8.7
3	PSYCHOL REV	0033-290X	9095	7.060	0.969	32	> 10.0
4	AM J PSYCHIAT	0002-953X	22478	6.991	0.943	283	7.3
5	PSYCHOL BULL	0033-2909	11572	6.038	1.286	34	> 10.0
6	J COGNITIVE NEUROSCI	0898-929X	1580	4.844	0.927	95	5.9
7	ANNU REV PSYCHOL	0066-4308	2087	4.841	1.409	22	7.7
8	J CLIN PSYCHIAT	0160-6689	6784	4.093	0.351	171	6.9
9	J AM ACAD CHILD PSY	0890-8567	6525	3.793	0.556	287	5.8
10	LEARN MEMORY	1072-0502	317	3.673	0.119	42	2.6
11	COGNITIVE PSYCHOL	0010-0285	2357	3.516	0.480	20	> 10.0
12	SCHIZOPHRENIA BULL	0586-7614	3755	3.589	1.380	50	7.1
13	J MEM LANG	0749-596X	1844	3.358	0.456	57	6.4
14	BRIT J PSYCHIAT	0007-1250	13851	3.265	0.744	195	8.3
15	CONTEMP PSYCHOL	0010-7549	107	3.200			4.9
16	DEV PSYCHOPATHOL	0954-5794	1153	3.192	0.419	43	4.3

5.12. ábra: A legmagasabb impakt faktorról rendelkező pszichiátriai és pszichológiai folyóiratok 1997-ben.

Journal Citation Reports - 1997 Social Sciences Edition

File Edit View Calculators Options Help

Journal Rankings

Mark	Rank	Journal Abbreviation	ISSN	1997 Total Cites	Impact Factor	Immediacy Index	1997 Articles	Cited Half-Life
	1	ARCH GEN PSYCHIAT	0003-998X	22238	10.751	1.790	164	9.4
	2	AM J PSYCHIAT	0002-957X	22478	6.581	0.943	283	7.3
	3	J PERS SOC PSYCHOL	0022-3514	17864	2.986	0.580	216	> 10.0
	4	BRIT J PSYCHIAT	0007-1250	13851	3.265	0.744	195	8.3
	5	PSYCHOL BULL	0033-2909	11572	6.038	1.206	34	> 10.0
	6	CHILD DEV	0009-3920	9372	2.067	0.365	85	> 10.0
	7	PSYCHOL REV	0033-299X	9085	7.060	0.969	32	> 10.0
	8	PSYCHOL BEHAV	0031-9384	8062	1.250	0.167	360	8.6
	9	AM PSYCHOL	0003-065X	7688	3.076	2.118	76	8.4
	10	J CLIN PSYCHIAT	0160-6689	6784	4.083	0.351	171	6.9
	11	J AM ACAD CHILD PSY	0890-8567	6525	3.793	0.556	287	5.8
	12	ACTA PSYCHIAT SCAND	0001-698X	6245	1.588	0.136	169	8.5
	13	J ABNORM PSYCHOL	0021-942X	5882	2.678	0.290	69	> 10.0
	14	PSYCHOL MED	0033-2917	5740	3.017	0.550	131	8.5
	15	DEV PSYCHOL	0012-1649	5532	2.263	0.283	90	9.3
	16	J APPL PSYCHOL	0021-9010	5371	1.815	0.283	79	> 10.0

Sort: Total Cites Filter: Multiple Categories 422 Journals 14 marked

5.13. ábra: A legtöbb hivatkozással rendelkező pszichiátriai és pszichológiai folyóiratok 1997-ben.

Journal Citation Reports - 1997 Social Sciences Edition

File Edit View Calculators Options Help

Journal Rankings

Mark	Rank	Journal Abbreviation	ISSN	1997 Total Cites	Impact Factor	Immediacy Index	1997 Articles	Cited Half-Life
	1	PERCEPT MOTOR SKILL	0031-5125	2286	8.272	0.060	486	> 10.0
	2	PSYCHOL REP	0033-2941	3411	0.243	0.078	473	> 10.0
	3	PSYCHOL BEHAV	0031-9384	8062	1.250	0.167	360	8.6
	4	AM J PSYCHIAT	0002-957X	22478	6.581	0.943	283	7.3
	5	PERS INDIV DIFFER	0191-8869	2382	0.786	0.113	231	6.4
	6	J PERS SOC PSYCHOL	0022-3514	17864	2.986	0.580	216	> 10.0
	7	J AM ACAD CHILD PSY	0890-8567	6525	3.793	0.556	287	5.8
	8	BRIT J PSYCHIAT	0007-1250	13851	3.265	0.744	195	8.3
	9	J CLIN PSYCHIAT	0160-6689	6784	4.083	0.351	171	6.9
	10	ACTA PSYCHIAT SCAND	0001-698X	6245	1.588	0.136	169	8.5
	11	NEUROPSYCHOLOGIA	0028-3932	5294	2.267	0.397	156	8.8
	12	INT J GERIATR PSYCH	0885-6230	1114	1.097	0.187	150	4.6
	13	SCHIZOPHR RES	0920-9964	1734	2.737	0.261	138	4.9
	14	PSYCHOL MED	0033-2917	5740	3.017	0.550	131	8.5
	15	WORKFORCE	0031-5745	4		0.016	125	
	16	PSYCHIAT RES	0165-1781	3187	1.327	0.099	121	7.8

Sort: Total Articles Filter: Multiple Categories 422 Journals 14 marked

5.14. ábra: A legproduktívabb pszichiátriai és pszichológiai folyóiratok 1997-ben.

DIALOGWEB

new search databases alerts cost logoff help

JOURNALS WITH THE HIGHEST POSTINGS IN 1997 FROM MHA

RANK: S1/1-3418 Field: JN= File(s): 86
(Rank fields found in 3238 records -- 211 unique terms)

RANK No.	Items	Term
1	137	INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY
2	76	JOURNAL OF CLINICAL PSYCHOLOGY
3	53	JOURNAL OF PERSONALITY ASSESSMENT
4	49	JOURNAL OF TRAUMATIC STRESS
5	45	AMERICAN JOURNAL ON MENTAL RETARDATION
6	44	AMERICAN JOURNAL OF ORTHOPSYCHIATRY
7	43	COGNITIVE THERAPY AND RESEARCH (NEW YORK)
8	42	JOURNAL OF ABNORMAL CHILD PSYCHOLOGY (NEW YORK)
9	40	HUMAN PSYCHOPHARMACOLOGY: CLINICAL AND EXPERIM
10	40	JOURNAL OF CLINICAL CHILD PSYCHOLOGY
11	39	ADOLESCENCE
12	38	SEX ROLES (NEW YORK)
13	37	INTERNATIONAL PSYCHOGERIATRICS

5.15. ábra: Az MHA által 1997-ben feldolgozott folyóiratok listája (részlet)

Egy adott diszciplína folyóiratainak nem hivatalos, de informatív presztízs listái szintén kiindulópontul szolgálhatnak adatbázisok folyóiratbázisának értékeléséhez. A könyvtár- és információtudomány területén elég sok presztízs lista létezik. Esteibar és Lancaster (1992) könyvtár- és információtudományi folyóiratok informatív rangsorát készítette el, az Illinois-i Egyetem Urbana-Champaignben található Könyvtár- és Információtudományi iskola (Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign) 131 kötelező irodalom jegyzékét, 41 doktori disszertációját és 13 oktató 114 publikációját vizsgálva.

A könyvtár- és információtudományi folyóiratokkal kapcsolatos, Kohl és Davis által végzett klasszikus vizsgálat (1985) most már valamennyire elavult, és nem tartalmazza a növekvő mértékben fontos információtechnológiai folyóiratokat. Továbbra is jól használható listát nyújt az elsődlegesen könyvtártudományi címekről, amelyeket az ALA által akkreditált könyvtárosképzők dékánjai és a Kutatókönyvtárak Egyesülete (Association of Research Libraries, ARL) 43 könyvtárának igazgatója a vezető folyóiratok közé sorolt. A Tjoumas és Blake által végzett felmérés (1992) ugyancsak eredményezett egy presztízs listát, amely a könyvtárosképzők iskolai és nyilvános könyvtári témákra specializálódott munkatársainak a szakmai folyóiratokra vonatkozó véleményén alapult. Jacsó (1998a) ezen listák alapján kombinált listát készített, hogy összehasonlítsa őket hat könyvtár- és információtudományi és -technológiai anyagot feldolgozó adatbázis folyóiratbázisával (5.16. ábra).

Az információtechnológiára specializálódott oktatók és gyakorló információs szakemberek körében elvégzett hasonló felmérés adhatná meg a szükséges egyensúlyt ezekkel a listákkal, amelyek a könyvtár- és információtudomány hagyományos oldalát reprezentálják.

Ojala (1992) jó példát adott azzal, hogy olyan módon vizsgált üzleti és gazdasági adatbázisokat, hogy a különböző egyetemi tanszékek oktatóinak presztízs listáin szereplő folyóiratok esetében a feltártság mértékét is meghatározta (nem csupán jelenlétüket). Jaguszewski és Kempf (1995) négy témafigyelő szolgáltatás folyóiratbázisát hasonlította össze. A teszteléshez 44 kémiai és 50 matematikai folyóiratot választottak ki. Bár a kémia területén a feltártság jó és összehasonlítható volt a négy adatbázis között, a matematikában két adatbázisban nagyon gyenge volt a feltártság, egyé jó, a CARL Uncover-é pedig nagyon jó (5.07. ábra)

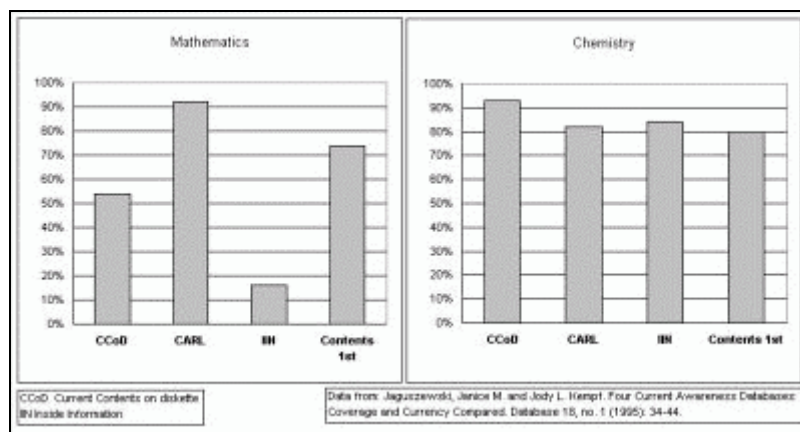
A rövidítések feloldása: CCoD - Current Contents on Diskette, IIN - Inside Information

Az adatok a következő forrásból származnak:

Jaguszewski, Janice M. és Jody L. Kempf: Current Awareness Databases: Coverage and Currency Compared. Database 18, no.1. (1995): 34-44. p.

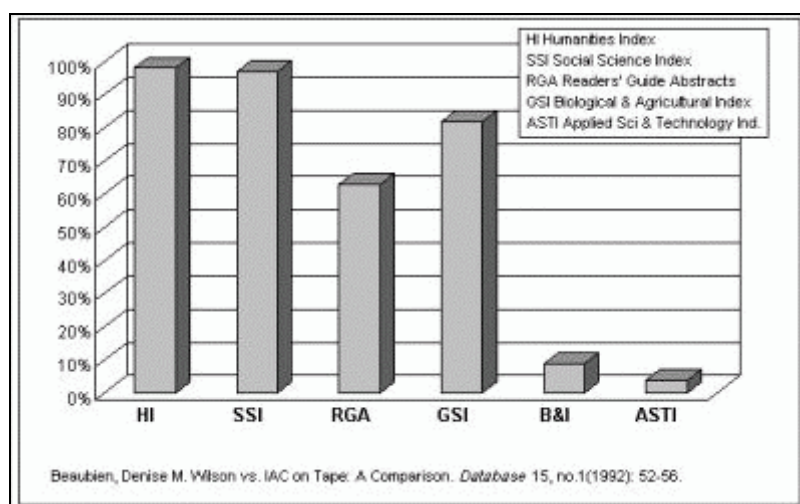
TABLE 1 Sample Journals with Postings for Databases in the Peer Group						
<i>Abbreviated Journal Titles</i>	<i>INSPEC</i>	<i>SSCI</i>	<i>LISA</i>	<i>Pascal</i>	<i>LibLit</i>	<i>ISA</i>
ANNU REV INFORM SCI		284	44	196		252
ASLIB PROC	654	1,282	938	1,005	493	698
B MED LIBR ASSOC	355	2,126	973	887	922	1,171
CAN J INFORM LIB SCI	16	94	70		82	32
<i>CAN LIBR J</i>	111	2,446	602	503	1,004	475
COLL RES LIBR	24	2,753	871	696	961	985
DATABASE	724	1,993	968	556	1,158	840
<i>EDUC INFORM</i>	51	622	280	77	411	92
ELECTRON LIBR	462	909	467	376	608	429
GOV INFORM Q	155	833	316		444	262
<i>GOV PUBL REV</i>	52	1,594	442	571	222	
<i>IFLA J-INT FED LIBR</i>		928	532	377	724	189
INFORM PROCESS MANAG	863	1,748	924	802	1,042	987
INFORM TECHNOL LIBR	400	1,022	527	431	809	467
<i>INT FORUM INFORM DOC</i>	192	550	413	424	284	184
INT J INFORM MANAGE	264	626	280	204		162
<i>INTERLEND DOC SUPPLY</i>	95	364	267	260	298	73
J ACAD LIBR		2,324	972	3	1,286	516
J AM SOC INFORM SCI	1,208	2,092	1,217	1,107	904	1,544
J DOC	260	1,737	485	366	793	903
J EDUC LIBR INF SCI	60	539	294	265	517	220
J INFORM SCI	503	1,007	685	677	536	617
<i>LAW LIBR J</i>	25	1,580	479		511	63
<i>LIBR ACQUIS PRACT TH</i>	112	1,235	773	553	914	459
LIBR INFORM SCI RES	51	530	301	216	337	302
<i>LIBR J</i>		27,108	1,407	1,383	7,346	1,015
<i>LIBR QUART</i>	26	2,487	412	265	934	397
<i>LIBR RESOUR TECH SER</i>	243	1,327	471	385	830	793
<i>LIBR TRENDS</i>	175	1,066	835	872	540	907
<i>LIBRI</i>	28	504	460	496	289	352
<i>NACHR DOK</i>	566	1,636	521	927	341	500
<i>NAUCH-TEKHN INFORM 1</i>	902	2,315	245		1,109	1,032
ONLINE	1,240	2,207	1,152	777	1,232	998
<i>P ASIS ANNU MEET</i>	36	2,342	106	120		1,527
<i>PROGRAM-AUTOM LIBR</i>	559	1,004	529	500	599	176
RQ	48	4,005	685		1,284	604
<i>SCHOLARLY PUBL</i>		743	210		330	159
SCIENTOMETRICS	45	1,124	689	689		644
<i>SERIALS LIBR</i>	374	1,159	827	376	730	691
SPEC LIBR		1,657	801	842	576	1,788
<i>WILSON LIBR BULL</i>		7,717	1,136		3,973	409
<i>Z BIBL BIBL</i>		1,270	210	626		23
Total for 42 sample titles	10,879	90,889	24,636	18,665	35,722	24,159
Total without LJ and WLB	10,879	56,064	22,093	17,282	24,403	22,735
Category A titles (bold)	7,332	30,892	13,479	10,095	13,984	14,421
<i>Category B titles (italics)</i>	1,197	49,773	8,656	5,713	18,777	6,064
Category C titles	2,350	10,224	2,501	2,857	2,961	3,674

5.16. ábra: Nagy presztízsű könyvtár- és
információtudományi folyóiratok feldolgozottsága
hat adatbázisban



5.17. ábra: 94 kémiai és matematikai folyóirat feldolgozottsága négy témafigyelő szolgáltatásban.

Beaubien (1992) az IAC Expanded Academic Index és az H. W. Wilson adatbázisai egy csoportja folyóiratbázisát hasonlította össze. Míg a folyóiratok szintjén nagyon nagy volt az átfedés a humán tudományok, társadalomtudományok és általános tudományok területén, jóval kisebb mértékű volt a nagyközönségnek szóló területeken (amelyeket a Readers' Guide Abstracts adatbázis reprezentál). Az alkalmazott tudományok és a technika, biológia és mezőgazdaság területén különösen alacsony volt az átfedés (4 és 9% között), ahogy azt az 5.18. ábra mutatja.



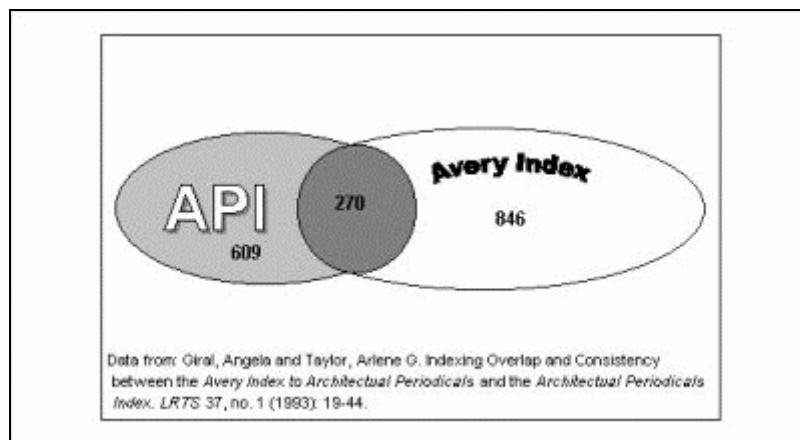
5.18. ábra: Folyóiratok közötti átfedés az IAC Expanded Academic Indexben és az H. W. Wilson cég adatbázisaiban

Az adatbázisok által feldolgozott folyóiratok összevetése egy könyvtár állományával döntést meghatározó tényező lehet egy előfizetendő adatbázis kiválasztásakor. Fry és Parsons (1994) a UMI, IAC és H. W. Wilson cégek különböző adatbázisainak folyóiratbázisát hasonlította össze az alapidomás kutatás számára a főbb tantervi kategóriák alapján, s arra az eredményre jutottak, hogy a Wilson adatbázis-család folyóiratai egyeztek meg legjobban az Iowa State Libraries állományában található címekkel.

Azt lehetne várni, hogy a specifikusabb területek adatbázisai által feldolgozott folyóiratok között nagyobb mértékű lesz az átfedés. Valójában nem ez a helyzet, ahogy azt Giral és Taylor (1993) kimutatta. Azt tapasztalták, hogy csupán 30%-os (270 folyóirat) volt az átfedés az Avery Index (846 folyóirat) és az Architectural Periodicals Index (609 folyóirat) között az

1979 és 1986 közötti időszakban (5.19. ábra). Bebizonyosodott, hogy gyakorlati okokból még szűkebb időtartamot érdemes vizsgálni összehasonlításra.

Az adatok a következő forrásból származnak:



5.19. ábra: Átfedések az Avery Index és az Architectural Periodicals Index által feldolgozott folyóiratok között

Még kivételesen széles folyóirat-bázisú adatbázisok esetében sincs garancia arra, hogy egy adott könyvtár olvasói számára releváns folyóiratok elérhetőek-e, nem is beszélve arról, hogy megfelelő mértékben vannak-e feltárva. Holt és Schmidt egy tanulmányában (1995) ugyanazt a stratégiát alkalmazó mintakeresésekkel azt találták, hogy csupán 30% átfedés volt a Faxon Finder és az Uncover2 folyóiratai között, pedig mindkettő több mint 10000 folyóiratot dolgoz fel.

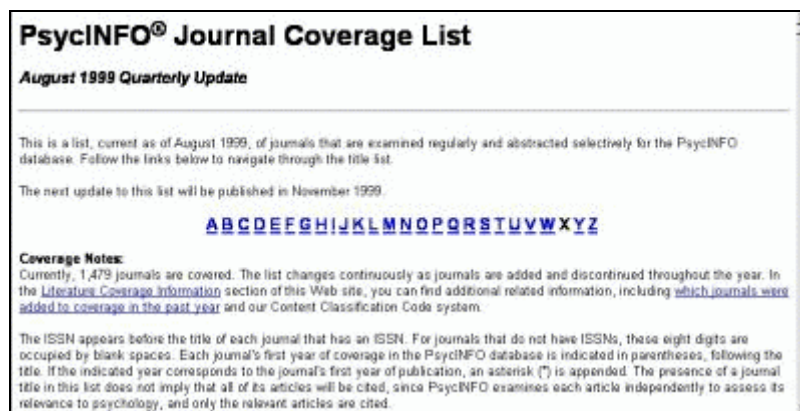
Az adatbázisok folyóiratainak összehasonlítása fárasztó művelet lehet, ahogy folyóiratcímek százait vagy akár ezreit kell összehasonlítani olyan listáról, amelyet tipikus esetben a fájlok készítői rendelkezésre bocsátanak. Nagyobb adatbázisoknál a lista terjedelmes nyomtatott kötetet tesz ki, ahogy ezt a BIOSIS, EMBASE és INSPEC esetében látjuk. A PubMed a MEDLINE folyóirat-listáját ingyen böngészhetővé és kereshetővé tette a weben a <http://www.ncbi.nlm.nih.gov:80/entrez/jrbrowser.cgi> címen. Bár HTML formátumban sokkal több információt tesznek hozzáférhetővé, mint más adatbázis-készítők egyszerű folyóiratlistái tartalmazznak, sok használó érdeke lenne, hogy arra ösztönözzék a fájlok előállítóit, hogy kereshető adatbázis változatban tegyék elérhetővé a weben folyóirataik listáját.

Egyre több fájlkészítő teszi közzé folyóiratlistáját a weben. A legjobbak a folyóiratlistákat számos feltétel kombinálásával teszik kereshetővé, s olyan táblázatos formájú outputot kínálnak, amelyben az oszlopokat a használó határozhatja meg. A Bell & Howell szolgált erre a legjobb példával a <http://www.umi.com/hp/Support/Titles/> címen, ahogy azt később illusztráljuk ebben a fejezetben. (A megadott cím eredeti funkciójában már nem működik, csak a Bell & Howell utódjaként létrehozott Proquest cég különböző részeire utal, de ProQuest a <http://www.proquest.com/tls/jsp/list/tlsSearch.jsp> címen pótolja ezt.)

A különböző rövidítési és leírási szabályok miatt a címek megtalálása bonyolultabb egy adatbázis nyomtatott változatában, mint ahogy az első pillantásra látszik. Az online változatok sem csodaszerek azonban hosszú HTML oldalak görgetésének nehézsége miatt. A betürendbe sorolt folyóiratlistán rengeteg belépési pontot kellene ellenőrizni ahhoz, hogy biztosak lehessünk benne, hogy nem mulasztunk el egyet sem egy viszonylag egyszerű cím, a *Journal of the American Oil Chemists' Society* különböző variációs lehetőségei közül. Ahogy azt a 8. fejezetben a következetességgel kapcsolatban megtárgyaljuk, egyedül a PASCAL

adatbázisban variációk tucatjai vannak erre a folyóiratcímre, amelyek közül sok nincs is közel egymáshoz, mint például az a változat, amely semmi más, mint a rövidítés: JAOCS.

Számos fájlkészítő kezdte ingyenesen elérhetővé tenni folyóiratlistáját a weben. A PAIS ennek egyik úttörőjeként elismerést érdemel. Folyóiratlistájuk nem adatbázis, de azon túl, hogy böngészhető, kereshető is - bár csak korlátozott mértékben - a böngésző Find parancsát használva. Az American Psychological Association szintén közzétette folyóiratainak listáját a weben (5.20. ábra). Ez nem csak a folyóirat címét és ISSN számát (ahol van ilyen) tartalmazza, de azt is, hogy melyik évtől dolgozzák fel a lapot (5.21. ábra).



5.20. ábra: Részlet a PsycINFO folyóiratlistájának bevezetőjéből.



5.21. ábra: A PsycINFO folyóiratlistája

Egyes online szolgáltatók úgy segítenek a használóknak, hogy kereshető folyóiratlistát tesznek hozzáférhetővé, még azok számára is, akik nem előfizetőik, vagy pedig hajlandóak arra, hogy összevegyék feldolgozott anyagukat egy potenciálisan jövedelmező üzlet érdekében, amikor szerződést kötnek egy nagy egyetemmel. Nem kell a Bell & Howell Learning and Information előfizetőjének lenni ahhoz, hogy megtudjuk, milyen folyóiratokat dolgoznak fel a cég különböző adatbázisai. Választani lehet, hogy melyik adatelemek (ISSN, UMI kód, a formátum elérhetősége, a feldolgozottság dátumai különböző formátumokban) szerepeljenek a listán, amit a keresés során generálnak sima szöveggént, HTML-ben vagy vesszőkkel körülhatárolt formátumban (5.22. ábra), amely az Excelbe importálható.

ProQuest databases: ABI/INFORM Research										
The list is current as of August 23, 1999										
Title [1994]	Cit First [1994]	Cit Latest	Abs First [1994]	Abs Latest	FT First [1991]	FT Latest	Ing First [1991]	Ing Latest	T-G First [1991]	T-G Latest
ABA Book Companion	31854	current	31854	current	10554	current	31195	current	11556	current
ASMA Computing Services	13190	current	13190	current	31566	current	13190	current	31566	current
ASMA Periodic Management Quarterly	11779	current	11779	current	11651	current	11885	current	41566	current
ASMA Systems Management	18179	current	18179	current	18167	current	18167	current	18167	current
ASCI Journal of Management	9180	31487	9180	31487						
ASTAH Business Quarterly	31880	41881	31880	41881						
AT & T Technology	11887	11886	11887	11886	11887	11886	11888	11886	11886	11886
AZB, Arizona Business [Selective Business Coverage]	18171	current	18171	current	11651	current	12167	current	41566	current
Abacus [Full Text/Range Available via Pdf]	12174	current	12174	current						
Administrative	3181	10168	3181	10168						
Academy of Management Executive, The	2187	current	2187	current	2161	current	2187	current	11165	current
Academy of Management Journal	9171	current	9171	current	3162	current	3187	current	10165	current
Academy of Management: The Academy of Management Review	1176	current	1176	current	1162	current	1187	current	1166	current
Academy of Management: The Academy of Management Review	4179	current	4179	current	6166	current	1196	current	6166	current
Academy	11172	current	11172	current	1162	current	1192	current	5166	current
Accounting Digest, The	6173	6173	6173	6173						
Accounting Horizons	12167	current	12167	current	12167	current	12167	current	3166	current
Accounting Review, The	18171	current	18171	current			1188	current		
Accounting Technology	11164	current	11164	current	1162	current	9196	current	8166	current
Accounting and Business Research	1182	current	1182	current	1187	current	1187	current	1187	current
Accounting and Finance [Full Text/Range Available via Pdf]	11164	current	11164	current						
Accounting, Organizations and Society	1180	current	1180	current						
Action de Road	3171	current	3171	current	22165	current	1187	current	11165	current
Ad Forum	12162	3183	12162	3183						
Adweek	9175	10169	9175	10169						
Adweek Jan	9181	current	9181	current	1162	current	9196	current	8166	current

5.22. ábra: A UMI folyóiratlistájának egyik formátuma

Természetesen, ha előfizetünk egy szolgáltatásra vagy a kipróbálási időszak során több választási lehetőség is van. Az H. W. Wilson cég az egyes adatbázisokkal kapcsolatos információ részeként folyóiratlistát tesz közzé. Gyakran használok ezt arra, hogy időszaki kiadványokkal kapcsolatos információkat ellenőrizzek.

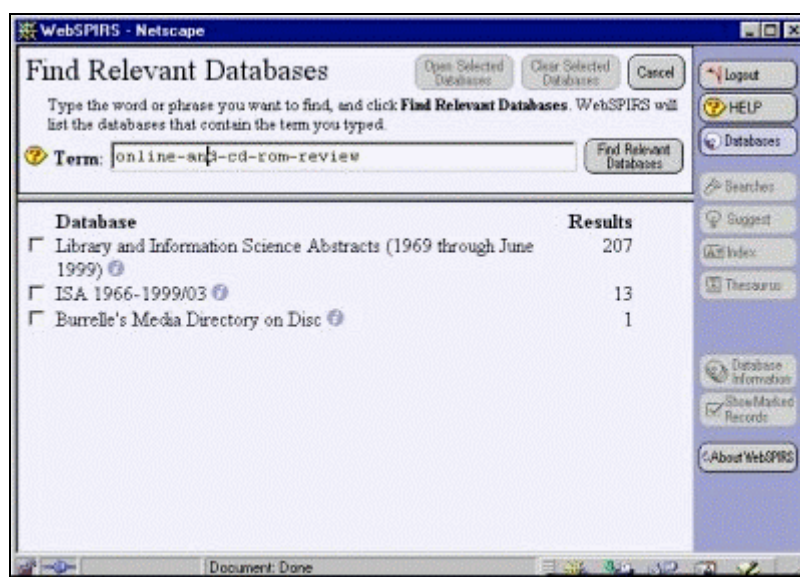
Library Literature & Information Science Journal List				
Updated: September 28, 1999				
Journal	ISSN	Peer Review	Index Start/End	Journal Note
AB Bookman's Weekly	001-0340		101894	
African Journal of Library, Archives & Information Science	0793-0771	B	111893	
Against the Odds	1045-2894	B	111893	
ALA Washington News	1525-6803		111896	Formerly ALA Washington Newsletter; name changed with January 26, 1996
ALA Washington Newsletter	001-1746		121894 121895	Name changed to ALA Washington News with January 26, 1996
Alabama Libraries	0012-4295		801890 831899	Ceased publication with Volume 46, No. 3, 1998
ALCTS Newsletter	1047-0402		111896 121896	Formerly RTSD Newsletter; name changed with Vol. 1, No. 1, 1998. Ceased publication in print form with Vol. 9, No. 6, 1999. Available in electronic format with 1999
ALCTS Newsletter (Online)	1047-0402		801899	Ceased ALCTS Newsletter in electronic format with 1999
Alexandria	0015-7408		111893	
Adis	0736-4073		831893	
The American Archivist	0360-9881	B	101894	
American Libraries	0012-0768		111894	

5.23. ábra: Az H. W. Wilson cég egy adatbázishoz kapcsolódó folyóiratlistája

Ezenkívül a legtöbb adatbázisban keresni is lehet a folyóirat címe szerint. A folyóiratcímek mezőjét azonban csaknem mindig kifejezésként indexelik, ez épp olyan nehézé teszi egy folyóirat megtalálását, mint egy nyomtatott listáról. Néhány speciálisan kereshető, később tárgyalandó folyóirat-adatbázist szavanként indexelnek, ez sokkal rugalmasabb keresést tesz lehetővé.

Az H. W. Wilson cég által készített minden CD-ROM adatbázisban (azokban azonban nem, amelyeket a SilverPlatter készít az H. W. Wilson adatbázisokból) van egy kevésbé ismert aranybánya: a folyóiratcímek adatbázisa. Ez az adatbázis lehetővé teszi a keresést bármely szóra vagy szavak kombinációjára (teljes vagy rövidített formában) a folyóiratcímek mezőjében. A folyóirat-adatbázis tartalmaz minden folyóiratot, amelyet az H. W. Wilson cég referál és indexel, nem csak azokat, amelyek a kezünkben tartott CD-ROM adatbázisban szerepelnek.

A WebSPIRS 4. 0 egy hasznos eszközt kínál az adatbázis-választáshoz. Ennek az a célja, hogy lehetővé tegye a kereső számára, hogy beírja tárgyszavak kombinációját annak érdekében, hogy megtalálja a tárgy szempontjából legígéretesebb adatbázisokat, de ez ugyanúgy használható arra is, hogy a keresést a folyóiratcímek között végezzük. Ez rendkívül kényelmes olyankor, amikor két vagy több egymással vetélkedő, számításba vett adatbázis érhető el a SilverPlatteren keresztül, mint például a LISA és az ISA. Sajnos tudni kell azt, hogy a folyóirat címe pontosan milyen formában jelenik meg az adatbázisban. Elég nehéz megjósolni, hogy egy olyan cím, mint a *Journal of the American Society for Information Science*, milyen formában fog megjelenni az adatbázisban (s hányféle változatban). Nem könnyíti meg a feladatot az sem, hogy a SilverPlatter a folyóiratcímekre a saját furcsa, kötőjeles változatát használja. Kevés használó találná ki pontosan, hogy az *Online & CD-ROM Review* ONLINE_AND_CD_ROM_REVIEW formában jelenik meg az adatbázisok SilverPlatter változatában (5.24. ábra).



5.24. ábra: Folyóiratok keresése a WebSPIRS adatbázisokban

A DIALOG Journal Name Finder (JNF) adatbázisa régóta elérhető, s messze a leghatékonyabb eszköz az adatbázisok által feldolgozott folyóiratok felderítésére. Ennek ez ereje két tényben rejlik. Az egyik az, hogy ez a DIALOG összes bibliográfiai adatbázisát tartalmazza, a második pedig az, hogy szavanként és kifejezésként is indexelik, JN= prefixszel és anélkül is. Ezek a változatos indexek nagy rugalmasságot ajánlanak még a legfurcsábban rövidített folyóiratcímek esetében is.

Van azonban egy probléma néhány fájlkészítő illogikus gyakorlata miatt. A Journal Name Finder adatbázist a rekordok JN vagy JO címkéjéből hozzák létre (5.25. ábra), így olyan esetekben, amikor a folyóiratcím csak SO (source, forrás) mezőben jelenik meg, a JNF eredményei megbízhatatlanok, mert az SO mezőt nem indexeli a JNF adatbázis. Ez akkor is probléma, ha közvetlenül az adatbázisban keresünk. Például az MHA és az ISA adatbázisban rekordok tízezreit vitték be az 1980-as évek előtt, amikor a folyóirat címe csak a forrás mezőben jelenik meg, a folyóiratcím mezőben nem. Egy folyóiratra vonatkozó teljes keresés esetében mindkét mezőben keresni kell. Növeli a zavart az, hogy két különböző stratégiát kell alkalmazni, mivel az adatbázison belül a JN mezőt csak kifejezésként indexelik, az SO mezőt pedig szavanként.

Az MHA és az ISA adatbázisnak ezt a súlyos fogyatékoságát már rég ki kellett volna javítania az IFI/Plumnak, amely az MHA kiadója (és az ISA-é volt 1998-ig), egy okos konverziós eszközt használva. Az adatbázisokban való dupla keresés, mind a JN=, mind a SO= prefixet használva extra erőfeszítést követel meg, s valószínűleg ismeretlen marad sok használó számára, akik azt feltételezhetik, hogy bármilyen eredményt kapnak a JN indexben keresve, az reprezentálja az adott folyóirat teljes feldolgozottságát az adott adatbázisban. Ez a gyakorlat az egyébként kiváló Journal Name Finder adatbázist is gyengíti.

Search History			
Database Details			
Set	Term Searched	Items	
S1	JN=JOURNAL OF SOCIAL PSYCHOLOGY	395	Display
S2	SO=(JOURNAL (1W) SOCIAL (W) PSYCHOLOGY)	2767	Display

5.25. ábra: Keresés a Mental Health Abstracts JN és SO indexeiben

1998-ban egy rövid időre a JNF adatbázis ingyenesen volt használható azok számára, akik DIALOG jelszóval rendelkeztek, de most komoly díjat kell fizetni érte. Mindig nagyon rendszeresen, havonta aktualizálták, a hónap első hétvégéjén. Az utóbbi időben jelentős késéseket lehetett tapasztalni. Ez speciális adatbázis, s a DIALOG-nak olyan kurrensen kellene tartania, amennyire csak lehetséges. Nagy érték a legmegfelelőbb adatbázis kiválasztásához, nem csak alkalmi kereséshez, hanem licenceléshez is, mivel lehetővé teszi a licencelni tervezett adatbázisok által feltárt folyóiratok megvizsgálását.

Földrajzi feltártság

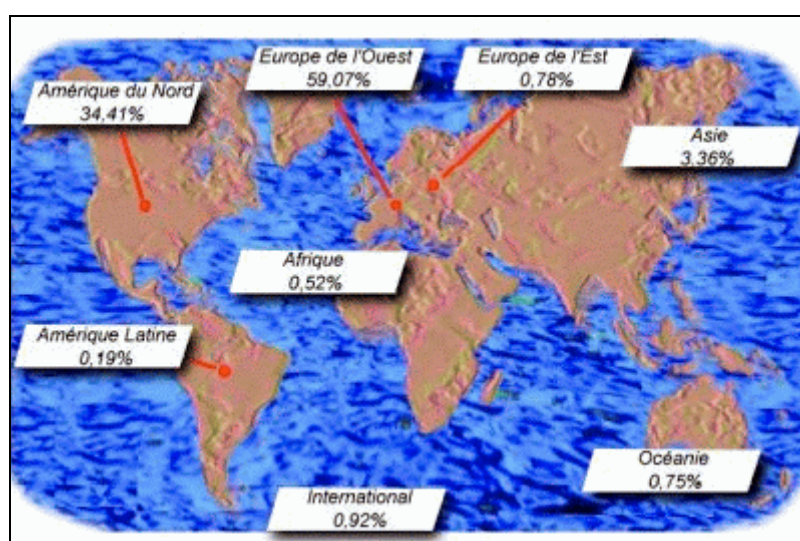
Egy adatbázisban feltárt anyagot sokszor írják le a „nemzetközi” jelzővel. A *nemzetközi* politikailag korrekt fogalom, de ahogy az a helyzet sok más politikailag korrekt fogalom esetében, ez is alapos vizsgálatot igényel. A DIALOG 426 adatbázisa közül, amelynek van leírása az adatbázis leírás (bluesheet) adatbázisban, 422-nek van Geographic Coverage (GC mezője). Nem kevesebb mint 349 adatbázis állítja azt, hogy nemzetközi a feltárása, 61 mondja azt, hogy csak az USA anyagát tárja fel, 12 pedig más országét (5.26. ábra). Ezeket az állításokat nem lehet egy az egyben elfogadni. Az még nem tesz egy adatbázist igazán nemzetközivé, hogy az USA-beli folyóiratok mellett egy kanadait is felvesz. Másrészt például az H. W. Wilson Education Abstracts adatbázis egy GC=other (egyéb) értéket ad meg, bár feldolgozott anyaga valóban nemzetközi, s igazán indokolt lenne, hogy a nemzetközi jelzőt használja a földrajzi feldolgozottságra

A nemzetközi feldolgozottság feltételezi olyan dokumentumok jelenlétét, amelyek 1, több országból származnak, vagy 2, több országra fókuszálnak a bibliográfiai adatbázisokban. Az útmutató adatbázisokban a *nemzetközi* szó arra utal, hogy a feltárt szervezetek, termékek vagy tárgyak az egész világból származnak. Sok „nemzetközi” adatbázis azonban nem felel meg ezeknek a követelményeknek. Csak azért, mert egy cégjegyzék tartalmaz USA-beli és kanadai vállalatokat, még nem igazán minősíthető nemzetközinek, bár ez szemantikusan igaz. Ugyanez érvényes arra az adatbázisra, amely tartalmaz féltucat brit periodikumot több száz amerikai mellett. Alapvető fontosságú lenne, hogy bizonyos statisztikai adatok legyenek a dokumentációban vagy a súgó fájlban a folyóiratok megjelenési helyéről (ahogy ez megtörtént a PASCAL adatbázisnak a fájl készítője által nyújtott változatban), ez az 5.27. ábrán látható. Ha ez az információ nem áll rendelkezésre, az adatbázis értékelőjének kell elvégeznie az erre vonatkozó kereséseket.

Search History			
Database Details			
Set	Term Searched	Items	
S1	UD=?	426	Display
S2	GC=?	422	Display
S3	GC=INTERNATIONAL	349	Display
S4	GC=US ONLY	61	Display
S5	GC=OTHER	12	Display

Show Database Details for:
415: DIALOG Bluesheets(TM)
[Bluesheet](#)
[Rates](#)

5.26. ábra: A DIALOG adatbázisainak nemzetközi feltártságára vonatkozó állításai



5.27. ábra: A PASCAL adatbázis folyóiratainak földrajzi megoszlása

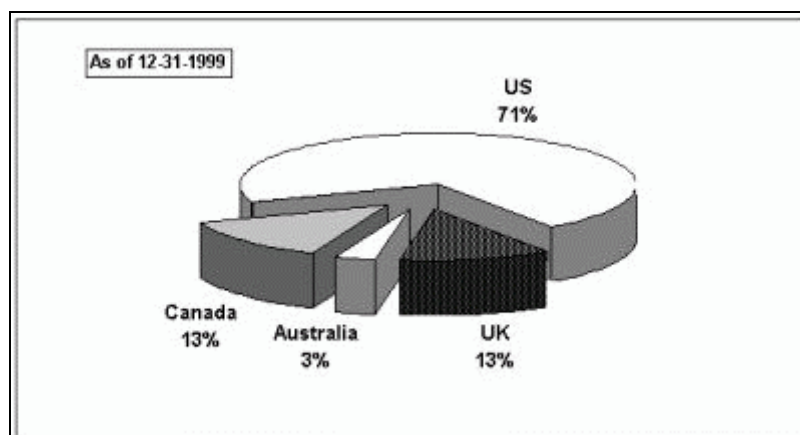
A fájlok készítőit nem lehet hibáztatni amiatt az egyoldalúság miatt, hogy elsősorban az Egyesült Államokból, Nagy-Britanniából, Kanadából és Ausztráliából származó folyóiratokat tárnak fel. Ezek azok az országok, amelyek a legérdekesebbek a legtöbb használó számára fejlett kutatásaik, üzleti kapcsolataik, kortárs művészetük, szórakoztató iparuk és kurrens eseményeik miatt. Azt is tudni kell, hogy a szakmai információs szolgáltatások és adatbázisok használóinak többsége is ezekből az országokból származik. Ezen kívül még más országokban is az angol lett lingua franca a tudományos, műszaki és üzleti kommunikációban (és a szórakoztatásban). A fájlok előállítóit nem köti az arányos földrajzi reprezentáció alapelve, ahogy az például az ENSZ és más nemzetközi szervezetek és bizottságok esetében érvényes, s ami munkájukat nem teszi mindig túl hatékonyá a nemes alapelvek miatt.

Elég sokatmondó az, hogy a korábban említett négy ország jelenteti meg az Ulrich's adatbázisban szereplő folyóiratok több mint 50%-át (5.28. ábra), amelynek az alcíme, helyesen - Nemzetközi Periodikum Címjegyzék (International Periodicals Directory).

Search History			
Database Details			
Set	Term Searched	Items	
S1	CN="UNITED STATES"	85891	Display
S2	CN="UNITED KINGDOM"	24184	Display
S3	CN="CANADA"	10197	Display
S4	CN="AUSTRALIA"	6566	Display
S5	S1.S4	126838	Display
S6	UD=?	230843	Display

5.28. ábra: Az USA-beli, brit, kanadai és ausztrál folyóiratok dominanciája az Ulrich's-ban

Az EBSCO Serials Directory (időszaki kiadványok címjegyzéke) hasonló megoszlást mutat (bár vannak jelentős különbségek a két címjegyzék között az egyes országok anyagának feltártságában). Az EBSCO - kisebb mérete ellenére - nagyobb számban veszi fel az USA, Kanada, Kína és Dél-Korea periodikumait, míg az Ulrich's vezet minden más ország és régió tekintetében.



5.29. ábra: Az időszaki kiadványok földrajzi megoszlása az EBSCO The Serials Directory adatbázisában

Ezeknek a számoknak az összehasonlításakor gondolni kell arra, hogy az Ulrich's esetében az újabb cím rekordja magában foglalja a korábbi címre vonatkozó információkat, míg az EBSCO megtartja a megszűnt címek rekordját. Ez tipikus gyakorlat az egyes országok időszaki kiadványokat feltáró katalógusaiban, köztük a Kongresszusi Könyvtár CONSER adatbázisában. Az OCLC közös katalógusa (Union Catalog) ugyanezt a gyakorlatot követi. Ezek a források különösen érdekesek, mert adott könyvtárak időszaki kiadvány gyűjteményeinek a földrajzi megoszlását reprezentálják, nem csupán az időszaki kiadványoknak az egész világra vonatkozó kiadási adatait.

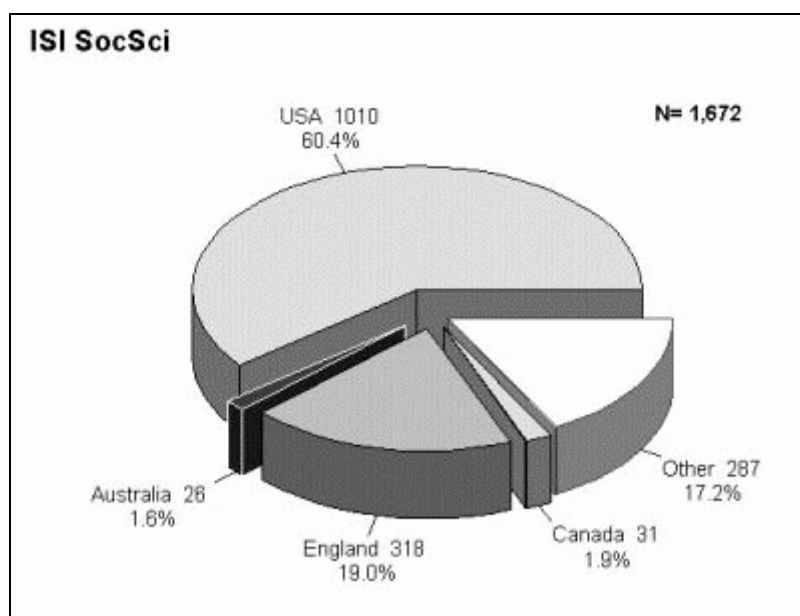
Az időszaki kiadványok nemzetközi nyilvántartó rendszere, az International Serials Data System (ISDS) által készített sokkal nagyobb (900000) és sokkal nemzetközibb címjegyzék statisztikai adatai megerősítik a többi időszaki kiadvány címjegyzékben tapasztalt megoszlási mintákat (5.30. ábra). Az ISDS számára csaknem száz nemzeti könyvtár jelenti az adatokat (amelyek nemzeti és néha regionális ISDS központokként működnek).

Poland	4,869	5,799	6,811	8,008	9,084	10,028	10,574	12,614
Portugal	1,675	1,938	2,207	2,396	2,621	3,069	3,242	3,447
Romania	937	1,239	2,043	2,363	2,363	2,365	3,996	5,410
Saudi Arabia	82	83	85	268	376	491	634	626
Senegal	90	89	91	91	91	94	93	91
Singapore	2,014	2,217	2,366	2,424	2,515	2,680	2,713	2,712
Slovakia	na	na	552	551	551	552	552	844
Slovenia	na	na	701	856	1,383	1,604	1,707	1,832
Spain	7,801	9,247	10,129	11,060	11,895	13,114	14,365	14,969
Sri Lanka	136	140	142	152	255	320	402	477
Switzerland	2,779	2,989	3,023	3,199	4,051	5,237	6,411	7,319
Sweden	12,033	13,140	13,854	14,531	15,256	16,061	16,846	17,133
Tanzania	340	395	406	436	436	479	499	515
Thailand	1,205	1,450	1,452	1,485	1,490	1,769	1,810	2,859
Tunisia	262	264	269	279	295	305	318	320
Turkey	498	498	498	663	1,260	1,383	1,395	1,534
United Kingdom	33,157	36,143	39,600	43,098	46,324	50,113	53,983	57,145
United States of America	113,101	116,989	124,728	130,775	136,053	140,372	143,977	150,924
Uruguay	697	782	825	900	965	991	1,127	1,264
Venezuela	420	484	652	931	1,062	1,428	1,529	1,698
Yugoslavia ¹	4,484	4,726	2,391	2,391	2,391	2,369	2,722	2,722

5.30. ábra: Az ISDS adatbázis időszaki kiadványainak földrajzi megoszlása

Egy dolog az, hogy felismerjük az időszaki kiadványok kiadásának mintáit, a másik az, hogy megismerjük példányszámukat (az előfizetések számát), amely jelezheti népszerűségüket. Az ilyen adatokhoz nem juthatunk hozzá közvetlenül egyik címjegyzékből sem, bár az Ulrich's és az EBSCO azt állítja, hogy megadják az ilyen adatokat. A 7., a pontossággal és a 9., a teljességgel foglalkozó fejezet illusztrálja, hogy ezek az adatok milyen pontatlanok és hiányosak (ebből következően megbízhatatlanok) lehetnek.

Az Institute for Scientific Information által figyelt folyóiratok földrajzi megoszlása közvetetten, de jól reprezentálja, hogy a szakmai tudományos fogyasztói piac hogyan értékeli az időszaki kiadványoknak a megjelenés országa szerinti fontosságát.



5.31. ábra: Az időszaki kiadványok földrajzi megoszlása az ISI Journal Citation Reports adatbázisának társadalomtudományi sorozatában

Ezek a bevezető adatok biztosan segítenek abban, hogy az adatbázisok forrásainak földrajzi megoszlását megfelelő perspektívába helyezzük. Alapvetően a megcélzott használói közönség érdekeinek kell tükröződnie egy adatbázis forrásainak földrajzi megoszlásában. A nemzeti

bibliográfiákon és a közös katalógusokon kívül sok adatbázis van még, amelyben egyetlen országból származnak a kiadványok vagy termékek vagy a vállalkozás helye egyetlen ország. Ez teljes mértékben érhető mindaddig, amíg kereslet van irántuk.

Az egyetlen forrásra épülő újság adatbázisok, mint például a *Guardian*, az *Irish Times* vagy a *China Morning Star* esetében nyilvánvalóan egyetlen ország az elsődleges földrajzi forrás (még akkor is, ha a nemzetközi eseményeket is feldolgozzák). A Canadian Newspapers Database számára elég nagy a helyi piac, hogy igazolja a kanadai megközelítést a földrajzi feltártság szempontjából. A SilverPlatter kanadai adatbázisa (Canadian Database) feltehetőleg a legnagyobb Kanadának szentelt digitális információforrás. A Dun & Bradstreet Canadian Market Identifier adatbázisa ideális forrás lehet azok számára, akiknek információra van szükségük ennek az országnak a cégeiről, amely a legnagyobb vagy második legnagyobb kereskedelmi partnere sok más országnak. Az Új Zéland-i újságok adatbázisa (New Zealand Newspapers database), amely 13 újságot tár fel, talán merészebb projekt a DIALOG-on, de igazán segít a kiviknek és ausziknak, hogy helyi szempontból releváns információhoz jussanak. Az ausztrál adatbázisok nem érhetőek el a legnagyobb nemzetközi online szolgáltatóknál, de az ausztrál adatbázisok helyi és CD-ROM-változatai bizonyosan segítenek orvosolni egy ország irodalma elvesztésének problémáját (amelyet nagyon meggyőzően írt le klasszikus írásában Byrne [1983], maga is ausztrál).

A földrajzi feltártság másik aspektusa az, amely független a megjelenés országától, s egy ország, régió vagy az egész világ tematikus feltárására utal. Az AGRIBUSINESS jó példa ennek illusztrálására. Csak USA-beli forrásokat tár fel; azonban ezek a források gyakorlatilag a világ minden országára vonatkoznak - legalábbis mezőgazdasági szempontból. Másik oldalról a World Databases - sok végzetes hiányossága közül ez az egyik - címe ellenére nem igazán tárja fel az egész világot. Egyértelműen az USA és Európa adatbázisaira koncentrál, így ez valójában sokkal kevésbé nemzetközi, mint a Gale Directory of Databases.

Az ALA könyvtári és információs szakembereket tartalmazó címjegyzéke (Directory of Library and Information Professionals) CD-ROM adatbázisa azzal dicsekszik, hogy anyaga az egész világra vonatkozik. Formálisan valóban sok ország van reprezentálva benne, de közülük sok csak egy személlyel (aki feltehetőleg egy testületi tag ALA összekötője). A rekordoknak kevesebb mint 10%-a vonatkozik az Egyesült Államokon kívüli szakemberekre. Még Nagy Britannia is csak 34 személlyel szerepel, pedig könyvtáros-egyesületének (Library Association) több mint 20000 tagja van.

Sok adatbázisra jellemző a regionális feltártság. Az Info Latino America adatbázisnak (amelyet Floridában készítenek) és a Latin American News adatbázisnak (Mexikóban készül) egyértelmű az irányultsága. Azonban, míg az előző nagyon sok lapot dolgoz fel (körülbelül 1500-at), addig az utóbbi eredetileg két újságra korlátozódott: az *El Norte* és *Reforma* című lapokra, amelyek csaknem 90%-át adják az adatbázis tartalmának (5.32. ábra).

Dialog Response		
Ref	Items	Index-term
E1	0	*JN=
E2	136	JN=ANDEAN GROUP REPORT
E3	135	JN=BRAZIL REGIONAL REPORT
E4	217	JN=CARIBBEAN & CENTRAL AMERICAN REPORT
E5	274059	JN=EL NORTE
E6	842	JN=LATIN AMERICAN ECONOMY & BUSINESS
E7	141	JN=MARKET: LATIN AMERICA
E8	234	JN=MEXICO AND NAFTA REPORT
E9	276283	JN=REFORMA
E10	181	JN=SOUTHERN CONE REPORT
E11	7733	JN=WEEKLY REPORT
E12	9619	LA=ENGLISH
E13	550339	LA=SPANISH

**5.32. ábra: A Latin American News adatbázis
forrásainak földrajzi összetétele**

A dél-amerikai üzleti információk adatbázisa másik oldalról címében valamivel többet ígér, mint amit tartalmaz - csupán öt országot Dél-Amerikából, ezek a következők: Argentína, Brazília, Chile, Peru és Uruguay. Az ázsiai, csendes óceáni üzleti folyóiratok adatbázisa (Asia-Pacific Business Journals) a tudomány, technika és üzlet másik fontos régióját dolgozza fel, mind látványos eredményei, mind pillanatnyi pénzügyi pangása miatt. Az Északi-sarkvidéki és Déli-sarkvidéki régiók adatbázisának (Arctic & Antarctic Regions database) hasonló regionális jellemzői vannak, az alaptudományokra és a technikára koncentrálnak. A kelet-európai cégek címjegyzéke (Eastern European Company Directory) nyilvánvalóan elsőszámú forrása a régió cégeire vonatkozó információknak.

Néha a készítő neve sejteti egy adatbázis földrajzi feltártságát, ahogy ez a helyzet a Royal Tropical Institute KIT adatbázisával, az ENSZ Élelmezési és Mezőgazdasági Szervezete (Food and Agricultural Organization) AGRIS adatbázisával és a Pánamerikai Egészségügyi Szervezet (Pan-American Health Organization) LILACS adatbázisával. A szervezet neve azonban nem mindig jelzi a szakterületet. A Nemzetközösségi Mezőgazdasági Iroda (Commonwealth Agricultural Bureaux, most CAB International) CAB Abstracts adatbázisának sokkal szélesebb a spektruma, mint a Brit Nemzetközösségnek, a dokumentumok származási helyét és szakterületét tekintve is. Az adatbázisok földrajzi feltártsága nem mindig felel meg az általános feltevéseknek. A PASCAL, Franciaország vezető multidiszciplináris adatbázisa 84 USA-beli információtudományi folyóiratot dolgoz fel, s csak 14-et Franciaországból. Ez alatta marad a két országban megjelentetett időszak kiadványok arányának is.

Önmagában az, hogy az adatbázis címe tartalmazza egy ország vagy országok földrajzi és politikai csoportjának nevét, nem jelenti azt, hogy tartalma csak azokra korlátozódik. A British Education Indexben van néhány ezer rekord más európai országokból, így például Belgiumból, Franciaországból, Németországból és Hollandiából.

A széleskörű földrajzi feltártság lehet előny vagy hátrány, függően a használók általános vagy adott kereséshez kapcsolódó preferenciáitól. Egy átlagos amerikai nem különösebben érdekelhet egy Magyarországon megjelent folyóirat, de érdeklődése rendkívüli mértékben megnőhet, ha Fulbright-ösztöndíjjal három hónapra Magyarországra kell utaznia, hogy a könyvtárgépesítéssel kapcsolatos tanácsadói munkát végezzen. Az az orvos, aki ópiumszármazékokat használó betegeket kezel, nem nagyon zargatja a könyvtárat azzal, hogy könyvtárközi kölcsönzéssel megkapja a thaiföldi Chiang Rai-ban élő emberek életstílusáról szóló konferenciakötetet, de bizonyára kíváncsi lesz rájuk, amikor az Arany Háromszögbe

utazik. Hátrány lehet, ha hiányzik az anyag egy vagy több országgal kapcsolatban - különösen egy szabadalmi adatbázisban, ahol kétségtelen érték az egész világ anyagának feltártsága. A Fast-Alert és a Patents Preview összehasonlításakor Cheeseman (1995) kimutatta, hogy az utóbbi nem tartalmazta a Franciaországban és Németországban megjelentetett szabadalmi leírásokat.

Az adatbázisok földrajzi feltártságának vizsgálata érdekes jellemzőket tárhat fel. Whitney (1990, 1992, 1993) végezte a legtöbb kutatást, a fizikai tudományok, társadalomtudományok, orvostudomány és műszaki tudományok legfontosabb adatbázisai által feldolgozott anyagot vizsgálva általában, s különös tekintettel a harmadik világból és Európából származó szerzőkre. Vizsgálatai elemzik az adatbázisok trendjeit és az egyes országok szintjén két évtized változásait.

Az, hogy egy adott ország irodalmának egy részét felveszik, még nem feltétlenül jelent megfelelő mértékű feltártságot. Amba és Nares (1994) mutatta ki, hogy bár néhány indiai folyóirat szerepel a CAB Abstractsban, s találhatóak benne a bőriparra vonatkozó cikkek, a bőripar magfolyóiratai közül egyik sem szerepel, annak ellenére, hogy India fontos helyet foglal el a bőripari kutatásban. A kevésbé fejlett országok folyóiratai ritkán szerepelnek a nemzetközi adatbázisokban, bár az AGRIS, CAB Abstracts, MEDLINE és LISA saját szakterületén jelentős mértékben feltárja anyagukat.

A legnagyobb valószínűséggel a nemzetközi együttműködéssel létrehozott adatbázisok tartalmazzák a harmadik világ országainak folyóiratait, ahogy azt Hitti (1995) és Thomas (1990) kimutatta. Az orvosi adatbázisok túlsúlyos kategóriájában az ExtraMED adatbázisnak sikerült egy rést találnia azzal, hogy csaknem 3000 olyan, elsősorban a fejlődő országokban megjelenő biomedikai folyóiratot feldolgoznak, amelyek nem szerepelnek más adatbázisokban. Az európai országokból származó folyóiratok egyes adatbázisok esetében a forgalmukat erősen növelő tényezők lehetnek. Az EMBASE specialitása, az európai folyóiratok feldolgozása szempontjából csökkenő tendenciákat mutat (Briggs és Crowlesmith, 1995). Néhány évvel ezelőtt az adatbázis rekordjainak csaknem fele származott Európában kiadott folyóiratokból. Világos azonban a tendencia, hogy az EMBASE-ben mind az USA-n kívüli folyóiratok, mind a nem angol nyelvű anyag aránya folyamatosan csökken (5.33. ábra).

Search History			
Database Details			
Set	Term Searched	Items	
S1	PY<1990	4183721	Display
S2	S1/NONENG	1041428	Display
S3	PY>1989	3519765	Display
S4	S3/NONENG	411669	Display

Show Database Details for:	
73: EMBASE® (1974-present)	Fullsheet Rates

5.33. ábra: A nem az USA-ban és Kanadában megjelent folyóiratokból származó rekordok csökkenő aránya az EMBASE adatbázisban

Freedman tanulmánya (1995) megmutatta, hogy a BIOSIS adatbázisban a források több mint 50%-a Európából és a Közép-Keletről származott, 30% Észak-Amerikából, 3% Közép-és Dél-Amerikából és 2% Afrikából. Csökkenés figyelhető meg a kelet-európai folyóiratok feldolgozottságában, mivel az ottani országok közül sok elvesztette kutatóit az agyelszívás miatt, s most fontosabb prioritásaik vannak, mint a kutatás. Ez később megváltozhat, ahogy ez

már megtörtént a Sociological Abstractsben feldolgozott kelet-európai szociológiai folyóiratokkal, amelyek feldolgozottsága visszaállt a korábbi szintre az 1990-es évek közepén (Chall és Owen, 1995).

A változó politikai és gazdasági körülmények hatással lehetnek egy meghatározott országból vagy régióból származó folyóiratok feldolgozottságára. Ubico, Baily és Weaver (1995) kimutatta, hogy az olajkutatással foglalkozó TULSA adatbázis a kínai anyagok feltártságának hirtelen növekedését jelezte 1982-ben, amikor Kína lehetőséget adott külföldieknek a petróleumkutatásra. A volt Szovjetunió üzleti lehetőségeinek változó valósága tükröződik az orosz anyagok feltártságának hullámszerűségében. Más területeken az 1990-es évek csökkenő orosz kutatásai tükröződnek az orosz folyóiratok csökkenő feltártságában.

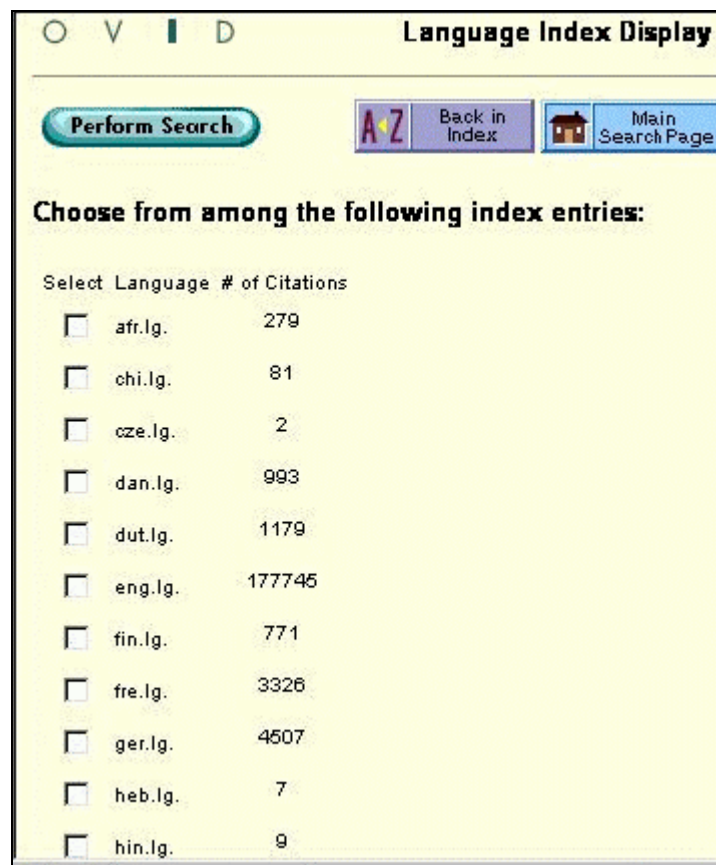
Bár első pillantásra a Social SciSearch könyvtár- és információtudományi részhalma továbbra is feldolgoz két orosz folyóiratot, 1990 és 1994 között a feldolgozott cikkek száma csupán 772-vel növekedett, szemben az 1990 előtről származó 3200 rekorddal. Nem meglepetés, hogy az orosz könyvtár- és információtudományi folyóiratokra vonatkozó hivatkozások száma rendkívüli mértékben lecsökkent a volt Szovjetunió szatellit országainak tényleges függetlensége óta. Az orosz nyelvű folyóiratok idézését - a szerző véleménye szerint - sokkal inkább a politikai szempontok motiválták, mint tudományos okok - legalábbis a könyvtár- és információtudomány területén.

A forráspublikációk megjelenési országát azonosító adatelemek hiánya meglepő a LISA adatbázisban, amely büszkén (és indokoltan) kérkedett azzal, hogy a könyvtár- és információtudományi folyóiratokat az egész világra kiterjedően feltárja. Az ExtraMED adatbázis nem jelzi a folyóirat megjelenési országát, de a cikkek földrajzi szempontjai kereszthetők a deszkriptorokon keresztül.

Néha meglepő egy adott országból származó rekordok viszonylag magas száma. A Responsive Database Services Business & Industry adatbázisában a cseh publikációkat (a magyarokkal és lengyelekkel összehasonlítva), rendkívüli mértékben dolgozzák fel (5.34. ábra). A bársonyos forradalom és a karizmatikus Václav Havel vonzásán kívül semmi sem látszik igazolni az adatbázisnak a Cseh Köztársaság iránti erős preferenciáit. A Library Literature jó keverékét mutatja a feldolgozott folyóiratok földrajzi megoszlásának (5.35. ábra). Bár a kiadás országa szerint nem lehet keresni, a nyelvi index elég jól felfedi ennek a két adatbázisnak a nemzetköziségét.

E3	0	*LA=
E4	326	LA=CHINESE
E5	29853	LA=CZECH
E6	944	LA=DANISH
E7	1393170	LA=ENGLISH
E8	2	LA=ENGLISH
E9	13445	LA=FRENCH
E10	18359	LA=GERMAN
E11	769	LA=ITALIAN
E12	36653	LA=JAPANESE
E13	11	LA=KOREAN
E14	1837	LA=NORWEGIAN
E15	24522	LA=PORTUGUESE
E16	18321	LA=RUSSIAN
E17	25910	LA=SPANISH
E18	15	LA=SWEDEN
E19	1863	LA=SWEDISH
E20	16098	MT=ACCOUNT ACTIVITY

5.34. ábra: Aránytalanul sok cseh nyelvű rekord található a Business & Industry Database-ben



5.35. ábra: A nemzetközi folyóiratok megfelelő keveréke figyelhető meg a LibLit adatbázisban

A két legnagyobb ingyenes filmes adatbázis jó képet ad a filmek származási helyéről, anélkül, hogy keresni kellene bennük. Az All Movie Guide (AMG) betűrendes listát szolgáltat (5.36. ábra), míg az Internet Movie Database (IMDb) (5.37. ábra) azoknak az országoknak a csoportját mutatja, amelyekből 500-nál több film szerepel. Első pillantásra is nyilvánvaló, hogy az IMDb több filmről ad információt Argentínából, Ausztriából, Brazíliából, Kelet-Németországból és Finnországból, mint az AMG.

COUNTRIES:			
Albania	14	Algeria	53
Argentina	416	Armenia	26
Australia	928	Austria	331
Azerbaijan	12	Bangladesh	6
Belgium	310	Bolivia	10
Brazil	326	Bulgaria	146
Burkina Faso	20	Byelarus	74
Cameroon	18	Canada	1,567
Chile	70	China	260
Colombia	39	Croatia	11
Cuba	90	Cyprus	7
Czech Republic	19	Czechoslovakia	464
Denmark	494	East Germany	178
Egypt	86	Estonia	27
Finland	206	France	11,847
Georgia (Republic)	66	Germany	1,263
Ghana	11	Greece	296
Guinea	9	Hong Kong (China)	34
Hong Kong (U.K.)	689	Hungary	649
Iceland	33	India	656
Indonesia	32	Iran	110
Iraq	13	Ireland	123

5.36. ábra: Az All Movie Guide országok szerinti listája

5.37. ábra: Az Internet Movie Database országstatisztikája

A feldolgozott anyag nyelvi megoszlása

A földrajzi megoszláshoz hasonló megfontolások jellemzőek az adatbázisok anyagának nyelvi megoszlására. Első látásra a változatosság vonzó. Hasznos lehet néhány potenciális használó számára, hogy megtalálhatja, mondjuk, magyar nyelvű cikkek bibliográfiai adatait, de a használók döntő többsége számára a változatosság negatív tulajdonsággá válhat. Míg a több forrásdokumentum több kontinensre kiterjedő földrajzi megoszlása nem feltétlenül jelent korlátozott relevanciát a megcélzott közönség számára, a nyelvi megoszlás változatossága nagy valószínűséggel az lehet.

Az esetek többségében az USA-beli, ausztrál és az Egyesült Királyságból származó periodikumok potenciálisan egyformán fontosak azon használók számára, akik angolul olvasnak. Ugyanez érvényes a kanadai folyóiratokra is, amíg vagy kétnyelvűek vagy angolul jelennek meg. Az olyan közönség számára, amely könnyen olvas franciául, egy kanadai folyóirat épp olyan jól használható lehet, mint egy Franciaországban kiadott, amíg francia nyelvű vagy kétnyelvű. Ugyanez érvényes a frankofón, Afrikából származó használó preferenciáira is.

Bibliográfiai adatbázisokban az idegen nyelvű anyag feltárásának van egy olyan további következménye, hogy nehéz a folyóiratot vagy a forrásdokumentumot megszerezni könyvtárközi kölcsönzés útján. Egy 43 felsőoktatási könyvtárosra kiterjedő felmérésben Hernon és Metoyer-Duran (1992) a LISA-ra vonatkozóan azzal a kritikával találkozott, hogy olyan idegen nyelvű folyóiratokat tartalmaz, amelyeket nehéz megszerezni.

Az értékelőtől függ, hogy egy adatbázis anyagának nyelvi megoszlása megfelelő. Spanyol nyelvű dokumentumok nagy száma érték a Spanyolországból, Közép- és Latin-Amerikából származó használók számára, továbbá mindenkinek, aki megfelelő szinten ismeri a nyelvet. Mások számára azonban ez csak ballaszt, amely lelassítja a keresést és felhívja az eredményeket. Ezt a panaszt orvosolhatja a rendszer, ha intuitív nyelvi korlátozást alkalmaz, ha azt megfelelő módon megjelenítik a keresési cella mellett a kereső űrlapon (ahogy ezt később ebben a részfejezetben illusztráljuk).

A nyelvek egyenlősége kardinális kérdés sok nemzetközi szervezetnél. A nyelvek közötti fordítás és a megfelelő képzettségű tolmácsok alkalmazása jelentős költségekkel jár. Az Európai Unió (EU) költségvetésének jelentős részét költi fordításra, például görögből az EU sok más hivatalos nyelvére.

Az adatfájlok készítőinek hasonló problémákkal kell szembenéznük, s egyensúlyozniuk kell az idegen nyelvű folyóiratok referálása és indexelése és teljes szövegű formában való beszerzésének költségei és az irántuk megmutatkozó valós igények között. A gyorsan növekvő költségek késztethették az Elseviert arra, hogy csökkentse a nem angol nyelvű dokumentumok számát az EMBASE adatbázisban. Korábban láttuk (az 5.33. ábrán) az EMBASE nyelvi összetételének drasztikus változásait 1990 előtt és után. Az 5.38. ábra 1975-től 5 évenkénti adatokkal mutatja a csökkenést 1975-től, míg az adatbázisba bekerülő rekordok éves száma azóta is folyamatosan növekszik. A francia, német és elsősorban az orosz nyelvű dokumentumok számának csökkenése különösen jelentős volt. Érdekes, hogy a lengyel nyelvű dokumentumok száma az 1990-es 254-ről 402-re nőtt 1998-ra.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	7815869
S2	S1/NONENG	1463013
S3	S2/1975	74019
S4	S2/1980	64591
S5	S2/1985	57752
S6	S2/1990	62825
S7	S2/1995	41111
S8	S2/1998	38324

5.38. ábra: Az EMBASE adatbázis által feldolgozott idegen nyelvű dokumentumok számának csökkenése öt évenkénti adatok alapján.

A Social SciSearch adatbázis világosan mutatja a nem angol nyelvű dokumentumok arányának csökkenését. 1990 előtt a nem angol nyelvű dokumentumok anyagának 13%-át képezték. 1990-től ez az arány csupán 3% (5.39. ábra). A Sociological Abstracts az 1960-as évektől jelentős mértékben tárt fel idegen nyelvű anyagot, később ez csökkent, de az 1990-es évek végén újra növekedett. A 20%-ról 15,4%-ra való csökkenés a Sociological Abstractsben messze nem volt olyan drámai, de feltétlenül figyelemre méltó (5.40. ábra).

A könyvtár- és információtudományi anyagot feltáró adatbázisok szintén csökkenést mutattak a nem angol nyelvű cikkek számában, bár különböző mértékben. A LISA-ban, amely a legnemzetközibb a könyvtár- és információtudományi adatbázisok között, 32%-ról 18,7%-ra csökkent a nem angol nyelvű anyagok aránya (5.41. ábra). A LibLitben az arány 32%-ról 23%-ra csökkent. Az ISA-ban soha nem volt jelentős mennyiségű a nem angol nyelvű anyag, de a 6,5%-ról 2%-ra való csökkenés nem olyan rossz, mintha azt mondanánk, hogy a nem angol nyelvű dokumentumok mennyisége csaknem 70%-kal csökkent az ISA-ban.

Search History			
Database Details			
Set	Term Searched	Items	
S1	SC=INFORMATION? AND PY<1990	61620	Display
S2	S1/NONENG	8082	Display
S3	SC=INFORMATION? AND PY>1989	79662	Display
S4	S3/NONENG	2207	Display

Show Database Details for:
 7: Social SciSearch® [Bluesheet](#) [Rates](#)

5.39. ábra: A nem angol nyelvű dokumentumok csökkenő aránya a Social SciSearch adatbázisban

Search History			
Database Details			
Set	Term Searched	Items	
S1	PY<1990	309543	Display
S2	S1/NONENG	62985	Display
S3	PY>1989	242883	Display
S4	S3/NONENG	37570	Display

Show Database Details for:
 37: Sociological Abstracts [Bluesheet](#) [Rates](#)

5.40. ábra: A nem angol nyelvű dokumentumok csökkenő aránya a Sociological Abstracts adatbázisban

Search History			
Database Details			
Set	Term Searched	Items	
S1	PY<1990	104333	Display
S2	S1/NONENG	33508	Display
S3	PY>1989	93749	Display
S4	S3/NONENG	17542	Display

5.41. ábra: A nem angol nyelvű dokumentumok csökkenő aránya a LISA adatbázisban

Amíg a keresés nyelv szerinti korlátozása könnyű és intuitív, az idegen nyelvű anyagok nem jelentenek problémát a csak angol nyelvet ismerő használó számára sem. Ez természetesen a keresőszoftver és az interfész megtervezésének funkciója. A DIALOG parancsnyelves keresésében a SELECT LIBRARIANSHIP AND LA=(ENGLISH OR GERMAN OR SPANISH) parancs könnyű a gyakorlott keresőnek, de távolról sem az alkalmi használóknak. Ők jobban szeretnék a nyelvi index legördülő menüit vagy a dokumentum nyelve szerinti keresési filterkockát.

A legördülő menük esetében a használok esetleg nem jönnek rá, hogy egynél több nyelvet is kiválaszthatnak, ha lenyomják a Ctrl (Control) billentyűt, amikor rákattintanak egy nyelvre, ezért az Ovid elismerést érdemel azért, hogy ezt érthetően elmagyarázza (5.42. ábra). A filterkocka alkalmazása kétségtelenül a legintuitívabb, de a választási lehetőség gyakran csak az angol és nem angol. Olyan adatbázisok esetében, amelyekben sok a nem angol nyelvű anyag, hasznos lenne, ha legalább az öt legfontosabb nyelvre vonatkozó filterkocka szerepelne. Bár az ISA nem igazán poliglott, a WebSPIRS változat ideális megoldást mutat erre (5.43. ábra).

☐ V
 ☐ I
 ☐ D

Limit a Search

Select	#	Search History	Results
<input checked="" type="checkbox"/>	1	19\$ yr.	195624

Limit to:

☐ Latest Update
☐ Book Reviews

☐ Articles
☐ Symposia

☒ English

☐ Feature Articles

From: To:

Hold down the shift, ctrl, or "apple" key to make multiple selections or remove selections from a list

Languages

None

Afrikaans

Chinese

Danish

Physical Description

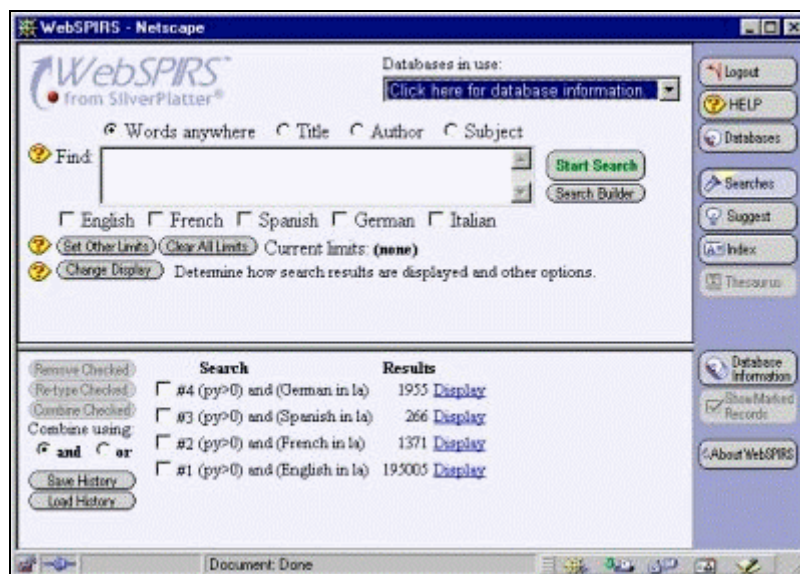
None

Bibliography

Chart

Color

5.42. ábra: A LibLit nyelv szerinti korlátozása a WebOvidban



5.43. ábra: Nyelv szerinti korlátozás az ISA WebSPIRS változatában

Meglepő dolog, amikor egy kifejezetten soknyelvű adatbázisban semmilyen formában nincs nyelvi index (mint az Ulrich's adatbázis esetében). Ez nem igazán jó ötlet, mert az időszaki kiadványok nyelv szerint való keresése nyilvánvalóan fontos (például akkor, ha angol nyelvű, Japánban megjelenő folyóiratokat akarunk találni, vagy francia nyelvű újságokat Nyugat-Afrika országaiból). Az Economic Literature Indexben van bizonyos mennyiségű nem angol nyelvű folyóirat, de arra nincs lehetőség, hogy a keresést a nyelv szerint korlátozhassuk. A vizsgált filmes adatbázisok közül csak az IMDb-ben lehet nyelv szerint keresni.

Mivel a kutatás és fejlesztés sok területén jelentős szerepe van Japánnak, a Japánból származó (különösen az angol nyelvű) kiadványok megismerése nagyon fontos sok szakterületen. Ez olyan életképes piac, hogy a Japán Tudományos és Műszaki Információs Központ (Japan Information Center of Science and Technology) a japánból származó angol nyelvű tudományos és műszaki publikációkról készített egy angol nyelvű adatbázis részhalmazt, a JICST-E-t. Dueltgen (1991) jó meglátásokat tartalmazó összehasonlítást készített négy japán folyóiratot tartalmazó adatbázisról, Bykikeva (1994) összehasonlította az INSPEC, a Compendex és a JICST-E által feldolgozott anyagot. A tesztkeresések során a Compendex mutatta a legalaposabb feltártságot. Sodha (1993) azt mutatta ki, hogy a JICST-E, a MEDLINE és az EMBASE adatbázisokban együttesen az angol nyelvű japán közlemények 70%-a található meg.

A könyvtár- és információtudományi anyagot feltáró adatbázisok közül csak a LISA tartalmazott jelentős mértékben japán nyelvű publikációkról rekordokat - 1999 elején valamivel több mint 3000-et. A LibLitben 135, az ISA-ban 138 ilyen volt, s az ISA régen abbahagyta az ilyen rekordok felvételét. Némileg meglepő módon a Social SciSearchben összességében is viszonylag mérsékelt a japán folyóiratok feltártsága (9 cím), s közülük egyik sem japán nyelvű. Az egyetlen terület, amelyen az MHA - bár minimális mértékben, de - jobb teljesítményt nyújt, mint a PsycINFO (mielőtt az IFI/Plenum átvette volna), a japán és koreai nyelvű dokumentumok halmaza (5.44. ábra).

A nyelvek, nyelvi kódok indexének vagy az adatbázis súgójának pusztá átnézése téves benyomásokkal szolgálhat a feltárt anyag nyelvi megoszlásáról. A Bowker Complete Video Directoryja felsorolja azokat a nyelveket, ahol a nyelv kódja nem azonos a nyelv nevének első három karakterével (5.45. ábra). Azok a használók, akik arra vágynak, hogy információhoz

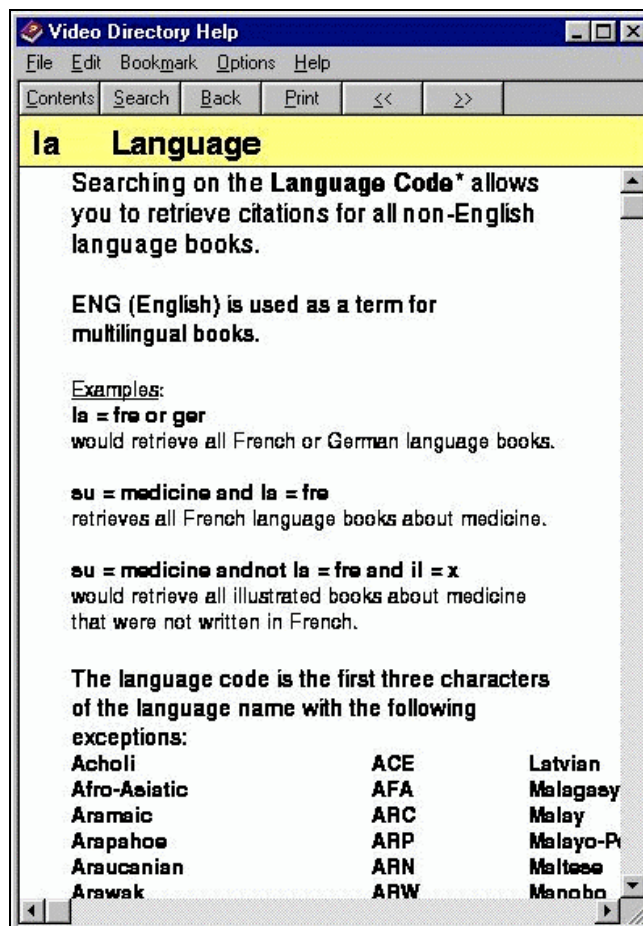
jussanak acholi, arámi vagy arapahoe nyelvű filmekről, csalódottak lehetnek, hogy egyáltalán nem találnak ilyeneket az adatbázisban (s valószínűleg az egész világon sem). A sűgő egyszerően felsorolja azokat a nyelveket és kódjaikat, amelyeket a nyelvészek valamikor is felfedeztek, s amelyekhez a Kongresszusi Könyvtár egy kódot alkotott. Ez a segítő gyöngyszem a *Books in Print* sűgőjából származhat, ahogy az az 5.45. ábrán látható.

A szerkesztő nem foglalkozott azzal, hogy megváltoztassa a sűgőt, s nyilvánvaló, hogy senki sem ellenőrizte, hogy az, ami mintapéldaként szerepel, hogyan működik a valóságban. Érezni lehet, ahogy a reklámszöveg írója feltupírozta a reklámanyagot, amely rendkívüli lehetőségeket sorol fel (5.46. ábra), s kissé túlmént a valóságon, amikor azt ajánlotta, hogy „Állítsa össze az olyan videóok listáját, amelyek megfelelnek meghatározott paramétereknek, mint például „az összes csak szűlői kísérettel látogatható, angol nyelvű, 1998-ban kiadott vígjáték”. A reklámszöveg írója minden bizonnyal nem volt tisztában azzal, hogy az angol nyelvet csak a többnyelvű filmek esetén adják meg. Csak 772 ilyen filmről van rekord ebben az adatbázisban. Nem csoda, hogy az eredmény szárnalmas. Csak egyetlen olyan rekord van, amely megfelel a keresési feltételeknek - ez aligha nevezhető listának (5.47. ábra).

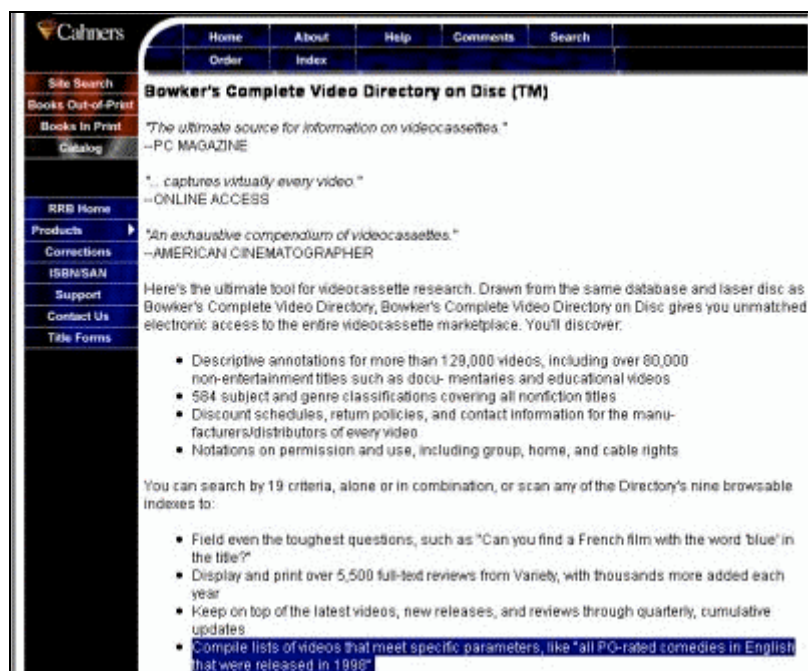
Search History				
Database Details				
Set	Term Searched	Items	File	
S1	LA=JAPANESE	13672		Display
S1	LA=JAPANESE	6289	11	Display
S1	LA=JAPANESE	7383	86	Display
S2	LA=KOREAN	558		Display
S2	LA=KOREAN	241	11	Display
S2	LA=KOREAN	317	86	Display

Show Database Details for:	
86: Mental Health Abstracts	Bluesheet
11: PsycINFO®	
86: Mental Health Abstracts	Limits

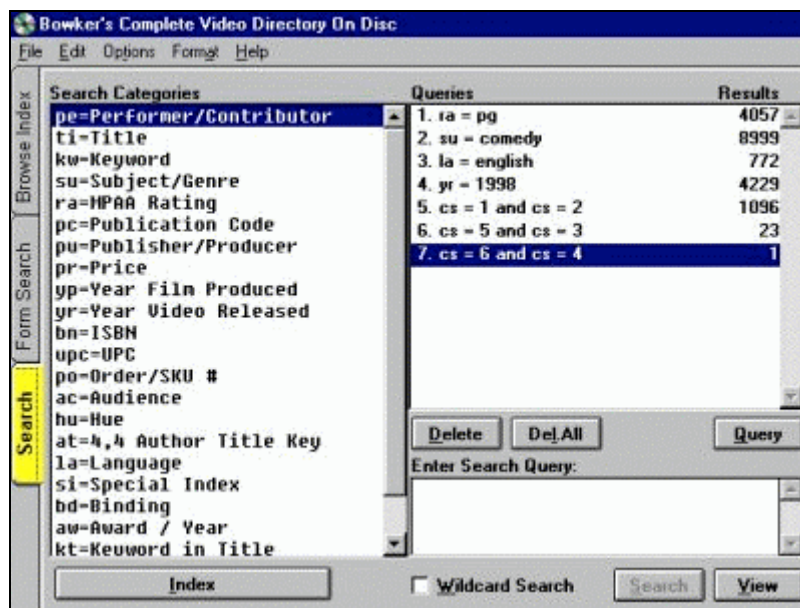
5.44. ábra: Japán és koreai nyelvű anyagok az MHA és a PsycINFO adatbázisokban



5.45. ábra: „Segítő” információ a nyelv szerinti keresésről a Complete Video Directoryban

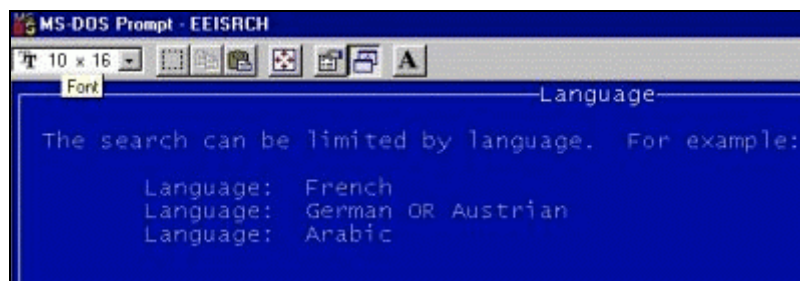


5.46. ábra: Pezsgő ígéretek a Bowker's Complete Video Directory lehetőségeiről



5.47. ábra: A tényleges keresés keserű valósága

Másoknak szintén nehézsége lehet azzal kapcsolatban, hogy megfelelő információkat adjanak az idegen nyelvekről. Az EBSCO-nak sikerült kiválasztania egy nem létező nyelvet (Austrian, osztrák) ahhoz, hogy a sűgőban illusztrálja a nyelv szerinti keresést. A jó hír az, hogy legutóbbi kiadásban a nem létező Austrian csak egyike a nyelveknek (5.48. ábra), s ha önmagában keresünk nyelvként az Austrian szóra, nem találunk egyetlen rekordot sem. Az Austrian (osztrák) épp annyira nyelv, mint mondjuk a Missourian, hiszen Ausztriában az emberek a német nyelv egyfajta dialektusát beszélik, nem pedig egy saját nyelvet.

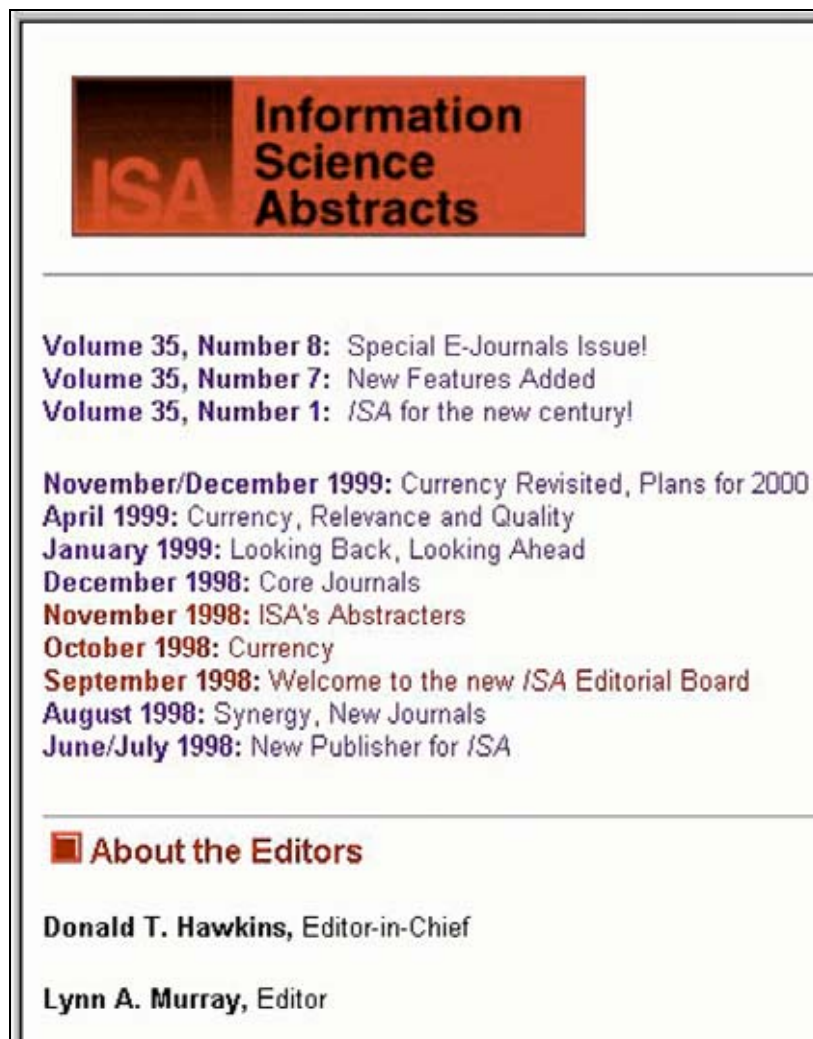


5.48. ábra: Segítő információ a nyelv szerinti keresésről a The Serials Directoryban.

Esettanulmány - Egy adatbázis szakterületének és feltárt forrásainak (disz)harmóniája

Csak szakterülete alapján lehet pontosan értékelni az egy adott adatbázis által feltárt forrásokat. Harmóniának kell lennie az adatbázis állítása szerint feldolgozott szakterület és a kiválasztott források, valamint az anyag következetessége és megfelelő mélysége között. A korábbi részben magukban néztük meg ezeket az elemeket. Ebben a részben azt mutatjuk be - esettanulmányként -, hogyan alakul ez az egyik olyan adatbázisban, amely potenciálisan az egyik legérdekesebb lehet a könyv olvasóinak: az Information Science Abstractsben (ISA). A dolgot nagyban megkönnyítette az a szokatlanul nagy publicitás, amelyben az adatbázis részesült szerkesztőitől az elmúlt három évben. Ez értékes betekintést nyújt azokba az erőfeszítésekbe, amelyekkel egy adatbázist megpróbálnak fenntartani, fejleszteni és népszerűsíteni. A főszerkesztő, Donald T. Hawkins és kisebb mértékben a szerkesztő, Lynne A. Murray sűrűn írták szerkesztőségi cikkeiket az adatbázisról az ISA nyomtatott változatának havi

füzeteiben, a DIALOG és az ISA weboldalára (<http://www.infotoday.com/isa/editorials.htm>), miután az Information Today, Inc. 1998 közepén megvásárolta az adatbázist (5.49. ábra). (Jelenleg nem érhető el, mivel az ISA átkerült az EBSCO-hoz.)



5.49. ábra: Az ISA-ról íródott szerkesztőségi cikkek listája az adatbázis saját weboldalán

Bár a szerkesztőségi cikkek száma az első néhány hónap után csökkenni kezdett, Hawkins továbbra is sok cikket publikált, amelyben bemutatta az adatbázis történetét, a megvalósított (megvalósítandó) változásokat, az adatbázis rossz minőségét (mielőtt a mostani tulajdonos megvásárolta volna) s az ISA helyzetét versenytársai között. Ezek a közlemények rendkívül sokat segítenek abban, hogy megtudjuk, milyen nehézségek adódnak, amikor harmóniát akarunk teremteni egy adatbázis állítólagos szakterülete és az általa feldolgozott források között, amely az egyik legalapvetőbb kérdés, amikor egy adatbázis minőségét értékeljük.

Különösen hálás vagyok Hawkinsnak az ISA-hoz kapcsolódó publikációi miatt, mivel azok folyamatosan arra inspiráltak, hogy tovább foglalkozzam egy olyan adatbázis anatómiájával, amelynek rangos neve van szakmánkban, s szembesíti gyakran változó véleményemet írott (azaz idézhető) állításaival a valóság ellenőrzése érdekében.

Tudjuk, hogy az anatómia tudomány mennyit fejlődött a boncolásoknak köszönhetően, s különösen olyankor, amikor az emberek napi tevékenységeikről készített naplójukkal együtt ajánlják fel holttestüket kutatási célokra, lehetővé téve a kutatóknak, hogy ezekkel az extra információkkal a kezükben elemezzék a boncolási eredményeket. A publikációk az említett

napló funkcióját töltik be az én halál előtti töprengéseimben és később talán a halál utáni következtetésekben - végső soron attól függően, hogy vajon az ISA él-e túl engem vagy én élem túl az ISA-t.

A szakmai elfogadottsággal és az adatbázis minőségével kapcsolatban a főszerkesztő és a szerkesztő azt írta, hogy „tisztában vagyunk az ISA örökségével és azzal a pozícióval, amelyet az információtudósok között elsődleges adatbázisként élvezett” (Hawkins és Murray 2000d). Ugyanabban a cikkben elismerik azt is, hogy „az adatbázis minősége gyenge volt.” Ez az összefoglaló ítélet megjelenik egy másik cikkben is (Hawkins 2001b), amely arra a kritikára íródott válaszként, hogy sok alapvető fontosságú folyóirat feltárása megszűnt vagy rendkívüli mértékben lecsökkent, miután az adatbázis és szerkesztője 1998-ban újra gazdát cserélt. Az ISA rangidős technikai szerkesztőjeként 1998 előtt és főszerkesztőként 1998-tól Hawkins abban az irigylésre méltó helyzetben van, hogy tudhatja, ahogy el is magyarázta, hogy ezeknek a címeknek a korábbi feltártsága „kiváló illusztrációja az ISA gyenge minőségének korábbi vezetősége alatt, amikor nem létezett válogatási politika.” Elmerenghetünk, hogy milyen státuszt szerezhetett volna az ISA az információtudósok között, ha minősége nem lett volna gyenge.

A korábbi vezetőség, Harry Allcock elnök és Anne Meagher szerkesztő javára szólva el kell mondani, hogy a Plenum (az adatfájl korábbi készítője) idején az ISA-nak *volt* válogatási politikája. 50 folyóiratot határoztak meg magfolyóiratként, amelyeket „teljes mértékben fel kell tární”, míg más folyóiratokat „válogatva kell referálni és indexelni”, ahogy azt a nyomtatott változat minden számában leszögezték. Magának Hawkinsnak (aki 1996-tól rangidős technikai szerkesztőként a korábbi ISA team része volt) is érdeméül lehet elismerni, hogy fenntartották a magfolyóiratoknak ezt a körét. Sajnos ezt a megkülönböztető jelzést eltüntették, amikor az Information Today, Inc. vette át az adatbázist. 1999-től az ISA korábbi magfolyóiratai közül sokat egyáltalán nem dolgoznak fel. Ahogy később megtárgyaljuk, az ilyen drámai változások súlyos ellentmondásra vezethetnek egy adott diszciplína alapvető témáinak állítólagos és tényleges feldolgozottsága között.

Nyilvánvaló, hogy Hawkins különösen jó - bár nem szükségszerűen pártatlan - helyzetben volt ahhoz, hogy megítélje az adatbázis minőségét, mivel „rangidős technikai szerkesztőként 1996-tól kapcsolatba került az ISA-val, s meglepve tapasztalta, hogy az információtudomány munkadefinícióját sohasem dolgozták ki ennek a fontos referáló és indexelő kiadványnak a számára”(Hawkins, 2001a). Ez a meglepetés érthető, ha figyelembe vesszük, hogy a tanácsadó testület tagjaiként nyolc könyvtár- és információtudományi egyesület képviselői találkoztak rendszeresen, éveken keresztül az ISA vezetőségének tagjaival.

Azt a nehézséget, amit egy adatbázis szakterületének meghatározása, s ennek megfelelően a primer dokumentumok kiválasztása és feldolgozása jelent, világosan illusztrálták a korábbi és újabban végzett tesztek, amelyekben az ISA és vetélytárs adatbázisai által feldolgozott anyagot vizsgálták független kutatók.

Ahogy azt korábban megmagyaráztuk és kvantitatív formában megmutattuk a 3.10. és 3.11. ábrán, az ISA által feltárt témákkal kapcsolatos állítások és a valóság szinte szöges ellentétben állnak egymással. Sajnos ez most még inkább uralkodó gyakorlat lett, és nagyon kétséges, van-e annak az esélye, hogy a kettő a jövőben összhangba kerüljön.

Az ISA és a LISA által feltárt szakterületek összehasonlítását (Hawkins 2001a) nem igazolják a tények, a cikk nagy mértékben hamis képet fest a valóságról. Ebben a bizonyos cikkben sok témáról állítja azt a szerző, hogy az ISA feldolgozza őket, míg a LISA nem (5.50. ábra), ezek az állítások azonban még tesztkeresés elvégzése nélkül is szembeszökően pontatlanok.

To help resolve this question, the subject classification of *Library and Information Science Abstracts (LISA)* was examined. Table 2 lists some of the differences. The left column lists the subjects that LISA covers that ISA does not. They are more related to librarianship. The right column lists the subjects covered by ISA but not by LISA. These terms are more focused on "pure" information science. In general, ISA concentrates on the theoretical and practical aspects of the knowledge transfer process and its supporting technologies, while excluding coverage of basic librarianship issues such as library facilities, budgeting, personnel management, etc. (Note 4)

Table 2. Librarianship vs. information science.

Librarianship subjects	Information science subjects
Archives	Artificial intelligence, expert systems
Buildings	Basic information science research
Exhibitions	Behavioral sciences
Furniture	Fuzzy logic/fuzzy searching
Library organization	Information industry/workstyle
Library use and users	Information professionals
Loans	Information technologies technical aspects (i.e. computing, Internet, telecommunications)
Management, budgeting, finance	Legislation/regulation
Materials	Natural language processing
Microforms	Reading (literacy)
Museums	Subject area databases
Nequest materials	Types of literature
Old and rare materials	
Promotions	
Removals	
Types of library staff	
Uses (various types)	
Vehicles	
World librarianship	

5.50. ábra: Állítások a kizárólag a LISA-ban vagy az ISA-ban feldolgozott témákról

Egy dolog lenne azt állítani, hogy az ISA a LISA-nál nagyobb mértékben dolgoz fel bizonyos témákat, mint például a mesterséges intelligencia, szakértői rendszerek, fuzzy keresés, távközlés és természetes nyelvi keresés. Hawkins megállapítása, hogy a LISA egyáltalán nem dolgozza fel ezeket a témákat, legfeljebb vágyálom lehet. Egy sommás megállapításban szerepeltetni olyan területeket, amelyeket a LISA jobban feldolgoz (s mindig is jobban feldolgozott), mint az ISA, mint az információipar és -piac, információs szakemberek, Internet, írni-olvasni tudás (olvasás), jog, törvényhozás és szabályozás, nem segít az ISA-n.

Elgondolkodtató, hogy egy ilyen cikk hogyan juthatott át a tekintélyes *Journal of the American Society for Information Science* (JASIS) szigorú lektorálási folyamatán. Egy ok lehet az - ahogy az 5.51. ábra megmutatja - hogy a JASIS által 1990 és 1999 között publikált 746 cikknek csak 5,23%-a származott a kereskedelmi szektorhoz tartozó szerző tollából (Koehler et al., 2000, 11. táblázat) - [<http://www.cindoc.csic.es/cybermetrics/articles/v4i1p3.html>]. Mindig megvan annak a veszélye, hogy a kereskedelmi érdekek, ha nem is szándékosan, de kisebb vagy nagyobb mértékben torzíthatják a kutatási módszereket és az eredmények értelmezését.

Különösen akkor lehet ez a helyzet, ha a szerzők olyan terméket mutatnak be, amelynek a népszerűsítéséhez egyéni pénzügyi érdekük fűződik. A vetélytárs termékeivel és szolgáltatásaival kapcsolatos torzítások sem kizárhatóak. Nem állnak rendelkezésre statisztikai adatok a JASIS-ben megjelent ilyen cikkekről, vagy hogy egyáltalán elfogadtak-e ilyen cikket korábban. Mindenképp kényes kérdésről van szó, amely rendkívül alapos vizsgálatot kíván meg a lektoroktól, hogy elkerüljék a tények hibás bemutatását, akár tudat alatt is. Túl nagy annak a csábítása, hogy valaki a saját lovát dicsérje. Az ilyen cikk súlyos etikai kérdéseket vethet fel, s alááshatja egy folyóiratnak a hitelét egy olyan helyzetben, amikor - a feltételezett lektorálási folyamat ellenére - nyilvánvalóan téves információk jelennek meg tudományos kutatásként.

Egy kutatást bemutató tanulmányhoz feltétlenül szükséges annak alapvető megállapításait támogató, megismételhető kutatási adatok összefoglalásának bemutatása. Sajnos a JASIS-ben megjelent cikkből hiányzik ez az összetevő. Viszonylag egyszerű keresések világosan jelzik, hogy nincs semmi alapja azoknak az állításoknak, amelyek az ISA adatbázis számára monopóliumot határoznak meg vetélytársával, a LISA adatbázissal szemben számos információ-tudományi területen.

Table 11							
AD/JASIS Corporate First Authors – Parent Organization Distribution by Decade							
Decades – Column Percent							All Years
Corporate Author	N	1950-59	1960-69	1970-79	1980-89	1990-99	1950-99
Commercial	386	27.38	40.68	16.06	16.67	5.23	17.79
Educational	1433	25.79	37.57	70.47	70.14	88.47	66.04
Government	147	23.81	7.63	4.92	6.02	2.01	6.77
Library	49	9.52	1.69	0.26	3.01	0.67	2.26
Hospital	1					0.13	0.05
Military	17	2.78	1.41	0.78		0.27	0.78
Organization	79	7.14	8.19	4.15	2.08	0.94	3.64
R&D	58	3.57	2.82	3.37	2.08	2.28	2.67
Total N	2170	252	354	386	432	746	

5.51. ábra: A JASIS-ben megjelent cikkek százalékos megoszlása a szerzők munkahelye szerint

Tesztkereséseket végeztünk 2001 nyarán, hogy összehasonlítsuk a két adatbázisban kapott találatok számát olyan kérdésekre vonatkozóan, amelyek reprezentálják azokat a témákat, amelyekről a JASIS-ben megjelent cikk azt állította, hogy azokban az ISA adatbázisnak van monopóliuma. A kereséseket az alapszótárban végeztem (amely a cím, a referátum és a deszkriptor mezőkből kiemelt szavakból áll).

Fontos megjegyezni, hogy összehasonlítás céljára a címek szavaira vonatkozó keresések a legmegbízhatóbbak, mivel a referátumok és különösen a deszkriptorok mezőjének eltérő szóválasztása és következtelenül alkalmazott szókincse jelentős mértékben torzíthatja az eredményeket. Például az egyik téma, amiről azt állították, hogy csak az ISA dolgozza fel, a LISA pedig nem, az information professionals (információs szakemberek) volt. Az ISA ezt a fogalmat csupán 11 rekordhoz kapcsolta különböző változatokban, míg a LISA több mint 1000-hoz. A deszkriptor mezőt is tartalmazó keresés a LISA-t abszolút győztessé tenné. Az ISA jobban szereti az „information specialists” (információs specialisták) deszkriptort, 1045 rekordhoz kapcsolták ezt, míg ezt a LISA csak 28 rekordban használta, így ebben az ISA arat kiütéses győzelmet. Lehetnek még további, alkalmasszerűen használt deszkriptorok, mint az information brokers vagy searchers (információs brókerek vagy információkeresők).

A cikk azon állításának fényében, hogy „az ISA kontrollált szótára messze túlélte használhatóságát”, s a gyakorlatból ismerve annak nagyon következtelen használatát, hiábavaló dolog lenne, ha a deszkriptor index alapján végzett tesztkeresésekben bízánk. Az indexelésnek ez az elégtelensége természetesen az alapszótár alapján végzett keresések eredményeit is torzíttja. A LISA-nak is megvannak a saját gondjai a deszkriptorok mezőjével kapcsolatban. Miután a Bowker-Saur megvásárolta az adatbázist, valódi tezaurszt fejlesztettek ki, amellyel felváltották a régi „kontrollált” szótárt, de a tezaursz fogalmait nem alkalmazták visszamenőlegesen a régi rekordokban, ebből következően a deszkriptor index nem megbízható eszköz a LISA tesztelésére sem.

Ugyanakkor azonban a címező a két adatbázisban összehasonlítható, mivel egyik sem alkalmaz címkiegészítéseket. A kereséseket az alapszótárban, a címek indexében és a deszkriptor indexben végeztük, de csak a címekre vonatkozó keresések eredményeit használtuk fel az összehasonlításra, mivel ezek mutatják a legmegbízhatóbban a különbségeket. (5.52. ábra).

Annak az erőfeszítésnek a jegyében, hogy az ISA „tisztá” információtudományi jellegét mutassa ki, a JASIS-ben megjelent cikk szemet hunyt több tízezer olyan ISA-rekord fölött, amelyek olyan könyvtártudományi témákkal foglalkoznak, mint a levéltárak, könyvtár-szervezet, könyvtárhasználat és -használók, múzeumok, nem nyomtatott dokumentumok). (A későbbi 5.68. ábra megmutatja ezekre vonatkozóan is a valós adatokat.)

Search terms	All Years						1997-2001					
	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA
	basic index		TI	TI	DE	DE	basic index		TI	TI	DE	DE
artificial intelligence	2,059	9,606	227	702	1,445	6,870	869	629	55	9	717	346
expert systems	1,278	3,249	269	694	887	2,964	246	101	28	15	182	83
fuzzy (logic or searching)	95	208	23	78	26	6	50	39	12	13	13	0
fuzzy	330	1,007	161	641	157	774	138	175	75	118	73	139
information (industry or marketplace)	1,462	1,815	175	132	866	1,496	602	542	34	19	489	441
information professionals	2,368	1,130	361	175	1,007	35	980	524	144	80	423	3
information specialists	795	1,473	148	73	28	1,045	122	411	23	9	6	336
internet	13,184	7,075	4,033	1,769	7,575	5,082	8,398	4,662	2,669	988	4,919	3,194
telecommunications	5,405	11,268	562	925	3,878	4,121	776	649	133	27	473	211
natural language	1,131	1,898	236	550	404	1,051	219	168	52	29	83	76
reading (2nd) literacy	224	85	12	2	164	70	34	10	0	0	15	8
information literacy	688	399	244	118	438	286	406	242	141	72	279	175
information science research	241	157	61	55	82	45	82	23	9	7	65	6
(behavioral or behavioural) sciences	72	116	19	29	14	31	9	7	2	2	2	0
behavi?	3,364	5,670	666	896	809	1,331	1,099	805	206	124	328	320
law or legislation or regulation	9,910	5,449	2,586	1,148	5,253	2,261	2,519	791	617	151	1,416	344
(scientific or technical or medical or business or government) databases	541	477	82	74	253	240	201	74	9	5	108	24

5.52. ábra: A LISA-ban és az ISA-ban végzett tesztkeresések eredményei

A táblázat jobb oldalán található teszteredmények azt is illusztrálják, hogy az ISA-ban feldolgozott anyag hogyan vált kisebbé az IFI/Plenummal kötött megállapodás lejártá után. A fuzzy logika/fuzzy keresés kivételével minden olyan témában, amelyben az ISA monopolisztikus feltárását állította a cikk, a LISA több és jobb eredményt hozott, még akkor is, ha nem számítjuk be a nem angol nyelvű cikkeket, amelyek csak a könyvtár- és információ-tudományi adatbázisok használóinak kis része számára lehetnek érdekesek. Ez nem azért történt, mert az új vezetés a legfontosabb forrásdokumentumokra korlátozta a feltárást, hanem mivel sok, az ISA adatbázis-leírásán hirdetett téma (látható a 4.31. ábrán) szempontjából alapvető folyóirat feltárását abbahagyták vagy rendkívül egyenetlenül végezték.

Míg néhány szempontból megfigyelhető volt fejlődés az adatbázisban (a duplikátumok egy részét megszüntették, az ISA korábbi vezetése által magfolyóiratoknak tekintett lapok feltárási hiányosságainak egy részét megszüntették), a minőség összességében még lejjebb süllyedt. Az az állítás, hogy „az ITI gyorsan észrevette és kijavította a Jacsó által megállapított minőségi hibákat” körülbelül annyira pontos, mint amilyen a szerző nevének a helyesírása a cikkben: egyetlen bekezdésben kétszer hibás, egyszer helyes (5.53. ábra).

1995-ben a Hawaii Egyetemen dolgozó Jascó Péter kapta meg a DAI ösztöndíjat az ISA adatbázis értékelésére. Jascó tanulmányozta az ISA által feldolgozott folyóiratokat, összehasonlította azokat más információtudományi referáló és indexelő adatbázisokkal, és azt jelentette, hogy az ISA minősége számos területen jelentős hiányosságokat mutat. Jascó jelentése (1997) végül azt eredményezte, hogy a DAI felbontotta szerződését a Plenummal és az Information Today, Inc. (ITI) 1998 június 1-jén megvásárolta az ISA-t. Egyedüli kiadványának eladásával a DAI a feloszlás folyamatában van, vagyonát a nyolc szponzoráló társaság között osztja fel.

Az ITI gyorsan észrevette és kijavította a Jascó által megállapított minőségi hibákat, s elkezdte az ISA-t marketing csatornáin keresztül népszerűsíteni. Jelenleg készülnek a tervek az ISA-nak az interneten való elérhetővé tételére.

In 1995, Péter Jascó of the University of Hawaii received the DAI grant. Jascó studied the journal coverage of ISA, compared it to other A&I information science databases, and reported that the quality of ISA was significantly lacking in several areas. Jascó's report (1997) ultimately led to the termination of DAI's contract with Plenum and the acquisition of ISA by Information Today, Inc. (ITI) on June 1, 1998.² With the sale of its only publication, DAI is in the process of disbanding and distributing its assets to the eight sponsoring societies.

ITI quickly addressed and corrected the quality deficiencies identified by Jascó, and began to market ISA through its own marketing channels. Plans are currently underway to make ISA available on the Internet.

5.53. ábra: A gyors javítás körülbelül annyira volt hatásos, mint magának a sikerekről beszámoló bekezdésnek a helyesírás-ellenőrzése

Az tény, hogy az adatbázis sok régi hibája megmaradt 2001 nyarán is, mint például 12000 szabadalmi rekord, amelyeket veszett tempóban átmásoltak az ISA-ba a szabadalmi adatbázisból (amelyet szintén az adatbázis régi készítője hozott létre), duplikált rekordok ezrei, s minden 1990 előtti hiány alapvető magfolyóiratokból. Feldolgozási ürok is felbukkantak. A magfolyóiratok listáját megszüntették, a hiányosságokat azonban nem, s ezzel az ISA-t teljes mértékben alkalmatlanná tették arra, hogy az információtudomány irodalmát követni lehessen a segítségével, pedig a JASIS-ben megjelent cikk címe ezt ígérte.

A 4. fejezetben komoly hiányosságokat láthattunk néhány, az ISA által kiemelt téma alapvető folyóiratainak feldolgozásában, mint a *Scientometrics* (4.25. ábra) és a *Program* (4.24. ábra), amelyeket évek óta nem indexelnek. Sok további folyóirat van, amelyek esetében súlyos múltbeli vagy újabb hiányosságok tapasztalhatók, például a *NFAIS Newsletter*-ből az összes 1999-es szám hiányzik (a 4.37. ábra mutatja), vagy a *Journal of Scholarly Publishing*, amelyből évek hiányoznak 1984 utántól, amikor régi címe, a *Scholarly Publishing* megváltozott.

Miután 1994-től rövid ideig feltárták, úgy tűnik, hogy 1998 után újra abbahagyták ezt (5.54. ábra), bár nem válhatott kevésbé fontossá az ISA által feltárként meghatározott szakterületek szempontjából. Végül is, ha az elektronikus publikálás a 22 felsorolt téma közé tartozik, s

magát a könyvkiadást is felvették az ISA témáinak listájára 2001-ben, ennek a folyóiratnak mindenképp relevánsnak kell lennie. És ez valóban így van, bár a használók az ISA-ból nem szerezhettek tudomást azokról a cikkekről, amelyeket a LISA adatbázisból példaként megadunk (5.55. ábra).

```

RANK: S7/1-202   Field: PY=   File(s): 202
(Rank fields found in 202 records -- 18 unique terms)

```

RANK No.	Items	Term
11	9	1970
14	5	1971
10	10	1972
9	11	1973
3	19	1974
6	12	1975
4	18	1976
2	20	1977
7	12	1978
12	6	1979
5	17	1980
8	12	1981
18	1	1984
15	5	1994
1	29	1995
13	6	1996
16	5	1997
17	5	1998

---end of results---

P = next page Pn = Jump to page n

5.54. ábra: Régi és új hiányok a *Journal of Scholarly Publishing* feldolgozásában

Mindig volt némi ellentmondás az adatbázisban feldolgozottak mondott témák és az ezeknek megfelelő szakterületek alapvető folyóiratai között. Nagyon sajnálatos, ha egy adatbázis fontos szakterületként sorolja fel a döntéstámogató rendszereket, de nem tár fel olyan tekintélyes folyóiratokat, mint a *Decision Sciences*, *Decision Support Systems*, *Foundation of Computing and Decision Sciences*, *Information and Decision Technologies*, hogy csak néhány olyan folyóiratot nevezzünk meg, amelyek a leghivatkozottabbak a kategóriában, s amelyeket nagyon jól feldolgoz az INSPEC, a Compendex és az ABI/INFORM. Az ISA-ban egyetlen rekord sem található ezekből a folyóiratokból, bár a döntéstámogatás az egyik feltártnak mondott terület.

DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
From Gutenberg to gateway: electronic publishing at university presses.
DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
Fair use: a doubled edged sword.
DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
Towards electronic journals: realities for scientists, librarians, and publishers.
DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
The utility of publisher websites.
DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
What do the readers think? A look at how scientific journal users see the electronic environment.
DIALOG(R)File 61:(c) 2001 Reed Reference Publishing. All rts. reserv.
Competition and cooperation: libraries and publishers in the transition to electronic scholarly journals.

5.55. ábra: Néhány releváns cikk a *Journal of Scholarly Publishing* című folyóiratból, amelynek feldolgozását 1998 után az ISA abbahagyta

A múltban azonban ez a kisebb problémák közé tartozott. A nagy probléma az, hogy az adatbázis által feltárt szakterületek magas minőségű folyóiratainak feltárása következtelen volt. Ez a következtelenség az új vezetés alatt is folytatódott, sőt, sok minőségi információ-tudományi folyóirat feldolgozottsága fokozatosan csökkent az ISA-ban. 1999 végére ez nyugtalanító mértéket ért el, 2000-re tovább romlott, vészesen lesóványítva az adatbázist (Jacsó 200b). Az 5.56. ábra azt az eredeti illusztrációt mutatja, amely a feldolgozott folyóiratok számának gyors csökkenését demonstrálta, amelyek többségét magfolyóiratként határoztak meg, amíg ezt a kategóriát meg nem szüntették. (2001-ben egyes folyóiratok esetében néhány rekord bekerült az adatbázisba, de a trend alapvetően nem változott meg.)

	1996	1997	1998	1999	2000
Artificial Intelligence	32	52	48	4	0
IEEE Communications Magazine	66	15	5	6	0
IEEE Transactions on Information Theory	241	196	102	4	0
IEEE Transactions on Systems, Man,...	170	177	50	5	0
Information Sciences	121	92	31	0	0
Information Society	13	23	12	8	3
Information Systems	31	25	19	8	6
Information Systems Management	52	56	42	19	0
Journal of Supercomputing	15	14	0	0	0
M.D. Computing	28	42	23	1	0
MIS Quarterly	21	20	9	7	1
Neural Networks	113	98	7	1	0

5.56. ábra: Folyóiratok feltártságának gyors csökkenést mutató trendje az ISA-ban az adatbázis által alapvetőnek mondott témákban

A főszerkesztő nem osztotta a gyors csökkenésre vonatkozó aggodalmakat (Hawkins, 2001 a). Kíváncsi volt arra, hogy van-e bárki, aki képes „racionális érveket felhozni amellet, hogy a *Journal of Supercomputing* vagy az *MD Computing* a nagypresztízsű, információtudományi magfolyóiratok közé tartozik?” Kevesen hajlanának arra, hogy pozitívan válaszoljanak erre a kérdésre. Végül is, ki vitatkozna egy rangidős technikai szerkesztővel, akinek a feladatai közé tartozott az is, hogy ellenőrizze az ISA-ba felveendő anyagok kiválasztását: Nos, én megkockáztatom azt, hogy a *Journal of Supercomputing* feltétlenül magfolyóirat, ha egy információtudományi adatbázis azt állítja, hogy nemcsak a számítógép-tudomány, hanem kifejezetten a szuperszámítógépek alkalmazása is a 22 alapvető téma közé tartozik. Pedig ez található a 2000 júliusi adatbázis-leírason, ahogy azt korábban, a 4.31. ábrán már megmutattuk. (Érdekes, hogy mindkettőt levették az adatbázis-leírás témái közül 2001 elején, miután a Savvy Searching rovat 2000 decemberi részében ezt a hiányosságot tárgyaltuk [Jacsó 2000d] - 5.57. ábra).

1. ☐ 7/TI, ID, PD, OB/1
INFORMATION SCIENCE ABSTRACTS - File 202

Subject Coverage:
Subjects indexed include: -- Abstracting and Indexing Publications and Services -- Artificial Intelligence and Expert Systems -- Bibliometrics -- Cataloging, Classification, and Indexing -- CD-ROM Databases and Technology -- Copyright and Intellectual Property -- Current Awareness -- Databases, Database Production, and Data Storage Technology -- Digital Libraries -- Document Delivery -- Electronic Information Systems and Services -- Electronic Publishing -- Informatics -- Information Access -- Information Industry -- Information Management -- Information Sources and Services -- Information Technologies -- Information Theory -- Internet and World Wide Web -- Libraries and Library Operations -- Management Information Systems -- Multimedia Information Systems -- Natural Languages Processing -- Online Information Retrieval -- Properties of Information -- Publishing -- Searching and Search Engines -- Types of Literature
DIALOG FILE DATA

BLUESHEET STATUS
Date of Printed Bluesheet: February 16, 2001
Online Bluesheet Status: February 16, 2001
DIALOG Bluesheets(TM) (Dialog® File 415): (c) 2001 The Dialog Corporation. All rights reserved.

5.57. ábra: Az ISA által legfontosabbnak mondott, feltárt területek az adatbázis-leírás 2001 februári változatában.

A Springer Verlag által kiadott *M.D. Computing*ről azt lehet mondani, ha az orvosi informatika az adatbázis-leírás 2001 februári változatában (amelyet az adatbázis új vezetősége idején többször aktualizáltak, elég lehetőség lett volna arra, hogy ezt is módosítsák) az egyik kiemelten említett téma, akkor az *M.D. Computing*ot nem lehet figyelmen kívül hagyni. Pedig kétségtelenül ez történt az első néhány olyan évben, amikor rangidős technikai szerkesztő bekapcsolódott az anyagok kiválasztásába (ahogy azt az 5.56. ábra illusztrálja), feltehetőleg „racionális érvekre” alapozva. A folyóirat azonban jó okból maradt rajta az ISA magfolyóiratainak listáján. [Az *M.D. Computing* a három leghivatkozottabb forrás közé tartozik az orvosi informatikával foglalkozó társadalomtudományi folyóiratok cikkei alapján, ahogy azt a Social SciSearch adatbázisban végzett hivatkozási keresés eredményei mutatják (5.58. ábra). (Két címváltozatot, az *MD Computing* és az *M.D. Computing* formát kombináltuk, hogy igazodjunk a helyesírási variációkhoz.)]

Később természetesen magát a magfolyóiratok listáját is eltüntették, feltehetőleg azért, mert kényelmetlen és kellemetlen bizonyíték maradt annak illusztrálására, hogy hány, az ISA által magfolyóiratként meghatározott forrást nem dolgoztak fel egyáltalán az új vezetés idején.

Dialog Response		
MD COMPUTING AS CW IN SOCSCI		

RANK: S10/1-121 Field: CW= File(s): 7		
(Rank fields found in 112 records -- 1497 unique terms) Page 1 of 188		
>>>The * indicates a user-precombined rank term;		
>>>use DETAIL option to see which terms are precombined.		
RANK No.	Items	Term
-----	-----	-----
1	50	J AM MED INFORM ASSN
2	44	METHOD INFORM MED
3	39	JAMA J AM MED ASSOC
4*	27	M D COMPUT
5	26	B MED LIBR ASSOC
6	23	ANN INTERN MED
7	21	ACAD MED
8	20	NEW ENGL J MED
P = next page Pn = Jump to page n		
P- = previous page M = More Options Exit = Leave RANK		

5.58. ábra: Az *M.D. Computing* az orvosi informatikával foglalkozó társadalomtudományi cikkek között a leghivatkozottabb források közé tartozik.

A többi, feldolgozottságának mélypontját 1998 óta elérő folyóirattal kapcsolatban Hawkins tovább megy, és azt állítja, hogy „az *Artificial Intelligence* és a *Neural Networks* csak alkalmasszerűen közöl olyan cikkeket, amelyek relevánsak az információtudósok számára, de aligha nevezhetőek a területen magfolyóiratnak.” Újra csak, ahelyett hogy elfogadnánk azt a megalapozatlan véleményt, amely a kialakult gyakorlatot védi, nézzünk széles körben ismert tudományos mérőszámokat. A Journal Citation Reports 2000-es kiadásában a mesterséges intelligencia kategóriájában az *Artificial Intelligence* című folyóirat a második helyre került a kapott hivatkozások száma szerint és ötödik lett az impakt faktor alapján. A JCR 2000 szerint 2000-ben 4320 hivatkozást kapott az előző két évben megjelent cikkekre. A *Neural Networks* a 13. volt az impakt faktor szerint és 6. a kapott hivatkozások száma (2481) alapján a mesterséges intelligencia kategóriában (5.59.a ábra). A mesterséges intelligenciáról szóló cikkekben hivatkozott munkákra vonatkozó egyszerű keresés egyértelműen megmutatja az *Artificial Intelligence* jelentőségét, amely messze a leghivatkozottabb az elmúlt 25 évben, akár a természettudományi (5.59.b ábra), akár társadalomtudományi (5.59.c ábra) folyóiratokat nézzük az Institute for Scientific Information nagyhírű adatbázisaiban. Irracionális dolog, hogy feldolgozottsága az ISA-ban rendkívüli mértékben lecsökkent, majd megszűnt (5.59.d ábra).

Journal Citation Reports - 2000 Science Edition

File Edit View Calculations Options Help

Journal Rankings

Mark	Rank	Journal Abbreviation	ISSN	2000 Total Cites	Impact Factor	Immediacy Index	2000 Articles	Cited Half-life
	5	ARTIF INTELL	0004-3702	4320	2.029	0.364	77	9.2
	6	DATA MIN KNOWL DISC	1384-5810	131	1.931	0.417	12	3.1
	7	IEEE T FUZZY SYST	1063-6706	992	1.873	0.014	72	5.2
	8	INT J COMPUT VISION	0920-5691	1681	1.835	0.100	60	7.0
	9	ARTIF INTELL MED	0933-3657	355	1.793	0.083	36	3.5
	10	CHEMOMETR INTELL LAB	0169-7439	1326	1.462	0.232	69	5.2
	11	AI MAG	0738-4602	505	1.447	0.036	28	7.3
	11	MACH LEARN	0885-6125	1663	1.447	0.209	43	7.6
	13	NEURAL NETWORKS	0893-6080	2481	1.221	0.157	83	7.6
	14	IEEE T NEURAL NETWORKS	1045-9227	3051	1.096	0.217	138	6.4
	15	COMPUT INTELL	0824-7935	326	1.091	0.346	26	6.6
	16	PATTERN RECOGN	0031-3203	3204	1.077	0.130	162	7.9
	17	J ARTIF INTELL RES	1076-9757	243	1.064	0.000	20	4.2
	18	KNOWL ENG REV	0269-8889	244	1.025	1.000	12	5.0
	19	APPL ARTIF INTELL	0883-9514	187	1.017	0.000	36	5.2
	20	IEEE INTELL SYST APP	1094-7167	120	1.000	0.161	56	2.0

Sort: Impact Factor Filter: COMPUTER SCIENCE. 71 Journals 0 marked

5. 59a. ábra: Folyóiratok és mérőszámaik a Journal Citations Reports 2000 mesterséges intelligencia kategóriájában

DIALOGWEB

Command Search new search databases alerts cost logoff help

Dialog Response


Adding title to results...

MOST CITED SOURCES IN ARTIFICIAL INTELLIGENCE

RANK: s2/1-4750 Field: CW= File(s): 34
(Rank fields found in 4247 records -- 42431 unique terms)

RANK No.	Items	Term
1	1380	ARTIF INTELL
2	615	COMMUN ACM
3	592	ARTIFICIAL INTELLIGE
4	527	IEEE T SYST MAN CYB
5	362	AI MAG
6	311	SCIENCE
7	294	COGNITIVE SCI
8	287	PARALLEL DISTRIBUTED
9	256	IEEE T PATTERN ANAL
10	245	INT J MAN MACH STUD

5. 59b ábra: Az *Artificial Intelligence* különböző címváltozatokkal a leghivatkozottabb folyóirat a mesterséges intelligenciával foglalkozó cikkekben a SciSearch adatbázisban

 new search databases alerts cost logoff help			
Dialog Response			
Adding title to results...			
MOST CITED SOURCES IN ARTIFICIAL INTELLIGENCE			

RANK: S1/1-1412 Field: CW= File(s): 7			
(Rank fields found in 1305 records -- 16156 unique terms)			
RANK No.	Items	Term	
-----	-----	-----	
1	398	ARTIFICIAL INTELLIGE	
2	277	ARTIF INTELL	
3	162	COGNITIVE SCI	
4	151	COMMUN ACM	
5	144	SCIENCE	
6	129	MANAGE SCI	
7	118	IEEE T SYST MAN CYB	
8	110	PSYCHOL REV	
9	103	INT J MAN MACH STUD	
10	87	OPER RES	

5. 59c ábra: Az *Artificial Intelligence* különböző címváltozatokkal a leghivatkozottabb folyóirat a mesterséges intelligenciával foglalkozó cikkekben a Social SciSearch adatbázisban

Set	Term Searched	Items	
S1	JN="ARTIFICIAL INTELLIGENCE"	581	Display
S2	S1/1997	52	Display
S3	S1/1998	48	Display
S4	S1/1999	4	Display
S5	S1/2000	0	Display
S6	S1/2001	0	Display

Show Database Details for:			
202: Information Science Abstracts		Bluesheet	Rates

5. 59d ábra: Az *Artificial Intelligence* feldolgozottsága teljesen megszűnt az ISA adatbázisban, annak ellenére, hogy a szakemberek abszolút mértékben elismerik.

Az *IEEE Transactions on Information Theory* című folyóirattal kapcsolatban Hawkins cikke egy szám tartalomjegyzékét használja annak illusztrálására, hogy „csak egy cikk kapcsolódik az információtudomány területéhez.” Ez nem nagyon lehet hatással azokra a lojális ISA előfizetőkre, akiknek az előfizetési díja biztosította 2035 dokumentum referálásának és indexelésének költségeit egyedül ebből az időszaki kiadványból a megelőző években, köztük abban az időszakban, amikor a szerző volt az ISA rangidős technikai szerkesztője, s a

kiadványt következetesen magfolyóíratként határozták meg. 1996-ban, 1997-ben és 1998-ban 241, 196 és 104 rekordot készítettek ennek az időszaki kiadványnak a közleményei alapján.

Azután, hogy sok éven keresztül teljes mértékben feltárták ezt a forrást, amelyet a 4. helyen rangsoroltak a JCR információs rendszerek kategóriájában (5.60.a ábra), valami meggyőzhette az ISA szerkesztőit, hogy az IEEE kiadvány mégsem releváns az információ-tudósok számára. Csupán 4 rekord került be belőle az ISA-ba 1999-ből. Ha az információelméletet az ISA által feltárt szakterületnek ígérjük, akkor a JCR információs rendszerek kategóriájában az információelmélettel kapcsolatban az egyik leghivatkozottabb forrást (amelynek impakt faktora negyedik legmagasabb) nem szabad hirtelen szinte teljesen figyelmen kívül hagyni.

IEEE About IEEE | IEEE Memberships | Products & Services | Conferences | IEEE Organizations

Search Join Newsroom Shop Site Map Tour IEEE Home

[IEEE Online Catalog & Store](#)

[What's New @ IEEE](#)

[Products & Services Index](#)

[More About Products](#)

[More About Services](#)

Request your [2000 IEEE Publications Catalog](#)

Overviews

- [Books](#)
- [Conference Proceedings](#)
- [Electronic Products](#)
- [Merchandise](#)
- [Self-Study Courses](#)
- [Standards](#)
- [Subscription Packages](#)
- [Videos](#)

[Doing Business with Us](#)

Year After Year, the Top-Cited Publications in the Field are from IEEE

According to ISI's annual Journal Citation Report, IEEE publications continually rank at the top of a variety of fields of technology. Following is a partial listing of some of the most cited publications based on the 1999 JCR.

[Read more](#) about IEEE's highly-cited journals.

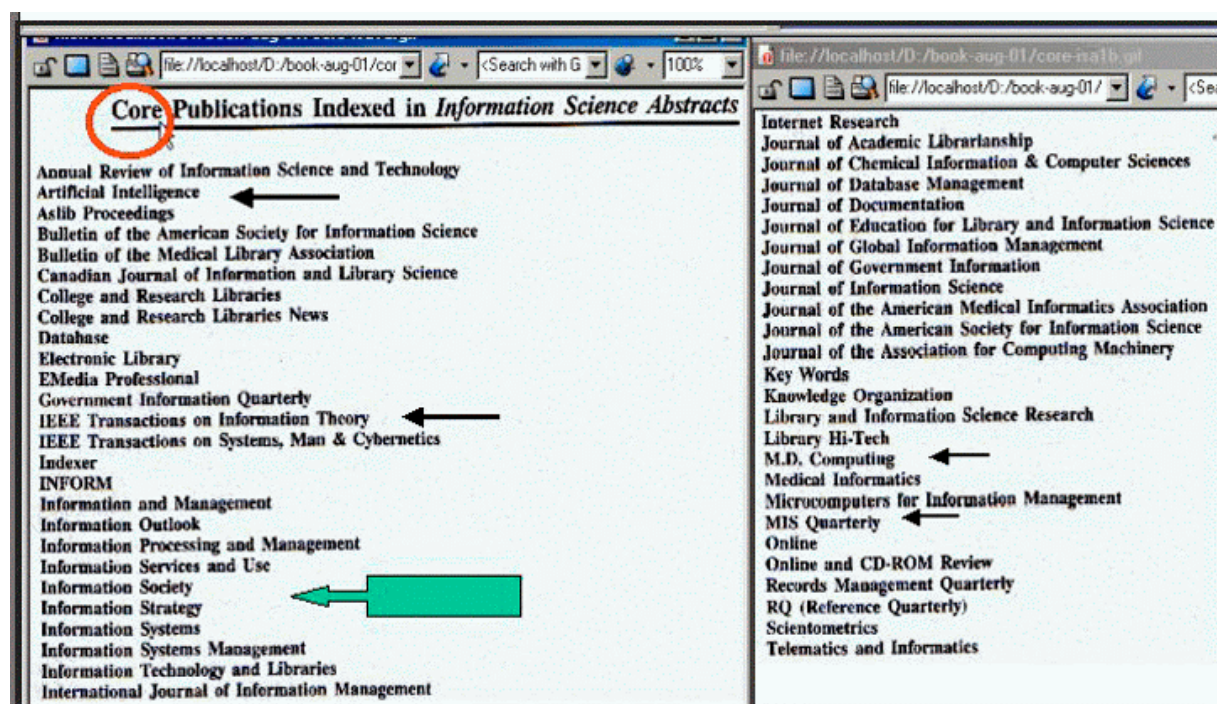
Field	Highly Cited Periodicals
9 of the top 10 in electrical & electronics engineering	Proceedings of the IEEE (#2) IEEE Electron Device Letters (#3) IEEE Transactions on Medical Imaging (#4)
8 of the top 10 in telecommunications	IEEE Journal of Selected Areas of Communications Magazine (#1)
In Computer Science	
7 of the top 10 in hardware and architecture	IEEE Transactions on Networking (#1)
3 of the top 10 in software, graphics and programming	IEEE Transactions on Image Processing (#3)
2 of the top 10 in computer theory & methods	IEEE Transactions on Image Processing (#1)
4 of the top 10 in information systems	IEEE Transactions on Information Theory (#4)

5. 60a. ábra: Az *IEEE Transactions on Information Theory* helye az impakt faktor szerinti rangsorban

A cikk egy hirtelen mondattal elintéz négy folyóíratot, amelyeknek a feltártsága nagyon lecsökkent, mióta a főszerkesztő működése elkezdődött az Information Today-nél. Azt állítja, hogy „jól hangzó címeik ellenére az *Information Sciences*, *Information Society*, *Information Systems* és *Information Systems Management* ... nem „információtudományi” folyóíratok abban az értelemben, ahogyan az IWR legtöbb olvasója definiálná ezt a fogalmat. Csak az információtudomány peremén vannak.”

Egyik olvasója vagyok az IWR-nek (és néhány cikkének szerzője is), és feltétlenül az információtudományhoz sorolom ezeket a folyóíratokat általában, s különösen az információ-menedzsment területéhez tartozónak tartom őket, vagyis ahhoz, amely egyike annak a 22 területnek, amelyeket az ISA adatbázis-leírása kiemelten felsorol. Valóban, nem sokkal korábban, amikor az ISA rangidős technikai szerkesztőjeként működött, minden bizonnyal Hawkins is így gondolta, ahogy azt látni lehet ezen folyóíratok feltártságának tendenciáiból az 5.56. ábrán, hogy ne említsünk további információtudósokat, akik kutatásaik során támaszkodnak ezekre a folyóíratokra, s publikálnak is bennük, mint például Blaise Cronin, az

információtudomány területének egyik legbefolyásosabb szerzője. Ezeknek a folyóiratoknak a feldolgozottsága azonban kezdett nagy mértékben csökkenni, miután az adatbázis gazdát cserélt, bár a négy folyóirat közül hármat az ISA magfolyóirataként határoztak meg 1997-ben (5.60. ábra), a magfolyóiratok listájának megszüntetése előtti évben.



5. 60b. ábra: AZ ISA magfolyóiratainak utolsó listája

Hawkins nem tagadja az *IEEE Transactions on Systems, Man, and Cybernetics* című folyóirat fontosságát (korábban szintén az ISA magfolyóirata volt), amelynek a feldolgozottsága megszűnt 2000-ben (legalábbis az ISA novemberi aktualizálásáig). A folyóiratból 1996-ban és 1997-ben 170 és 177 rekord került be az adatbázisba. Az *MIS Quarterly* sem minősíti relevánsnak Hawkins, pedig ennek az ISA-beli feldolgozottsága is 1 rekordra csökkent 2000-ben az 1996-os és 1997-es 21-ről és 20-ról (ez akkor a folyóirat teljes feldolgozottságát jelentette). Nem volt ez legszerencsésebb idő a *MIS Quarterly* feldolgozásának megszüntetésére, amikor ez a JCR könyvtár- és információtudományi kategóriájában az első, majd a következő évben a második helyen végzett az impakt faktor alapján. Egy elfogulatlan szemlélő számára ezek a jó helyezések azt jelentik, hogy az információtudósok olvassák és idézik ezeket a folyóiratokat, amelyeknek a feldolgozottsága olyan nagy mértékben lecsökkent, miután az ISA gazdát cserélt. Vannak mélyebb problémák is, amelyek nem kerülnek felszínre, amikor az ISA magfolyóiratai alapján alkotott rekordok számát nézzük, de amelyek kényelmetlen helyzetbe hozhatnak egy könyvtárost.

Ezen tudományos folyóiratok megfelelő szintű feltárásának fenntartása mellett maradt volna még hely az „új folyóiratoknak”, amelyeket a tulajdonosváltás utáni szerkesztőségi cikkben jelentettek be. Ahogy megnézzük azoknak az új címeknek a listáját, két dolog ragadhatja meg a figyelmünket. Az első az, hogy számos, az Information Today, Inc. által publikált forrás szerepel rajta, mint például a *Multimedia Schools*, *Information Today* és a *Link-Up*. Ezek hasznos források a gyakorló szakemberek számára, s nekem abban a kiváltságban volt részem, hogy hozzájárulhattam néhány cikkel az elsőhöz és számos rendszeresen megjelenő rovattal a másik kettőhöz (s remélem, hogy a könyv megjelenése után továbbra is megtehetem ezt). Mégsem nevezném őket információtudományi folyóiratoknak. S ugyanígy szinte egyik „új folyóíratra” sem mondanám ezt, amelyeknek a listája megjelent a beharangozóban,

különösen azután, hogy nyomtatékosan hangsúlyozták a „No pasaran!” elvet, azzal a határozottsággal, amivel a rövid életű Spanyol Köztársaság védői, azt állítva „egy dolog van, amit *nem* fogunk csinálni, nem veszélyeztetjük a minőséget azzal, hogy „töltelék” tételeket adunk az adatbázishoz vagy hamis módon megnöveljük az ISA-ban feldolgozott anyagot az információtudományhoz nem kapcsolódó szakterületekkel (Hawkins és Murray, 2000d) és kissé módosított változatban a <http://library.dialog.com/products/f202.html> címen. (Mivel az ISA jelenleg nem érhető el a DIALOG-nál, így természetesen ez a dokumentum sem.)

A második dolog az, hogy az „új folyóiratok” közül 27-et feldolgoznak a Microcomputer Abstractsben (később Internet and Personal Computing Abstracts). A szinergia azt jelenti, hogy újra felhasználják az IPCA referátumait az ISA-ban. Ezek jó minőségű referátumok, de töredékáron elérhetők közvetlenül az IPCA-ból - akár nyomtatott, akár online formátumban, ezért nem sok haszna van a használnak, aki könnyen végezhet keresést egyszerre több adatbázisban. Nem igazán érzékelhető a szinergia az IPCA szempontjából, amelynek a vezetősége kiváló munkát végez New Jerseyben, ugyanabban az irodaépületben, szerényen, és feltehetőleg a „The Sopranos” című televíziós showban néha látható harc nélkül eltekintett attól, hogy elemezze, milyen előnye származik az IPCA adatbázisnak ebből.

Ha tisztán a befektetés megtérülése szempontjából nézzük, akkor a szinergia terv természetesen vonzóan látszhatott. Igazán életerős üzleti terv részét képezheti az, ha meglevő bibliográfiai rekordokat használnak fel minimális módosítással (mint az ISA deskriptorainak és rovatcímeinek hozzáadása), nagyobb mértékben építenek a házon belül kiadott folyóiratokra vagy a fogyasztóknak szánt olcsó folyóiratok (mint a *PC World*, *PC Magazine*) meglevő előfizetésére, csökkentik a tudományos folyóiratok referálási kiadásait és előfizetési díjainak költségét (ez a legkomolyabb költségtényező egy referáló és indexelő adatbázis fenntartásában), miközben megtartják vagy akár még növelik is a nyomtatott és CD-ROM-változatok előfizetési díját és az online változat használati díját.

Elég sokat elmond az ISA 1998-ból származó folyóirat-kiválasztási politikájának kiválóságáról, hogy egyetlen rekord sem került be ezeknek a folyóiratoknak a 2001-es számaiból az adatbázisba; legalábbis 2001 júliusáig. Olyan is csak nagyon kevés volt, amelyből néhány 2000-es rekordot felvettek. (5.61. ábra)

Az is sokat elárul az adatbázist fenntartása mögött rejlő filozófiáról, hogy miután megjelent „Az adatbázis veszélyeztetett fajai”-ról szóló cikkünk, amelyben az ISA a veszélyeztetett fajok listájára került, az ISA adatbázis-leírását 2001 februárban újra kicserélték. Nem volt erre vonatkozó bejelentés és a DIALOG nem küldte el a nyomtatott adatbázis-leírásokat (blue-sheeteket) a szokásos módon a *Chronolog Newsletter*rel együtt. Érdeklődésemre azt válaszolták a DIALOG ügyfélszolgálatának képviselői, hogy az adatbázis-leírást az információ-szolgáltató kérésére cserélték ki, s erről azért nem adtak ki közleményt, mert a változások nagyon kicsik voltak.

Ez figyelemreméltó, mert korábban a DIALOG akkor is kiadta és elküldte az új adatbázis-leírást, amikor csak egyetlen szót változtattak meg az ISA adatbázis-leírásában. Például a 2000 májusi adatbázis-leíráson az aktualizálási gyakoriságot rendszertelenként határozták meg (5.62.a ábra), a DIALOG 2000 júliusában ezt egy új adatbázis-leírással váltotta fel, amelyen az aktualizálás gyakoriságát évi 9 alkalomra változtatták (5.62.b ábra). Ez talán érthető lett volna, ha a módosítást tények igazolnák. Ahogy az UD indexmező egyértelműen megmutatja a történetet (5.63.a ábra), az aktualizálások éves számát már az előző évben nyolcra csökkentették, így az évi 9 aktualizálás nem volt igazán nagy hír 1999 közepén, bár az igazán rossz hír volt a használók számára, hogy ez nem is fog változni.

NEW journals announced for ISA	1998	1999	2000	2001
AI Magazine	5	0	0	0
Booklist	5	3	0	0
Bottom Line: Managing Library	16	0	0	0
Bulletin of the Japan Spec. Libr. Associ.	18	18	18	0
Byte	1	0	0	0
Computer Life	2	0	0	0
Computerworld	10	4	3	0
DBMS	1	0	0	0
Digital Publishing Strategies	0	0	0	0
Fee for Service	0	0	0	0
Forbes	1	1	1	0
Fortune	2	0	2	0
Home Office Computing	1	0	0	0
Imaging	4	2	0	0
Information Today	134	84	56	0
InfoWorld	2	2	3	0
Inside the Internet	0	0	0	0
International Journal of Information Sciences and Decision Making	3	2	0	0
Internet World (was WebWeek)	44	16	2	0
IntraNet Professional	8	11	6	0
LAN Times	2	0	0	0
Link-Up	18	26	19	0
Multimedia Schools	5	6	27	0
Network	3	1	0	0
New Media/NewMedia	2	0	0	0
PC Magazine	29	2	2	0
PC Week	9	6	1	0
PC World	2	2	0	0
Proc Geoscience Info Society	8	0	0	0
Proceedings of the Off-Campus Library	27	0	0	0
School Libraries Worldwide	6	0	0	0
Software Magazine	6	0	0	0
WebWeek (now Internet World)	0	0	0	0
Windows Magazine	2	0	0	0

5.61. ábra: Az „új” folyóiratok feldolgozottsága 2001 júliusában

Information Science Abstracts					
Last Loaded on Web: Friday, May 26, 2000 Last Update To Bluesheet: May 17, 2000					
Bluesheet Contents		PDF version			
File Description	Dialog File Data	Special Features	Sample Record	Sort	Rates
Subject Coverage	Database Content	DIALINDEX/OneSearch Categories	Basic Index	Rank	
Sources	Document Types Indexed	Contact	Additional Indexes	Map	
Print Counterparts	Geographic Coverage	Terms and Conditions	Limit	Predefined Format Options	
Dates Covered: 1966 to the present File Size: More than 208,000 records as of December 1999 Update Frequency: Irregular					

5. 62a. ábra: Az adatbázis-leírás helyesen rendszertelennek jelöli az ISA aktualizálási gyakoriságát

		202
Information Science Abstracts		
FILE DESCRIPTION Information Science Abstracts provides references and abstracts in the fields of information science and library science. International in scope, the database indexes and abstracts articles from over 300 journals, as well as from books, research reports, conference proceedings, and patents. Each record includes the bibliographic citation and abstract as well as descriptors drawn from a controlled vocabulary. The Information Science Abstracts database corresponds to the printed publication of the same name, which is published by Information Today, Inc.		
SUBJECT COVERAGE Subjects indexed include: <ul style="list-style-type: none"> • Abstracting and Indexing • Artificial Intelligence • Bibliometrics • Cataloging, Classification, and Indexing • CD-ROM Technology • Computer Science • Decision Support • Document Delivery • Educational Technology • Electronic Publishing • Information Management • Information Systems Design • Information Theory • Library Administration • Management Information Systems • Medical Informatics • Multimedia Systems • Online Information Retrieval • Optical and Laser Technology • Standards • Supercomputers • Telecommunications 		
DIALOG FILE DATA Inclusive Dates: 1966 to the present Update Frequency: 9 times per year File Size: More than 208,000 records as of December 1999		
CONTACT Information Science Abstracts is produced by Information Today, Inc.. Questions concerning file content should be directed to: Information Today, Inc. Department ISA 143 Old Marlton pike Medford, NJ 08055-8570 Phone: 609-654-6266 Toll Free: 800-300-9868 Fax: 609-654-4309 Donald T. Hawkins Editor-in-Chief Phone: 908-647-4754 Fax: 908-647-4754		

5. 62b. ábra: Az ISA adatbázis-leírásának következő változata új hírként közli, hogy évente 9 alkalommal fogják aktualizálni

select	Update	Records
<input type="checkbox"/>	199901	609
<input type="checkbox"/>	199902	469
<input type="checkbox"/>	199903	433
<input type="checkbox"/>	199904	495
<input type="checkbox"/>	199906	452
<input type="checkbox"/>	199908	422
<input type="checkbox"/>	199910	385
<input type="checkbox"/>	199912	419

5. 63a. ábra: Az ISA adatbázis tényleges aktualizálási gyakorisága

Az aktualizálás rendszeressége mindig nagy probléma volt az ISA számára, és sohasem volt meghatározott, rögzített. Ígéreteket tettek, majd ezeket az ígéreteket megszegették. A legutolsó nyilvános ígéretet az ISA egy szerkesztőségi cikkében tették, az aktualizálás időközének csökkentésével együtt, körülbelül 6 hetes időközöket terveztek a nyomtatott változat számára, és hasonlót az online verzióra (5.63.b ábra). A valóság ettől nagy mértékben különbözött. Nem sokkal azt követően, hogy a szerkesztőségi cikket megírták, az ISA (File 202) aktualizálása újra rendszertelenné vált, s nem aktualizálták márciusban, áprilisban és májusban, míg vetélytársait, a LISA-t (File 61) és a Library Literature-t (File 438) megfelelő rendszerességgel aktualizálták.

Egy korábbi szerkesztőségi cikkben megtárgyaltuk a frissesség problémáját, ebben a számban, amely az utolsó 1999-ben, szeretnénk újra foglalkozni a témával, s bemutatni az ISA-ra vonatkozó 2000-es terveinket. Szerkesztőként növekvő aggodalommal figyeltük azt a gyenge frissességre utaló téves benyomást, amit az ISA aktualizálási dátumai keltenek használóinkban. Az ISA csupán azért látszik vétkesnek, mert lassan kapjuk meg a folyóiratokat a kiadóktól, nem pedig azért, mert bármi elmaradásunk lenne az adatbázis előállításában. Valójában az ISA-nak soha nem volt elmaradása; miután megkaptuk a folyóiratokat, azok anyaga bekerül az adatbázis következő számába. 1999 nyarán a probléma különösen kritikussá vált; a beérkezett folyóiratok száma a szokásos mennyiségnek csak a fele volt. Nem tudjuk, hogy milyen különleges körülmények magyarázhatnák meg ezt.

A fent leírt problémához kapcsolódóan a következő lépéseket tesszük meg:

A mostani az utolsó 1999-es szám, november/december megjelöléssel

2000-ben kilenc alkalommal fog megjelenni az ISA, megközelítőleg hathetes időközökben. A számozás nem fog megfelelni meghatározott megjelenési időnek vagy hónapnak. A következő szám, amelyet meg fognak kapni, a 35. évfolyam, 1. sz. 2000 jelölést fogja tartalmazni. Az online adatbázis aktualizálása hasonló ütemű lesz. A CD-ROM adatbázis továbbra is negyedévenként fog megjelenni.

Azt tervezzük, hogy jelentős mértékben kiterjesztjük az ISA által feltárt anyagot azzal, hogy más típusú dokumentumokat is felveszünk (az ISA-ban jelenleg feldolgozott anyag folyóiratokra és konferenciakötetekre korlátozódik). Reméljük, hogy az irodalom mennyisége kellő mértékben fog növekedni ahhoz, hogy lehetővé tegye 2001-ben a korábbi, havonkénti megjelenéshez való visszatérést.

Mivel az információtudomány területe növekszik, s olyan új területeket is tartalmaz, mint a tudásmenedzsment, az ISA által feltárt anyag is növekedni fog ezeknek az új területeknek a felvételével.

Egyetlen dolog van, amit *nem* fogunk csinálni, nem veszélyeztetjük a minőséget azzal, hogy „töltelék” tételekkel egészítjük ki, vagy hamis módon megnöveljük az ISA által feldolgozott anyagot olyan szakterületekkel, amelyek nem kapcsolódnak az információtudományhoz. Teljes mértékben tudatában vagyunk annak, hogy az adatbázis minősége rendkívül fontos az adatbázisok használói számára, és a frissesség, bár fontos, de csak az egyik szempont, amit a keresők figyelembe vesznek, amikor eldöntik, hogy melyik, az információs igényeiknek megfelelő adatbázist használják. Tisztában vagyunk az ISA örökségével és azzal a pozícióval, amelyet elsődleges adatbázisként élvezett az információtudósok között. Mi ezért különösen kérjük ennek a szerkesztőségi cikknek az olvasóit, ha vannak megjegyzéseik vagy javaslataik az ISA-hoz kapcsolódó bármelyik kérdéssel kapcsolatban, vegyék fel velünk a kapcsolatot. Nagyon várjuk az ilyen megjegyzéseket és javaslatokat. A használók visszajelzései messze a hozzájárulás legértékesebb formái az adatbázis szerkesztői számára.

Donald T. Hawkins, főszerkesztő

Lynn A. Murray, szerkesztő

We have discussed the problem of currency in a previous Editorial, and in this final issue of 1999, we would like to revisit the issue and present our plans for ISA in 2000. As Editors, we have become increasingly concerned about the false impression of poor currency that ISA's updating dates give to our users. The reason that ISA appears to be delinquent in publication is purely because of the slow receipt of journals from the publishers, not because of any backlog in database production. Indeed, ISA has no backlog; journals are entered into the next issue of the database following their receipt. During the summer of 1999, the problem became especially critical; the receipt rate of journals fell to approximately half its normal rate. We are unaware of any special circumstances which would explain this.

We are addressing the problem described above by instituting the following measures:

- This is the final issue of 1999, and it is dated November/December.
- In 2000, ISA will be published nine times, at approximately six week intervals. The issue numbers will not correspond to a specific publication date or month. The next issue you receive will be designated "Volume 35, Number 1, 2000". The online database will follow a similar updating schedule. The CD-ROM database will continue to be published at quarterly intervals.
- We are planning a significant expansion in ISA's coverage by including other types of literature (ISA is currently comprised totally of journals and conference proceedings). We hope that the volume of literature will grow enough to support resumption of the previous monthly publication schedule in 2001.
- As the field of information science grows to encompass new areas, such as knowledge management, ISA's coverage will also grow as these new areas are included.

One thing we will not do is to compromise quality by adding "filler" items or by falsely expanding the coverage of ISA into subject areas unrelated to information science. We are acutely aware that database quality is of paramount importance to the users of databases, and currency, albeit significant, is only one consideration that searchers make when deciding which database to use to meet their information needs. We are conscious of the heritage of ISA and the position that it has enjoyed as a premier database among information scientists. We therefore especially invite readers of this Editorial who have comments or suggestions on any of the issues pertaining to ISA to contact us. Such comments and suggestions are most welcome. User feedback is by far the most valuable form of input to database Editors.

Donald T. Hawkins, Editor-in-Chief

Lynn A. Murray, Editor

5. 63b. ábra: A megközelítőleg hathetes aktualizálási időköz ígérete

Amikor több adatbázisban végzett keresést kezdünk, a rövid fájlazonosító címke utal az ISA adatbázis időbeli lemaradásának problémájára (5.63.c ábra). Az optimista ember azt remélheti, hogy csupán a DIALOG felejtette el aktualizálni a címkét (különben ennek automatikusnak kellene lennie). Sajnos a fájl címkéjén szereplő aktualizálási dátum helyesen figyelmeztette a használót a problémára, ahogy az a ZD indexben is látható, mivel az a három könyvtár- és információtudományi adatbázis tényleges aktualizálási dátumát tükrözi (5.63.d ábra).

Dialog Response	
ABI/INFORM(R) 1971-	2000/May 25
(c) 2000 Bell & Howell	
Social SciSearch(R) 1972-	2000/May W3
(c) 2000 Inst for Sci Info	
Gale Group Trade & Industry DB 1976-	2000/May 25
(c)2000 The Gale Group	
LISA(LIBRARY&INFOSCI) 1969-	2000/Apr
(c) 2000 Reed Reference Publishing	
Library Literature 1984-	2000/Apr
(c) 2000 The HW Wilson Co	
Internet & Personal Comp. Abs. 1981-	2000/May
(c) 2000 Info. Today Inc.	
Information Science Abs. 1966-	2000/Jan
(c) Information Today, Inc	

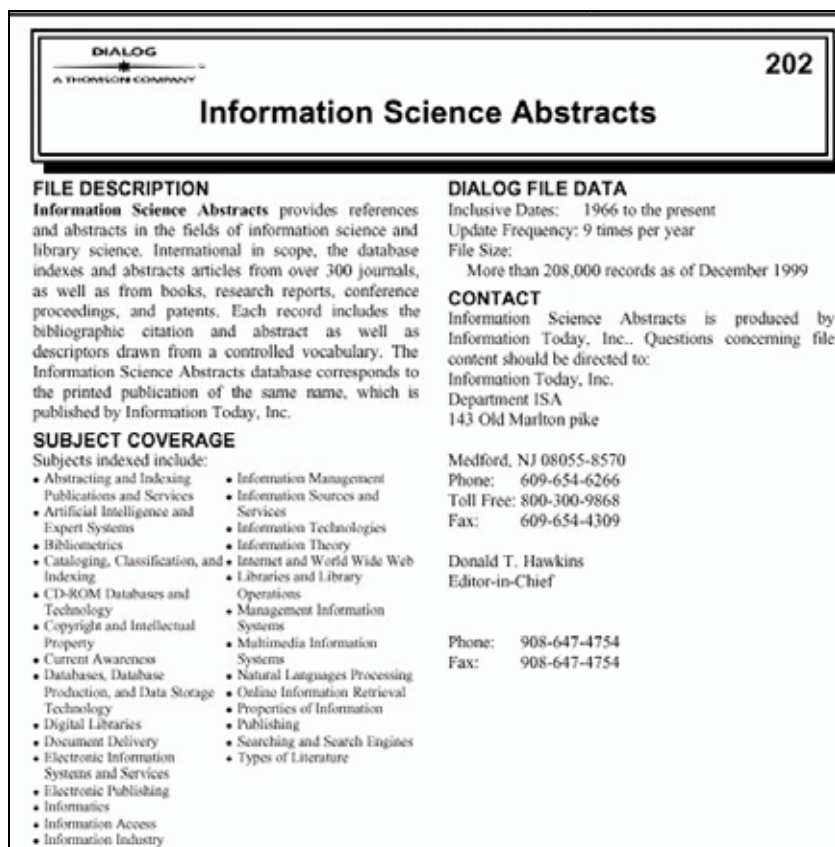
5. 63c. ábra: Az aktualizálási címkék jelzik a frissesség problémáját

Dialog Response			
File	Items	Total	Index-term
61		1708	ZD=19991207
438		1681	ZD=19991218
-----		0	*ZD=2000
61		1152	ZD=20000106
438		1338	ZD=20000119
202		1374	ZD=20000201
61		1018	ZD=20000204
438		2434	ZD=20000301
61		1116	ZD=20000307
61		1520	ZD=20000407
438		2531	ZD=20000425
438		3312	ZD=20000426
438		2350	ZD=20000510
61		1018	ZD=20000516

5. 63d. ábra: 2000 első félévének tényleges aktualizálási időközei

A DIALOG-nak az adatbázis-leírások kinyomtatásával és terjesztésével kapcsolatos hagyományos gyakorlatának fényében különösen érdekes az, hogy amikor az adatbázis-leírásnak a feltárt szakterületre vonatkozó része jelentősen megváltozott azt követően, hogy két cikk bemutatta azt a diszharmóniát, amely a beígért feltárandó szakterület és szakterület fontos folyóiratai feltártságának rendkívüli mértékű csökkenése közt volt. Ez nem volt elegendő ok ahhoz, hogy egy új adatbázis-leírást kinyomtassanak és azt a *Chronolog Newsletter*rel terjesszék.

Mi számít jelentős változásnak? A 22 korábban felsorolt szakterület közül hetet minden ceremónia nélkül eltüntettek, s újakkal egészítették ki (5.64. ábra). Ez jelentős változás. Melyik szakterületeket vették le a listáról? Számítógép-tudomány, döntéstámogatás, oktatástechnika, információs rendszerek tervezése, orvosi informatika, optikai és lézertechnika, szabványok, szuperszámítógépek, távközlés. Közülük sok pontosan az volt, amelyet két hónappal azelőtt két cikkben (Jacsó, 2000b és Jacsó 2000d) használtunk példaként annak illusztrálására, hogy néhány elsődleges forrás és az adatbázis-leírásban kiemelt szakterület - számítógép-tudomány, szuperszámítógépek, döntéstámogatás és orvosi informatika - feltárt-sága milyen drámai mértékben csökkent vagy került mélypontra azóta, hogy az ISA adatbázis tulajdonosa megváltozott.



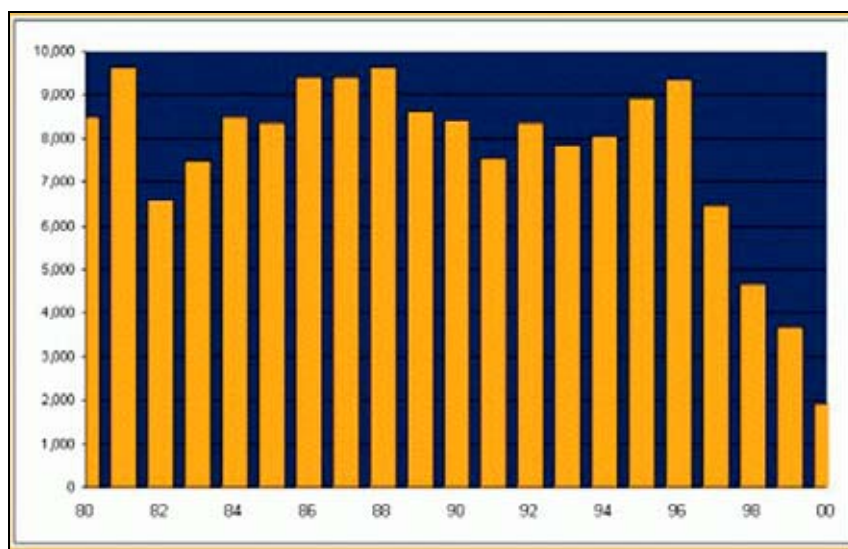
5.64. ábra: Az ISA korábbi szakterületeinek csaknem egyharmadát eltüntették az ISA adatbázis-leírásának 2001 februári változatában.

Az adatbázis-leírás négy egymás utáni számában kiemelt szakterületek csaknem 30%-ának eltüntetése nem olyasmi, amit könnyedén lehetne fogadni. Ha megfelelő felvilágosítást kapnak, a használók - különösen információtudósok és kutatók, és általában azok, akiket különösen érdekelnek azok a témák (és forrásaik), amelyeket eltüntettek vagy amelyeknek a feldolgozását drasztikusan csökkentették - úgy dönthetnek, hogy inkább más adatbázisokat használnak, mint például a kiváló INSPEC adatbázist, amely sokkal jobban (ha nem is tökéletesen) tárja fel az információtudomány és információs rendszerek legfontosabb forrásait, s amely sokkal kurrensebb, mint az ISA, nagyon jó tezaurusza van, míg ára csaknem azonos az ISA-éval.

Még sokkal jobban motiválhatja őket a váltásra az, ha a főszerkesztő kijelenti, hogy sok információtudományi folyóirat, amelyet jelentős mértékben (ha nem is következetesen) feltártak az elmúlt évtizedek során, az információtudomány peremére sorolhatónak tekinthető, vagy csak alkalmanként tartalmaz az információtudósok számára fontos cikkeket, s ezért csak

minimális mértékben kell feltárni őket. Ha ilyen döntést hoznak, akkor azt nem az *Information World Review*-ban, egy az Egyesült Királyságban megjelenő folyóiratban kellene közölni, amely nem sok példányban jut el az Egyesült Államokba, hanem egy amerikai forrásban, mivel az ISA használói bázisának döntő többsége az USA-ban található. Mostanában maga a DIALOG rendkívüli mértékben igazodott az ISA vezetőségének kéréseihez és bejelentéseihez egyéb ügyekben, így a feltárt anyagban bekövetkező ilyen mértékű változás esetében az egyetlen megfelelő mód az lenne, ha elsődlegesen a DIALOG kiadványaiban és hirdetményeiben közölnék. Még az ISA házi orgánuma sem közölt semmi információt arról, hogy ezeket a területeket törölték az ISA-ból.

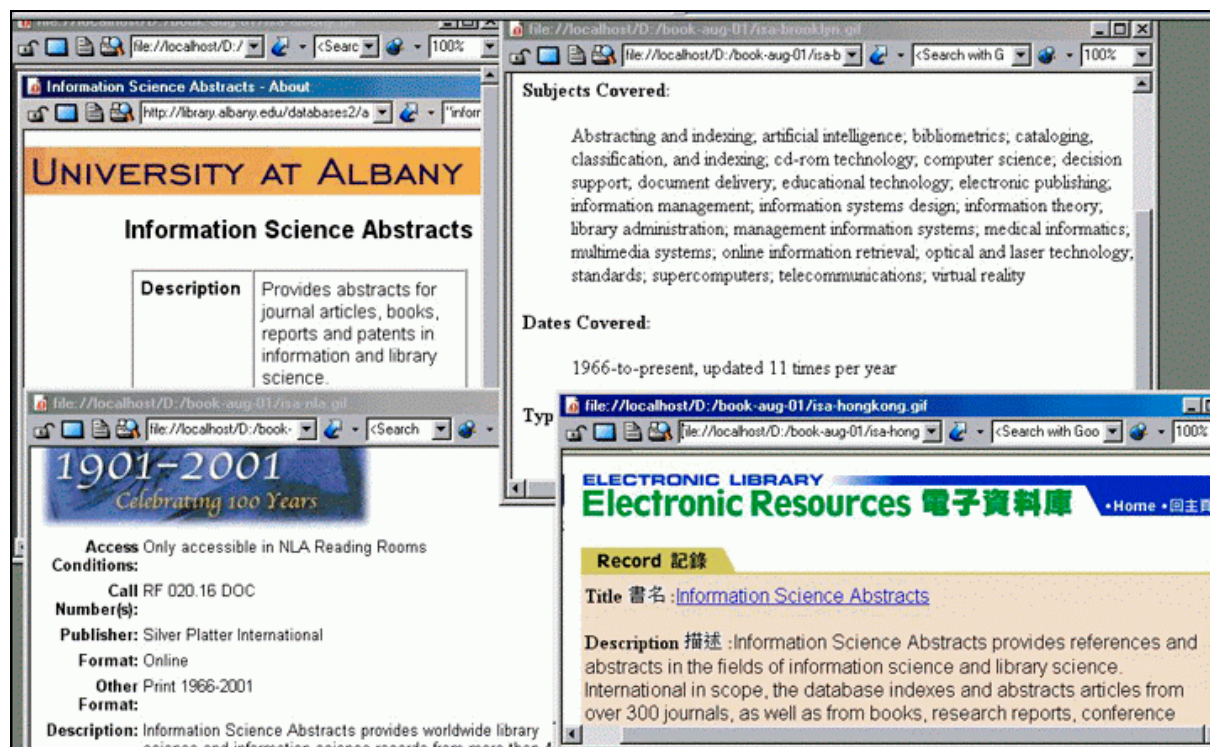
Fontosabb kérdés az ISA adatbázis valódi és hirdetett szakterülete, amely először az *Online Information Review* oldalain folyó vitában merült fel, amely eredetileg az ISA adatbázis csekély növekedési ütemével és frissességének hiányával foglalkozott, összehasonlítva a LISA-val és a Library Literature-rel. Az ISA főszerkesztője azt írta, hogy „az ISA az információtudományt dolgozza fel, amely a legkisebb diszciplína (az irodalom méretét tekintve) azok közül, amelyeket Jacsó összehasonlított, így tehát nem meglepő, hogy adatbázisa, az ISA adja a legkevesebb rekordot. Az, hogy egyszerűen összehasonlítjuk az új rekordok teljes számát számos nem azonos ligába tartozó [sic!] adatbázisban, az olyan mintha almákat és körteket hasonlítanánk össze, ahogy Jacsó csinálta, figyelmen kívül hagyva az egyes szakterületek szakirodalma méretének különbségeit. Jacsó megállapítja, hogy az ISA „könyvtár- és információtudományi adatbázis”, de tudni kell, hogy az ISA információtudományi adatbázis” (Hawkins és Murray 200d). Azt látva, hogy a három adatbázist nem összemérhetőnek nyilvánítják, könnyen azzal magyarázhatják a tapasztalt gyakorló szakemberek és a könyvtár- és információtudományi adatbázisok szakértői, hogy ez kétségbeesett próbálkozás az ISA-ban feldolgozott könyvtártudományi anyag letagadására annak érdekében, hogy megmagyarázzák az ISA-hoz évenként adott rekordok számának a tulajdonosváltás óta tapasztalható rendkívüli mértékű csökkenését. (5.65. ábra).



5.65. ábra: Az ISA adatbázis növekedésének csökkenő üteme a tulajdonosváltás óta

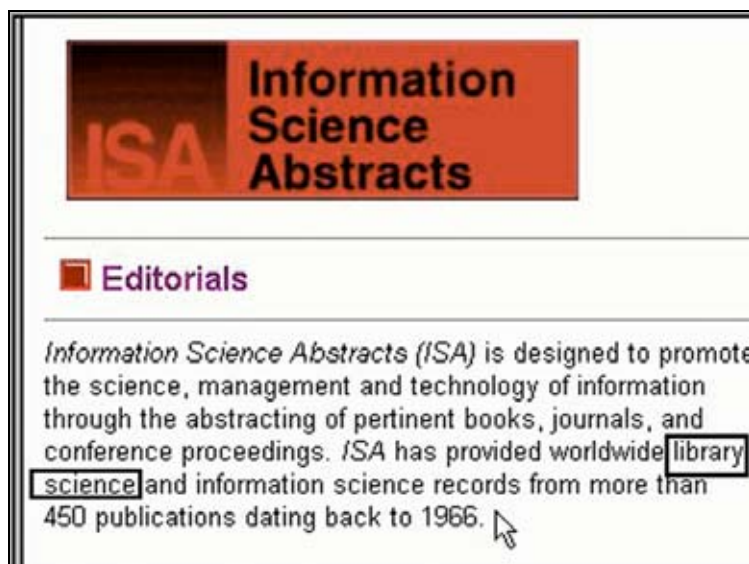
A LISA és a LibLit összehasonlításának „almák és körtek” összehasonlításaként való jellemzése klasszikus hírmagyarázói manőver arra, hogy elutasítsák az ellenfél munkáját, de ez nem nagyon tetszhet sok információtudósoknak, akik könyvtár- és információtudományi anyagot feldolgozó adatbázisokként évtizedeken keresztül hasonlították össze a LISA-t és az ISA-t, s több mint egy évtizede a LibLitet is. A használók furcsának találhatják a könyvtártudományi anyag feltárásának letagadását - mert nem csak a két kereskedelmi adatbázis-

szolgáltató, a DIALOG és a SilverPlatter, hanem az ISA-ra előfizető főiskolai könyvtárak is azt állítják róla, hogy könyvtár- és információtudományi adatbázis - néha különböző szörendet vagy frazeológiát alkalmazva (például könyvtárosság, librarianship), de sohasem hagyva ki a könyvtártudományt vagy annak valamilyen szinonimáját (5.66. ábra).



5.66. ábra: Kollázs az ISA adatbázisának főiskolai weboldalakon található leírásairól

Maga az ISA honlapja (5.67. ábra) szintén elismeri, hogy „az ISA 1966-ig visszamenőleg több mint 450 kiadványt dolgozott fel a könyvtár- és információtudomány világirodalmából.” A JASIS-ben megjelent cikkkel (Hawkins 2001a) egy újabb erőfeszítést tett arra, hogy erősítse az adatbázis információtudományi fókuszát, tagadva azt, hogy feldolgoz olyan könyvtártudományi témákat, mint a levéltárak, könyvtári szervezet, könyvtárhasználat és -használók, mikroformátumok, múzeumok és nem nyomtatott dokumentumok. Ez azonban nyilvánvalóan nem igaz, ahogy azt a mintakeresések példái mutatják (5.68. ábra). Nem foglalkoztunk olyan témákkal, mint a régi és ritka dokumentumok, selejtezés, bútorzat, a könyvtári személyzet típusa és egyéb járművek, amelyeket a cikk nagylelkűen a LISA kizárólagos területének tekint (kihagyva azonban olyanokat, mint a porvédelem, a rozsdá megelőzése és féregirtás).



5.67. ábra: Utalás az ISA által feldolgozott anyagra az ISA honlapján

Searched terms	All years						1997-2001					
	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA	LISA	ISA
	basic index		TI	TI	DE	DE	All	All	TI	TI	DE	DE
archives	7,526	3,263	2,278	683	5,198	1,690	1,794	527	473	75	1,230	278
library organi?ation	317	222	64	29	3	38	49	23	6	1	1	0
library (use or users)	3,212	1,447	515	244	41	130	648	227	74	27	10	0
library and loan?	4,590	1,924	2,468	1,035	2,152	1,487	522	220	271	85	252	177
museums	1,298	1,964	203	97	559	439	408	234	59	17	192	81
(nonprint or non()print) materials	151	207	33	31	32	57	24	4	3	1	11	0

5.68. ábra: A JASIS-ben megjelent cikkben kitagadott könyvtári témák feldolgozottsága az ISA-ban

Az ISA feldolgozza a cikkben felsorolt könyvtártudományi témákat, sokkal nagyobb mértékben, mint bizonyos információtudományi témákat, amelyeket sok éven keresztül fontosként soroltak fel az adatbázis-leíráson, mint például a multimédia információs rendszerek. Ez a kereső-kifejezés összesen két rekordot hoz elő az egész adatbázisból (5.69. ábra). A könyvtártudomány kizárása az ISA által feltárt anyagból nincs összhangban a legújabb megjelent adatbázis-leírással sem, amelyből töröltek hét témát, de megmaradt a dokumentumellátás vagy könyvtári adminisztráció, míg az újonnan felvett témák között is szerepelnek a könyvtárakra és a könyvtárak működésére vonatkozók.

Set	Term Searched	Items	
S1	MULTIMEDIAQINFORMATIONQ SYSTEM?/2001	0	Display
S2	MULTIMEDIAQINFORMATIONQ SYSTEM?/2000	0	Display
S3	MULTIMEDIAQINFORMATIONQ SYSTEM?/1999	2	Display
S4	MULTIMEDIAQINFORMATIONQ SYSTEM?/1998	0	Display
S5	MULTIMEDIAQINFORMATIONQ SYSTEM?/1997	0	Display

Format:

Number of Records:

5.69. ábra: Az egyik, az ISA adatbázis-leírásában kiemelten említett téma (multimédia információs rendszerek) feldolgozottsága

Az ISA szakterületének ez a látszólagos változása, miszerint kizárólag az információtudománnyal foglalkoznak, sértheti az adatbázis üzleti érdekeit, mert a könyvtárosok elhagyatva érezhetik magukat (ha elfogadják az ISA-nak azt az állítását, hogy csak az információtudományt tárják fel, a könyvtártudományt nem), s lemondhatják az ISA előfizetését, s helyette a LISA vagy a Library Literature Information Science Full Text adatbázist rendelhetik meg. Ez elkéséríthet olyan üzleti partnereket, mint az ISA termék menedzsere a SilverPlatternél, aki a szerkesztőhöz küldött levelében azt az erős meggyőződését fejezte ki, hogy „A SilverPlatter kiváló forrást ajánl a könyvtártudományi kutatáshoz” (Hern 2001). Ha az ISA valóban kizárólagosan információtudományi adatbázis lenne, s nem könyvtár- és információtudományi adatbázis, akkor aligha lehetne olyan kiváló forrás a könyvtártudományi kutatáshoz, és a SilverPlatternek a Hern által említett intenzív együttműködése az ISA szerkesztőivel nem lenne a legjövedelmezőbb befektetés.

Egy-két egyszerű teszt is csökkentheti ezeket az aggodalmakat. Ha megvizsgáljuk azt, hogy az ISA hogyan dolgozta fel a hagyományos könyvtártudományi folyóiratokat az elmúlt három vagy öt évben, akkor azt láthatjuk, hogy az ISA nem szüntette meg vagy csökkentette rendkívüli mértékben azon, elsődlegesen a könyvtártudományhoz kapcsolódó folyóiratok feldolgozását, amelyeket az 5.70. ábrán felsorolunk. Ténylegesen szinte mind az ISA-ban a legkövetkezetesebben feldolgozott folyóiratok közé tartozik, sokkal inkább, mint sok minőségi információtudományi folyóirat.

ACQUISITIONS LIBRARIAN
AGAINST THE GRAIN
AMERICAN LIBRARIES
ART LIBRARIES JOURNAL
BEHAVIORAL AND SOCIAL SCIENCES LIBRARIAN
CATALOGING & CLASSIFICATION QUARTERLY
COLORADO LIBRARIES
IFLA JOURNAL
JOURNAL OF LIBRARY ADMINISTRATION
KÖNYVTÁRI FIGYELO (LIBRARY REVIEW)
LIBRARIAN'S WORLD
LIBRARIES & CULTURE
LIBRARY & ARCHIVAL SECURITY
LIBRARY ACQUISITIONS: PRACTICE AND THEORY
LIBRARY ADMINISTRATION & MANAGEMENT
LIBRARY COLLECTIONS, ACQUISITIONS, & TECHNICAL
LIBRARY JOURNAL
LIBRARY MANAGEMENT
LIBRARY PHILOSOPHY AND PRACTICE
LIBRARY RESOURCES & TECHNICAL SERVICES
LIBRARY REVIEW
LIBRARY TRENDS
LIBRI
NEW LIBRARY WORLD
PUBLIC LIBRARY QUARTERLY
RECORDS MANAGEMENT JOURNAL
SERIALS LIBRARIAN
SERIALS REVIEW
TECHNICAL SERVICES QUARTERLY
TUDOMANYOS ES MUSZAKI TAJEKOZTATAS
VJESNIK BIBLIOTEKARA HRVATSKE

5.70. ábra: Elsődlegesen a könyvtártudományhoz kapcsolódó folyóiratok, amelyeket továbbra is referálnak és indexelnek az ISA-ban

Az ISA-nak a könyvtártudomány iránti elkötelezettségét jelezheti az is, hogy olyan idegen nyelveken megjelenő folyóiratok, amelyeket nem sokan beszélnek az ISA elsődleges célközönségéből, kiválasztottak vagy újra kiválasztottak az ISA 1998-as tulajdonosváltása után. A lista utolsó két tételére különösen érdemes figyelni a feltárt forrás földrajzi és nyelvi szempontjából. Ezek nem fogják képessé tenni az ISA-t, hogy versenyezzen a világ könyvtárügyének témájában, amelyet a JASIS-ben megjelent cikkben a LISA kizárólagos területeként határoztak meg. Azok az ISA előfizetők, akik nem perfekt használói a magyar és a horvát nyelvnek, csodálkozhatnak, hogy ennek a két folyóiratnak a relevanciája valóban magasabb-e, mint a korábban említett folyóiratoké.

Bár a folyóiratok magyar, illetve horvát nyelven íródnak, legalább a magyar folyóiratban található informatív angol nyelvű összefoglalók, amelyek különösebb plusz erőfeszítés nélkül felhasználhatók, szinte azonnal az eredeti szám megjelenése után elérhetők, s ezzel gyors másodlagos szolgáltatást nyújtanak a használói közösség számára (feltételezve, hogy az ISA-nak sikerül megvalósítania a rendszeres aktualizálást). Nem csoda, hogy úgy tűnik, hogy a folyóirat ténylegesen magfolyóirat státuszát élvezzi 1999-től, az ISA korábbi vezetőségének ideje alatt tapasztalt 10 év szünet után. [Úgy érzem, kötelességem elmondani, hogy a magyar folyóirat tanácsadó testületének tagja vagyok, de semmi közöm ahhoz, hogy ennek feldolgozását újra elkezdték az ISA 1998-as megvásárlása után.]

Nyugodtan feltételezhetjük, hogy egy másik magyar folyóirattal, a Könyvtári Figyelővel együtt az ISA elsődleges használói bázisa számára a magyar könyvtárügyi és bizonyos mértékig információtudományi és technikai fejleményeire vonatkozó igények többsége

kielégíthető. Bár az orosz, lengyel és cseh nyelvű könyvtártudományi folyóiratok feldolgozását nem kezdték újra, a horvát nyelvű könyvtártudományi folyóirat friss felvétele megnyugtathatja az ISA használói közösségét, amely bizonyára gyászolta az *Informatologia Yugoslavica* (és folytatása) feldolgozásának megszüntetését. Most örülhetnek annak, hogy a Vjesnik Bibliotekara Hrvatska (magyarul horvát könyvtári hírek) az 1997-ben kezdődött, következetlen feltárás után 2000-től a magfolyóiratoknak megfelelő bánásmódban részesül.

Ennek a potenciálisan a könyvtár- és információtudományi szakemberek érdeklődésének középpontjában álló adatbázisnak a látszólagos szakterületi feltártsága és ténylegesen feltárt forrásai közötti diszharmóniája kétségkívül növekedett az legutóbbi néhány évben. Ez sok olyan angol nyelvű információtudományi folyóirat feldolgozottságának a gyors csökkenésével indult, amelyek az ISA adatbázis-leírásán elsődlegesként felsorolt szakterületeken jelennek meg. A gyakorlatilag szinte teljesen feltáratlanul maradt információtudományi folyóiratok között sok olyan van, amelyet az információtudósok a rájuk vonatkozó hivatkozások és (ebből kikövetkezhetően) olvasottságuk alapján az elsők közé rangsorolnak.

Az adatbázis főszerkesztőjének legutóbbi állításai két szempontból további zavaró tónust adnak ehhez a diszharmóniához. Az egyik az, hogy nem veszi észre, a császár ruhája egyre kevesebb, s ennek a trendnek a folytatódása után csupán egy fügefalevele maradhat, mert sok szakterületen fő vetélytársa, a LISA adatbázis az elmúlt öt évben messze az ISA elé került. A másik aspektus az ISA-nak a könyvtártudományi területről való furcsa kivonulása.

Ezt még nem tagadták meg háromszor, de a két megtagadás, különösen a második, nagyon erős a JASIS-ben azt is eredményezheti, hogy a könyvtárosok - helytelenül - lemondanak ennek az adatbázisnak a használatáról. Valójában feldolgozza a könyvtártudományt, sok olyan szakterületet, amelyet a JASIS-ben megjelent cikk a LISA-nak tulajdonított (bár nem egyforma mélységgel). Az információtudományi témákhoz kapcsolódó folyóirathoz képest a könyvtártudományi folyóiratokból jobb választék található az ISA-ban, s következetesebben és jobban fel is tárják őket. Furcsa módon azon a területen, amely harmóniát mutatott a feldolgozott szakterületre vonatkozó hivatalos állítások és az annak megfelelő folyóiratbázis között, kakofónia alakulhat ki hamis hangok miatt, szinte önsorsrontó módon.

Az ISA adatbázis megvásárlása után, 1998 júniusában az első szerkesztőségi cikk az év végére ígerte az adatbázis újra betöltését. Három évvel később még mindig nem került sor rá, sok hiányosság megmaradt, s újak is megjelentek. Remélhető, hogy sor kerül majd az újra betöltésre, ezt követően egy aktualizálást teszünk közzé a <http://www2.hawaii.edu/~jacso/isa-faqs> címen, a hosszú ideje beharangozott és hosszú ideje várt változásokat értékelve. (Jelenleg nem elérhető.)

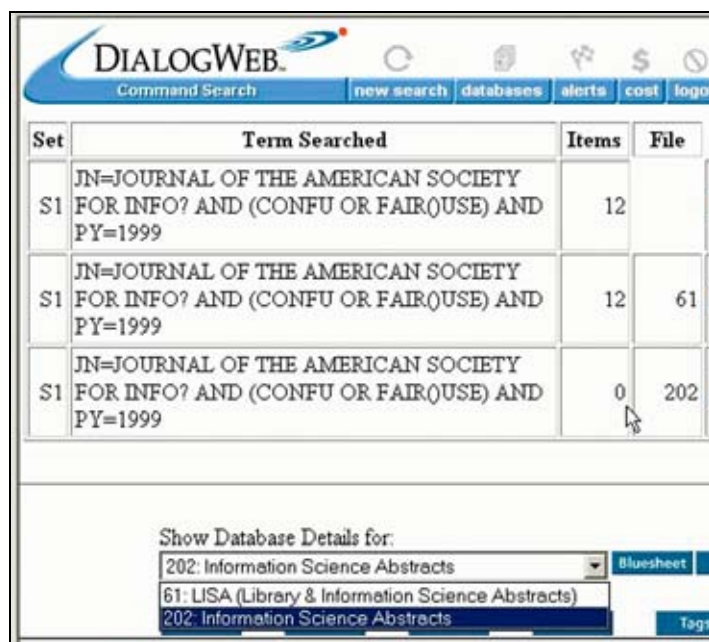
Vannak jelei annak, hogy amikor az adatbázist újra betöltik, mérete jelentős mértékben csökkenni fog. Azt gondoltam, hogy az adatbázis növekedni fog az új rekordok hozzáadásával, amelyekkel pótolják a jelentős hiányokat sok olyan alapvető fontosságú információtudományi folyóiratból, ellensúlyozandó azt a csökkenést, amelyet - remélhetőleg - a duplikátum rekordok ezreinek és az irreleváns szabadalmi rekordok döntő többségének a törlése okoz. Remélhető, hogy az 1998-ig magfolyóiratként megjelölt információtudományi folyóiratok rekordjainak száma nem fog lecsökkenni a folyamat során, egyre inkább sekély feltártságukat kevésbé nyilvánvalóvá téve.

Az is remélhető, hogy a kétségtelenül teljes mértékű feltártságot érdemlő folyóiratok indexelésében és referálásában újonnan keletkezett hiányosságokat is pótolni fogják. Ezek kevésbé nyilvánvalóak, mint a más folyóiratokra vonatkozó drámai változások, de a hiányok kétségkívül léteznek. Vegyük példaként a *Journal of the American Society for Information Science*-t, amelyet évtizedeken keresztül az 1. vagy a 2. helyen rangsoroltak a JCR könyvtár és információtudományi kategóriájában. Egy, az ISA-ban végzett eredménytelen keresés a CONFU néven ismert Conference on Fair Use-ról (Konferencia a szerzői jogot nem sértő

használatról) vezetett 1999-ben annak felfedezésére, hogy az egész szám kimaradt az ISA-ból, az elmúlt néhány év sok más cikkével együtt.

Szerencsére az összes cikket indexeli és referálja a LISA (5.71. ábra), amelyet sok információs szakember használ, mert alacsonyabb az ára, megjósolhatóbb (de nem tökéletes) feltártsága, jó a teaurusza (a Cambridge Scientific Abstracts-féle változatban) és nagyobb frissessége.

Még a LISA sem mindig tud segíteni az ISA indexelő és referáló feltártsága hiányosságain egy másik magasan rangsorolt periodikum, az *Annual Review of Information Science & Technology (ARIST)* esetében, mert azt a LISA, megmagyarázhatatlan módon, egyáltalán nem dolgozza fel. A probléma az, hogy az ISA sem tárja fel megfelelően. Azonnali figyelmeztető jel az, hogy az erre a kiemelkedő kiadványra vonatkozó keresés csupán 94 rekordot eredményezett (5.72. ábra). Bárki, aki csak kis mértékben is ismeri az *ARIST*-ot, tudhatja, hogy 8-10 cikk (valójában fejezet) található minden egyes kötetben. Már a 34. köteténél tart, így a rekordok számának közel kellene lennie a 300-hoz. Ironikus, hogy az *ARIST* évek óta az Information Today, Inc. legértékesebb tulajdona, s ennek ellenére az adatbázis - Information Today, Inc. termék - szerkesztői nem veszik észre kifogásolható feltártságát. Ha van olyan forrás, amelyet teljes mértékben fel kellene tártani az adatbázisban, akkor az *ARIST* olyan.



Set	Term Searched	Items	File
S1	JN=JOURNAL OF THE AMERICAN SOCIETY FOR INFO? AND (CONFU OR FAIRQUSE) AND PY=1999	12	
S1	JN=JOURNAL OF THE AMERICAN SOCIETY FOR INFO? AND (CONFU OR FAIRQUSE) AND PY=1999	12	61
S1	JN=JOURNAL OF THE AMERICAN SOCIETY FOR INFO? AND (CONFU OR FAIRQUSE) AND PY=1999	0	202

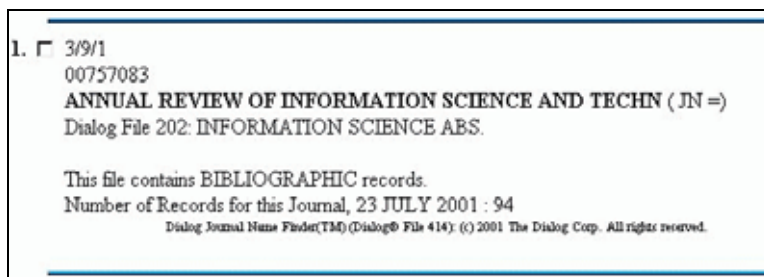
Show Database Details for:

202: Information Science Abstracts [Bluesheet](#)

61: LISA (Library & Information Science Abstracts) [Tags](#)

202: Information Science Abstracts

5.71. ábra: A JASIS-nek egy egész 1999-es száma hiányzik az ISA-ból, ugyanakkor megtalálható a LISA-ban



1. ☐ 3/9/1
00757083
ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHN (JN =)
Dialog File 202: INFORMATION SCIENCE ABS.

This file contains BIBLIOGRAPHIC records.
Number of Records for this Journal, 23 JULY 2001 : 94
Dialog Journal Name Finder(TM) (Dialog® File 414): (c) 2001 The Dialog Corp. All rights reserved.

5.72. ábra: A magasan rangsorolt *Annual Review of Information Science and Technology (ARIST)* átlag alatti feldolgozottsága az ISA-ban

Az ISA szakterülete és a benne feldolgozott források közötti súlyos diszharmonia fényében az 1999 januári, nagyon örömködő és elégedett szerkesztőségi cikkben említett eredmények egy kicsit idő előttinek látszanak (5.73. ábra).

Az 1998-ban bejelentett új folyóiratok felvétele szemmel láthatóan hirtelen lépés volt, s egyikük sem élt 2001-ig. Az ezen folyóiratok némelyikéből felvett rekordok relevanciája nagyon megkérdőjelezhető az információtudomány szempontjából, különösen akkor, ha a tudományra tesszük a hangsúlyt. A hiányok pótlására vonatkozó erőfeszítések teljesítésére való utalás nem különbözik attól, mint amikor Dewey bejelentette győzelmét Truman felett.

Szemmel láthatóan nem működik a feldolgozott folyóiratok komolyabb ellenőrzése annak biztosítása érdekében, hogy a jövőben ne következzenek be hiányok a feldolgozásban, még olyan prominens folyóirat esetében sem, mint a JASIS, ahol a válogatás nem lehet mentség. Annak a reménye, hogy egy teljesen új kontrollált szótárt és tárgyi struktúrát vezessenek be 1999-ben, csak remény maradt 1999-ben, majd 2000-ben, s utána 2001 első felében. Ezt megvalósítható, ahogy láthattuk, hogy a Bowker-Saur egy éven belül, 1992-ben képes volt rá. Az elkészítéshez szükség lehet némi időre, meg még egy kicsire.

Bár ezek a megjósolt fejlesztések nem valósultak meg, az évente az adatbázishoz adott rekordok száma 1998 óta a felére csökkent. A főszerkesztő megfogalmazta ezzel kapcsolatos aggodalmait az 1999 áprilisi, a frissességgel kapcsolatos szerkesztőségi cikkben, mondván, hogy „a múltban az ISA-hoz adott tételek számát jobban befolyásolta egy numerikus „referálási kvóta”, mint a szakterületi relevancia; az eredmény az volt, hogy sok irreleváns rekord került be az ISA-ba. Ilyen irreleváns rekordok többé nem szerepelnek az ISA-ban, így annak minősége javult. Ennek a fejlődésnek azonban nemkívánatos mellékhatásai is vannak. Mivel kevesebb az új rekord, az ISA havi füzetei nemcsak kisebb terjedelműek, mint a régebbiek, hanem tovább is tart, amíg elegendő anyag összegyűlik az egyes számokhoz.” Azután továbbmegy, s elismeri a kiadott számok késését.

Ahogy az ISA első 1999-es számát készítjük, itt az alkalom arra, hogy reflektáljunk arra, amit a múltban teljesítettünk és terveket készítsünk a jövőbeni javítás érdekében. Az ISA hosszú utat tett meg 1998-ban, különösen mióta az Information Today, Inc. megvásárolta. Néhány fejlesztés, amit már elvégeztünk:

Az adatbázisnak a *Microcomputer Abstracts*ből származó referátumokkal való kiegészítése, amely lehetővé tette számunkra olyan témák feldolgozását, mint az internetes keresőgépek, információküldő szolgáltatások és az információknak a használókhoz való eljuttatásának új módjai,

Számos, releváns cikket tartalmazó új folyóirat feldolgozása, köztük számos nem angol nyelvű

Az információtudomány irodalma szempontjából legfontosabb folyóiratok feldolgozási hiányainak pótlására vonatkozó erőfeszítések megvalósítása

Az időszaki kiadványok ellenőrzésének megszigorítása annak érdekében, hogy a jövőben ne fordulhassanak elő feltárási hiányok, továbbá

az ISA szerkesztőségi testületének felállítása.

Tisztában vagyunk azzal, hogy a munka még semmiképp nem fejeződött be, és sok további lépés van, amit még megtenni tervezünk. Az ISA által feldolgozott folyóiratok körét növelni fogjuk további olyan folyóiratokkal, amelyek az információtudományhoz kapcsolódnak. Azt is tervezzük, hogy teljesen új előállítási rendszert vezetünk be az ISA-nál, amely lehetővé tegye számunkra, hogy felvegyünk és pontosan dolgozzunk fel tételeket a nem-folyóirat irodalomból, ez a referátumok fejlesztett minőségellenőrzését is lehetővé fogja tenni. Tervezzük egy teljesen új kontrollált szótárt és tárgyi struktúrát, s reméljük, hogy ezek 1999-ben meg is valósulnak.

Izgattottak vagyunk ezek miatt a fejlesztések miatt, s biztosak vagyunk benne, hogy az ISA használói értékesnek fogják tartani őket. A minőség-ellenőrzés nagyon fontos számunkra, s mi elköteleztük magunkat az iránt, hogy fenntartsuk az ISA előállításának legmagasabb minőségét. Önök, a használók nagy szerepet játszhatnak ebben a folyamatban, és segíthetnek nekünk visszajelzéseikkel és kommentárjaikkal.

Donald T. Hawkins, főszerkesztő

Lynn A. Murray, szerkesztő

Editorial • January 1999**• About the Editors •**

As we produce the first 1999 issue of *ISA*, it is appropriate to reflect on accomplishments of the past and make plans for future enhancements. *ISA* has come a long way in 1998, especially since its acquisition by Information Today, Inc. Some of the improvements we have made are:

- Addition of abstracts from *Microcomputer Abstracts*, which has given us coverage of topics such as Internet search engines, "push" services, and new ways of delivering information to users,
- ▪ Addition of many new journals containing relevant articles, including several in non-English languages,
- ▪ Completion of the effort to fill gaps in coverage of journals that are important to the literature of information science,
- ▪ Institution of tighter serial control to ensure that coverage gaps will not occur in the future, and
- Appointment of an *ISA* Editorial Board.

We are conscious that the work is by no means finished, and there are many more significant steps that we are planning to take. The journal coverage of *ISA* will be enhanced with the addition of still more journals relevant to information science. We are also planning an entirely new production system for *ISA* which will allow us to add and correctly cite items from the non-journal literature and which will also provide enhanced quality control of the abstracts. A completely new controlled vocabulary and subject structure are in the planning processes, and we hope to roll these out in 1999.

We are excited about these improvements and are confident that the users of *ISA* will find them of value. Quality control is very important to us, and we are committed to maintaining the highest quality in the production of *ISA*. You, the users, can play a large part in this process and help us in this effort by giving us feedback and comments.

Donald T. Hawkins, Editor-in-Chief

Lynn A. Murray, Editor

5.73. ábra: Az ISA 1999 januári, rendkívül optimista szerkesztőségi cikke

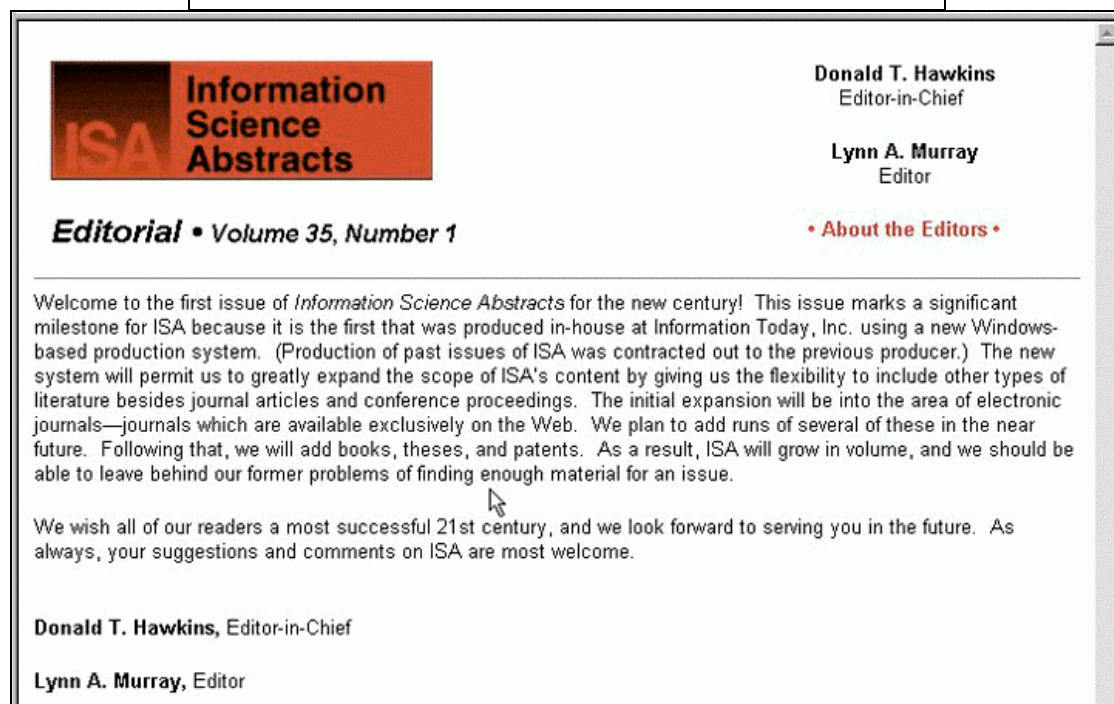
Ezt a pontot később még jobban kidolgozták (Hawkins és Murray 200d) a kritikára adott válaszként: „Jacsónak az a gondolata, hogy az ISA-nak van egy kvótája, nem pontos. Az igaz, hogy a korábbi tulajdonosoknak szerződéses kötelezettsége volt, hogy havonta legalább 800 rekordot adjanak hozzá az adatbázishoz. Ezek a tulajdonosok nehéznek találták, hogy a referátumok készítésének ezt a mértékét fenntartsák (ahogy mi is így lennénk ezzel) az információ-tudomány korlátozott területe miatt, s az a gyakorlatuk alakult ki, hogy „töltelék” rekordokat adtak minden aktualizáláshoz, hogy a hiányt kipótólják. Fontos megjegyezni, hogy a legtöbb „töltelék” rekord nem volt releváns az információ-tudomány szakterületével, ezért az adatbázis minősége gyenge volt.”

„Üdvözljük az *Information Science Abstracts* első számában az új évszázadban. Ez a szám fontos mérföldkő az ISA számára, mert ez az első olyan, amelyet házon belül, az Information Today, Inc.-nél készült, az új Windows-alapú előállítási rendszert használva. (A ISA korábbi számainak készítése szerződés alapján a korábbi készítőnek lett kiadva.) Az új rendszer lehetővé teszi számunkra, hogy nagy mértékben kiterjesszük az ISA által feldolgozott tartalmat, amely megadja nekünk azt a rugalmasságot, hogy a folyóiratcikkek és a konferenciakötetek mellett más típusú irodalmat is felvegyünk. Először az elektronikus folyóiratokra terjed ki ez, - olyan folyóiratokra, amelyek kizárólag a weben érhetők el. Azt tervezzük, hogy a közeljövőben számos ilyent dolgozunk fel. Ezt követően könyveket, szakdolgozatokat és szabadalmi leírásokat is fel fogunk venni. Ennek eredményeként nőni fog az ISA terjedelme, s magunk mögött hagyhatjuk azt a régi problémát, hogy elegendő anyagot találjunk egy szám számára.

Minden olvasónknak a lehető legsikeresebb 21. századot kívánunk, s örömmel szolgáljuk önöket a jövőben. Mint mindig, az ISA-ra vonatkozó javaslataikat és kommentárjaikat most is nagyon várjuk.

Donald T. Hawkins, főszerkesztő

Lynn A. Murray, szerkesztő



Information Science Abstracts

Donald T. Hawkins
Editor-in-Chief

Lynn A. Murray
Editor

Editorial • Volume 35, Number 1

[• About the Editors •](#)

Welcome to the first issue of *Information Science Abstracts* for the new century! This issue marks a significant milestone for ISA because it is the first that was produced in-house at Information Today, Inc. using a new Windows-based production system. (Production of past issues of ISA was contracted out to the previous producer.) The new system will permit us to greatly expand the scope of ISA's content by giving us the flexibility to include other types of literature besides journal articles and conference proceedings. The initial expansion will be into the area of electronic journals—journals which are available exclusively on the Web. We plan to add runs of several of these in the near future. Following that, we will add books, theses, and patents. As a result, ISA will grow in volume, and we should be able to leave behind our former problems of finding enough material for an issue.

We wish all of our readers a most successful 21st century, and we look forward to serving you in the future. As always, your suggestions and comments on ISA are most welcome.

Donald T. Hawkins, Editor-in-Chief

Lynn A. Murray, Editor

5.74. ábra: Az ISA 1999 áprilisi szerkesztőségi cikke annak okáról, miért csökkent a referátumok száma

Két nagyon zavaros pont van ebben az egyetlen bekezdésben. Az egyik a tulajdonosokra való utalás. A Documentation Abstracts, Inc. (DAI) volt az adatbázis tulajdonosa, amelyet hét könyvtár- és információtudományi egyesület képviselői segítettek. Ezeknek az embereknek nem volt kötelezettségük arra, hogy fenntartsák a referátumkészítés mértékét, egyszerűen azért, mert nekik egyáltalán nem kellett semmilyen referátumot készíteniük. A fájl előállítója,

a Plenum végezte ezt és a DAI fizetett érte, és a pénz a rangidős technikai szerkesztőhöz jutott, akinek többek között a kiadást kellett felügyelnie. Ezért az elődök gyakorlatára való utalás furcsa ujjal mutogatás. Az adatbázis gyenge minőségére való utalás helyes.

Nehéz megérteni, hogy nem lehet elég forrásdokumentumot használni az ISA adatbázis szakterületeként meghatározott témákban. Az elmúlt néhány évben hullámvész volt megfigyelhető azoknak a számoknak a kiadásában. Csak az ISA által figyelt folyóiratokból fel nem vett releváns cikkek száma is könnyen több százra tehető. Ezen felül vannak további folyóiratok és konferenciakötetek, amelyeket az ISA nem dolgoz fel (mint például az ACM és az IEEE digitális könyvtárakkal foglalkozó konferenciái, amelyek bőven enyhíthetik a főszerkesztő azzal kapcsolatos aggodalmait, hogy hiányzik a releváns anyag az adatbázis számára). Ha a szerkesztők figyelmüket a fontos kérdésekre összpontosítanák az adatbázis túlélése érdekében, s pótolnák a múltbeli és újabb hiányokat az adatbázis megtisztítása során, ez csökkentené az adatbázis szakterülete és feltárt forrásai közötti diszharmóniát.

Másik oldalról, ha az energiákat túlságosan az adatbázis felpörgetésére fordítják, az kisodorhatja őket a játéktérrel, ahol az INSPEC már jobban feldolgozza az információtudományt; a LISA a könyvtártudományt; a Library Literature a teljes szövegű könyvtár- és információ-tudományt; és az Internet & Personal Computing Abstracts az információs technikát. Ez figyelmeztető jel, és az ISA halálának legfőbb oka az ISA adatbázis állítólagos szakterülete és a ténylegesen feldolgozott források közötti diszharmónia lesz, továbbá ennek hosszú időn át való tagadása.

6. A rekord tartalma

Az adatbázisok jelentős mértékben különböznek egymástól abból a szempontból, hogy mennyi információt szolgáltatnak a forrásdokumentumokról, bemutatott személyekről vagy tárgyakról. A megjelenített információ struktúráltága szempontjából is nagy különbségek lehetnek. Mindezeknek hatása lesz arra, hogy a tartalom hogyan indexelhető, kereshető vagy jeleníthető meg a képernyőn vagy nyomtatásban. A legfontosabb az, hogy ez lesz a legfőbb oka a végeredmény különbségeinek, annak, amit a használó kap.

Bibliográfiai adatelemek

Azoknak az adatelemeknek a köre, amelyeknek szerepelniük kell egy rekordban, nagyban függ az adatbázistól. Az alapvető bibliográfiai adatokon túl (szerző, cím, folyóiratcím, kronológiai vagy numerikus adatok, oldalszám) a kiegészítő adatelemek azok, amelyek leginkább megkülönböztetik az adatbázisokat. Ezeket az adatelemeket értéknövelt információknak nevezzük, s olyan elemek tartoznak közéjük, amelyek növelik a korábbiakban említett bibliográfiai adatokat, például a következőkkel:

- Címkiegészítés, amely megmagyarázza a címet
- Az oldalak, sorok vagy szavak teljes száma
- A grafikonok, fotók és illusztrációk jelzése.

Jól illusztrálja a különbségeket, ha ezeket a bibliográfiai adatokat három bibliográfiai adatbázisban hasonlítjuk össze ugyanarra a cikkekre vonatkozóan. Az oldalszám a ProQuest-ben (6. 1a. ábra) és a sorok és szavak száma a Magazine Database-ben hasznos tájékoztatást ad. A ProQuestben azonban a különböző azonosítók zavaróak és nem kellene szerepelniük egy bibliográfiai rekordban. A Readers' Guide Abstracts (RGA) (6. 1b. ábra) a legvilágosabb a forrás bibliográfiai adatai szempontjából, de más szempontból ez a formátum nem jeleníti meg az RGA legértékesebb tartalmi elemét - a nagyon informatív referátumokat (ezeket később tárgyaljuk meg). A Gale Group Magazine Database-e tartalmazza a legjobb cím-információkat, hosszú alcímmel és informatív címkiegészítéssel (6. 1c. ábra).

A cégek címjegyzékeiben az alapvető elemek közé tartozik a cégek neve, címe, telefon- és faxszámai, e-mail címe (mostanában az URL-je is). Az időszaki kiadványok címjegyzékeiben a minimálisan megkövetelt bibliográfiai adatok közé tartozik a folyóirat címe, korábbi címe, újabb címe, ISSN száma, a kiadás országa, kiadója és indulási éve (megszűnésének éve, ha az is van) (és itt is a kiadónak vagy magának a folyóiratnak az URL-je). Bár ez szoftver kérdése, érdemes megemlíteni, hogy a DIALOG-nak nagy hibája az, hogy nem ajánl fel közvetlenül kattintható ugrópontokat a megjelenített rekordokból azokhoz, amelyekben ugyanaz a szerző, tárgyszó vagy folyóiratcím szerepel, pedig ez általános gyakorlat a legtöbb online információszolgáltatónál. Ennek a funkciónak a hiánya azt is jelenti, hogy a rekordban szereplő URL-ek hidegek (azaz passzívak) a DIALOG-nál. Az Ulrich's Ovid változatában ugyanazon időszaki kiadvány rekordjában olyan közvetlenül kattintható URL van, amely elvisz minket a kiadó vagy a folyóirat weblapjára (6.2. ábra).

New Search by word for publication ProQuest

RESULTS LIST | MARKED LIST | PREFERENCES

Article 51 of 100 BUILD PAGE MAKE LINK PREVIEW LINK MODIFY LINK HELP ?

Burial at sea
Maclean's; Toronto; Aug 2, 1999; [Andrew Phillips](#); [Susan Oh](#);

Volume: 112
Issue: 31
UMI Publication No.: 04403445
Start Page: 26-29
Page Count: 4
Text Word Count: 2258
Document Type: Feature
Source Type: **PERIODICAL**
ISSN: 00249262
Subject Terms: [Death & dying](#)
[Celebrities](#)
[Aircraft accidents & safety](#)
[Memorial services](#)
Personal Names: [Kennedy, John F Jr](#)
[Bessette, Carolyn](#)
[Bessette, Lauren](#)
Companies: [Navy-US](#)
UMI Article Re. No.: GMAC-2178-12
UMI Journal Code: GMAC

Mark article ☐
[Cite/Abstract](#)
[Full Text](#)
[Text+Graphics](#)
[Page Image](#)
[Publisher Info.](#)

6. 1a. ábra: Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (ProQuest)

☐ **Record 1 of 3 in Readers' Guide Abstracts FT 1/91-9/99 Updated Sep 27**
[Full Text HTML](#) or [Full Text PDF](#)

TITLE
Burial at sea

OTHER TITLES
 Augmented title: J. F. Kennedy, Jr.

PERSONAL AUTHOR
[Phillips, Andrew](#)

SOURCE
[Maclean's](#) v. 112 no31 Aug. 2 1999 p. 26-9.

6. 1b. ábra: Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (RGA)

05406400 Supplier Number: 55272717 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Burial At Sea: A massive search locates the bodies of John F. Kennedy Jr., his wife and his sister-in-law.(ceremony from USS Briscoe at recovery site)

PHILLIPS, ANDREW, OH, SUSAN
Maclean's , 26
 August 2 , 1999
Document Type: Biography
ISSN: 0024-9262
Language: English **Record Type:** Fulltext
Word Count: 2289 **Line Count:** 00175
Geographic Codes: 1U1MA Massachusetts
Biographer: Kennedy, John F., Jr.--Rites and ceremonies; Kennedy, Carolyn Bessette--Rites and ceremonies; Bessette, Lauren G.--Rites and ceremonies
File Segment: MI File 47

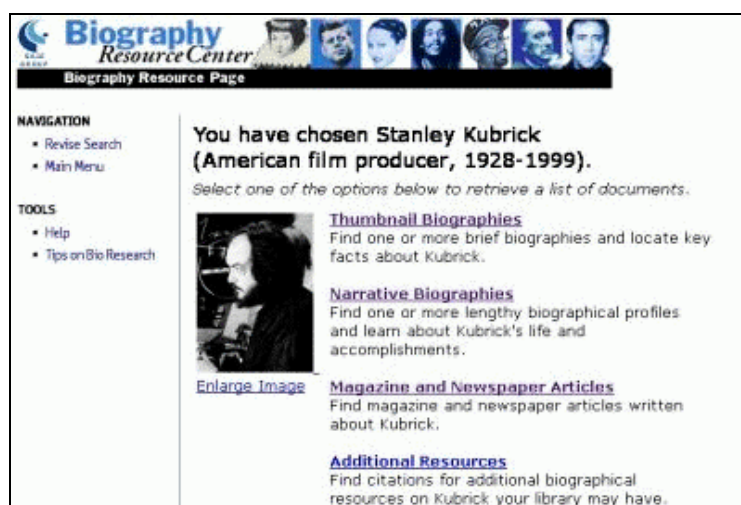
Gale Group Magazine DB(TM) (Using® File 47) (c) 1999 The Gale group. All rights reserved.

6. 1c. ábra: Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (Gale Group Magazine Database)

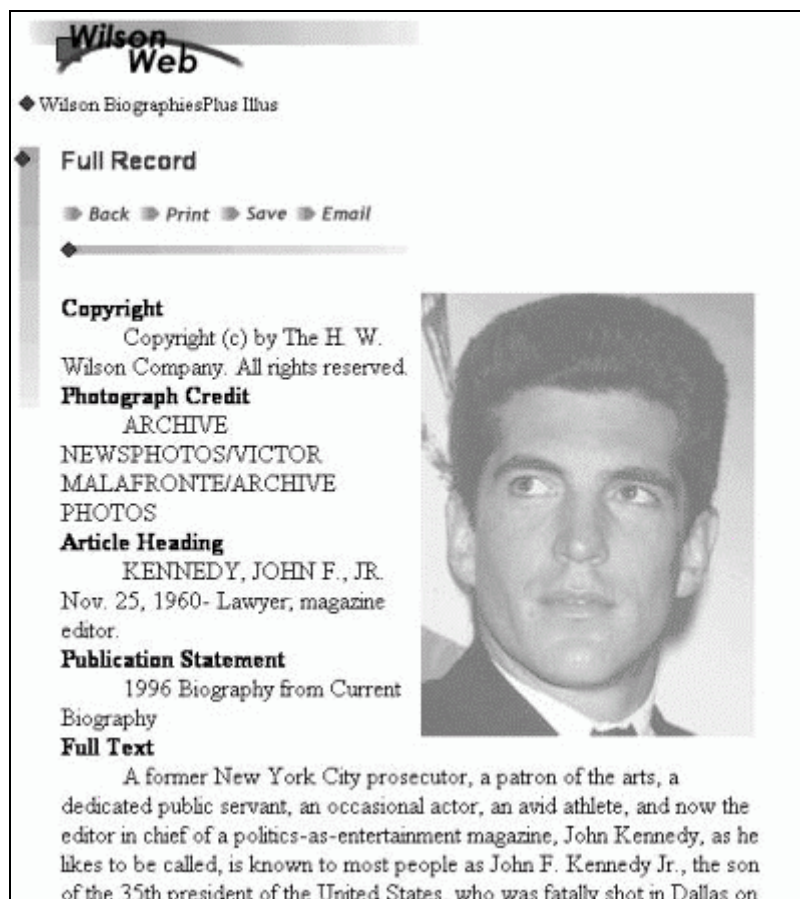
Title	E Media Professional
Alternate Title	Variant title: EMedia Professional. Former titles (until 1997): C D - R O M Professional. ISSN 1049-0833. Incorporates (in 1994): C D - R O M News Extra. ISSN 1075-1106. (until May 1990): Laserdisk Professional. ISSN 0896-4149
Publisher	Online, Inc.
Status Code	Active
Country Code	United States (US)
ISSN	1090-946X
Dewey Decimal Number	025.04
Circulation	17,000 (paid)
Editor	Stephen Nathans
LC Call Number	TA1635
E-mail Address	E-mail: emedia@onlineinc.com ; URL: http://www.emediapro.net

6.2. ábra: Az Ulrich's Ovid változata közvetlenül kattintható ugrópontot szolgáltat a kiadó és a folyóirat weblapjához.

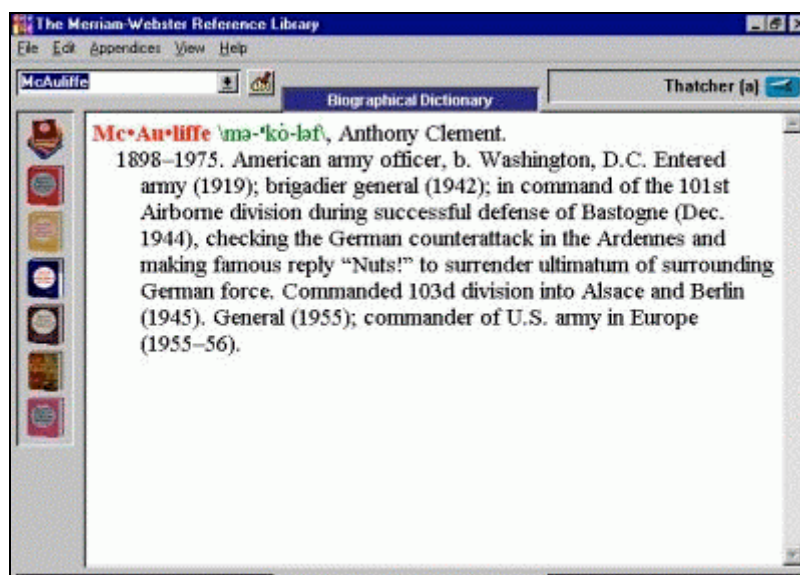
Egy életrajzi gyűjteményben a legalapvetőbb elemek közé tartozik a név, nem, születési és halálozási dátum, foglalkozás vagy szakma, a híresség oka. A 6.3. ábra megmutatja a Gale Group Biography Resource Centerben elérhető kategóriák választékát. Amit a Gale tömör (egy hüvelykujj körmére ráférő) életrajznak hív, az gyakran sokkal több ennél - közel áll egy életrajzi esszéhez. Az H. W. Wilson cég Biographies Plus Illustrated adatbázisából a 6.4. ábra egy részletet mutat ifjabb JFK jól szerkesztett életrajzából. Ezzel a két életrajzi forrással éles ellentétben állnak - elég csak az alapvető életrajzi információkat megnézni - a Merriam-Webster Biographical Dictionary (6.5. ábra) fájdalmasan rendszerezetlen életrajzi tételei. Ezek a nyomtatott kiadás zsúfolt szerkezetét tükrözik, s nem ajánlanak fel semmilyen külön hozzáférési pontot a tárgyként szereplő személy családi nevén kívül. Ez tankönyvbe illő példája annak, hogyan nem szabad digitális változatot készíteni egy nyomtatott referensz-forrásból.



6.3. ábra: Részlet a Gale Group Biography Resource Centerének egy tételéből.



6.4. ábra: Részlet az H. W. Wilson cég Biographies Plus Illustrated adatbázisának egy tételéből.



6.5. ábra: Egy tétel a Merriam-Webster Biographical Dictionaryből.

Értéknövelt adatelemek

Látszólag apró elemek tehetik a keresést sokkal hatékonyabbá, s könnyíthetik meg a legrelevánsabb cikkek kiválasztását az eredmények halmazából. Egyeseket igazán könnyű megadni (mint például a nyelvet vagy a dokumentumtípust), mások több munkát igényelhetnek (mint például a szerző munkahelye, egységesített formában). Meglepő, hogy miért nem használnak egyszerű elemeket rekordok tökéletesítéséhez. Például nehéz megérteni, hogy az Ulrich's International Periodicals Directory miért nem tartalmaz nyelvi mezőt; miért nincs a LISA-ban dokumentumtípus mező (kivéve a folyamatban levő kutatások jelentéseit), hogy el lehessen különíteni a konferencia előadásokat a folyóiratcikkektől és a szemléktől; miért nem szerepel a LISA adatbázisban a kiadás (megjelenés) országa mező; a Merriam-Webster Biographical Dictionary miért nem tartalmazza a nemet, amely gyakran szükséges keresési szempont. Az ilyen adatelemek szerepeltetése a rekordban és indexelése rendkívül egyszerű módja annak, hogy értéket adjunk a rekordokhoz és a keresést sokkal hatékonyabbá tegyük.

A további adatelemek iránti igény az adott adatbázistól függ. Az ember nem várna nyelvi mezőt az ABI/INFORM adatbázisban (de mégis van benne), mert a rekordok 99,99 százaléka angol nyelvű dokumentumokra vonatkozik.

A cikkek hosszúság szerinti osztályozása a UMI Magazine Express adatbázisában (rövid, közepes, hosszú) kiváló eszköz ahhoz, hogy a keresést csak terjedelmes cikkekre korlátozhassuk, elkerülve a rövid cikkek özönét olyan aktuális eseményekről, mint az elnök elleni felelősségrevonási per. A Magazin Express érdemének lehet azt is tekinteni, hogy egyes adatbázisaiban jelzi a folyóirat típusát (akadémiai, a nagyközönségnek szóló stb.), hogy arra lehessen szűkíteni a keresést és orientálni a használót a keresés eredményének áttekintésekor.

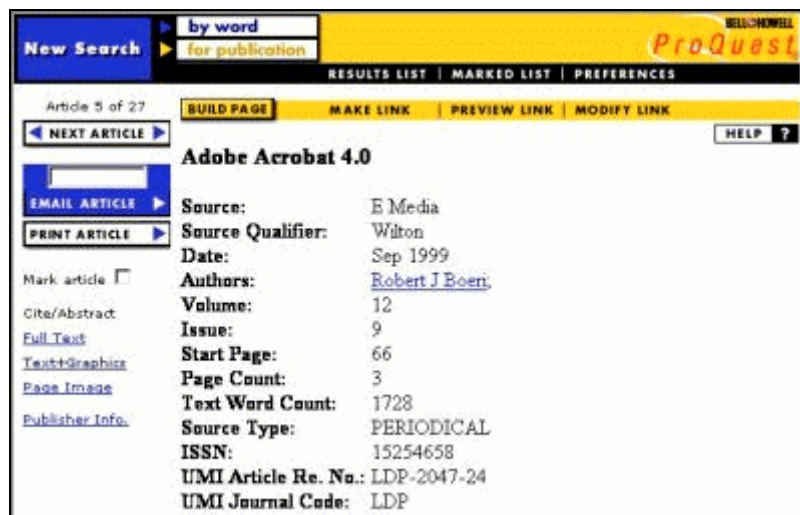
Az EBSCO valami hasonló lehetőséget ajánl, lehetővé téve, hogy a keresést a lektorált folyóiratokra lehessen szűkíteni (ennek a kiegészítő adatelemnek a használatával). (A folyóirat lektorált voltáról szóló információ megbízhatósága más kérdés.) Ahogy azt a 7., a pontossággal foglalkozó fejezetben megtárgyaljuk, ez az információ gyakran pontatlan az EBSCO Serials Directory című adatbázisában (és az Ulrich's adatbázisban is). Nyilvánvaló, hogy ez ellehetetleníti ennek az adatelemnek a szűrőként való használatát. Az H. W. Wilson cég adatbázisainak saját WilsonWEB-es változatában szintén szolgáltat olyan értéknövelő információt, mint a folyóiratok lektoráltsága, és - mint általában - ez pontos is a könyvtár- és információtudomány területén, ahol meg tudtam ítélni. Ez a fajta értéknövelt adatelem egyedi, mert nem minden egyes rekordhoz kapcsolják, csak a folyóirathoz.

Az értéknövelt információ hasznos lehet a böngészés, a keresés és az eredmény megjelenítése során. Bár valószínűtlen, hogy bárki böngészni akarná a folyóiratok ISSN mezőjét egy referáló és indexelő adatbázisban annak érdekében, hogy egy ígéretes ISSN számot fogjon ki, a folyóiratok ISSN számának megjelenítése különösen hasznos olyankor, amikor a használónak könyvtárközi kölcsönzési kérésre van szüksége. Az a mód, ahogy a Bell & Howell nem változtatja meg a folyóirat kódját olyankor, amikor a cím megváltozik, egyszerűsítheti a keresést.

Ahelyett, hogy meghatároznánk egy folyóirat korábbi és mostani címét egy keresésben, mint például FIND JN=(Laserdisk Professional OR CD-ROM Professional OR E Media Professional) a DIALOG-nál, a ProQuestnél a folyóirat kódjának JC (LDP) használatával végzett keresés meg fogja találni a rekordokat a folyóirat címe három verziójának bármelyikéből (6.6. ábra). Az más kérdés, hogy a folyóirat új címe sohasem válthatja fel a korábbi címet a rekordokban visszamenőlegesen, mint ahogy az megtörtént a Proquestben néhány folyóirattal, mint például a *Database*-zel, amely 1999-ben *EContentre* változtatta címét. Helytelen dolog

megjeleníteni rekordokat, mondjuk a folyóirat 1996-os számaiból úgy, hogy a folyóirat címeként *EContent* szerepel.

A könyvkritikák esetében a kiegészítő adatelemek között szerepelhet a bírált könyv ISBN száma, és az az osztályzat is, amelyet az indexelő adott a kritikus kommentárjainak elemzése után. Ez az utóbbi alkalmazható filmek, színházi előadások és kiállítások kritikáira is (6.7. ábra).



6.6. ábra: Állandó folyóiratkódok használata a korábbi és a mostani címek megtalálásához.

```
Adding title to results...
REVIEW SUMMARIES FOR BRIDGET'S DIARY
-----
RANK: S4/1-30   Field: DT=   File(s): 484
(Rank fields found in 30 records -- 4 unique terms)

RANK No.   Items   Term
-----
1           29    PERIODICALS
2           19    BOOK REVIEW-FAVORABLE
3           9     BOOK REVIEW-COMPARATIVE
4           2     BOOK REVIEW-MIXED
---end of results---
```

6.7. ábra: A kritikák osztályzatai a Periodical Abstracts adatbázisban.

A dokumentumtípusok meghatározása gyakran alkalmazott s jó eszköz a használó számára nem releváns dokumentumtípusok, például jelentések vagy könyvkritikák kiszűrésére. Kevésbé általános az, amikor a cikkek között megkülönböztetik az elméleti és a gyakorlati jellegűeket, ahogy az az INSPEC adatbázisban történik, amely a treatment (téma megközelítési mód) mezőt használja (6.8. ábra). Az ERIC hasonló értéknövelt információt használ a feldolgozott dokumentumok célközönségének (target audience) megkülönböztetésére. Ez az adatbázis Ovid változatában kereshető és böngészhető is (6.9. ábra). A két adatbázis egyikében sem érhetők el ezek az értéknövelt információk minden rekordban.

		Citation 1
Title	The value of interdisciplinarity: a study based on the design of Internet search engines.	
Treatment	Experimental.	
Go to ... Complete Reference Help Logoff		
		Citation 2
Title	Constraint propagation applied to citation analysis: an example.	
Treatment	Theoretical or mathematical.	
Go to ... Complete Reference Help Logoff		
		Citation 3
Title	Citation analysis as a tool in journal evaluation.	
Treatment	General or Review.	

6.8. ábra: Az INSPEC adatbázis cikkeinek téma megközelítési mód kódja.

O V I D


Audience Type Index Display

?

Perform Search

A-Z

Back in Index

 Main Search Page

Forward in Index

A-Z

Choose from among the following index entries:

Select	Audience Type	# of Citations
<input type="checkbox"/>	administrators.at.	19287
<input type="checkbox"/>	community.at.	2310
<input type="checkbox"/>	counselors.at.	1893
<input type="checkbox"/>	media staff.at.	2980
<input type="checkbox"/>	parents.at.	3751
<input type="checkbox"/>	policy makers.at.	13954
<input type="checkbox"/>	practitioners.at.	110157
<input type="checkbox"/>	researchers.at.	23644
<input type="checkbox"/>	students.at.	7362
<input type="checkbox"/>	support staff.at.	650

6.9. ábra: Az ERIC adatbázis célközönség indexe az Ovidnál.

Az értéknövelt információk következő szintjét a forrásdokumentumhoz kapcsolt osztályozási jelzetek és tárgyi kifejezések reprezentálják. Ebből a szempontból nagy változatosság van az adatbázisok között, ahogyan azt a 10., a tárgyi indexelés minőségével foglalkozó fejezetben megtárgyaljuk.

Az eredeti dokumentumok referátumai és összefoglalói értéknövelő elemek maradnak akkor is, amikor a dokumentum teljes szövege elérhető. A referátumok szerkezetének, olvashatóságának és informatív jellegének fontos megkülönböztető funkciója van a legtöbb, a nagyközönségnek szóló periodikumokat feldolgozó adatbázisban. Az ilyen típusú dokumentumokat feldolgozó adatbázisok primer forrásaiban nincs szerzői referátum. Ez különbözik a tudományos és akadémiai folyóiratok többségétől, amelyeket a tudományos, műszaki és orvosi adatbázisok dolgoznak fel. Ezeknek a fájloknak a készítői általában szerepeltetik az eredeti referátumokat, s csak alkalmanként módosítják őket, s nem nagyon készítenek sajátokat.

A referátumok értékelését a 11., a referátumok minőségével foglalkozó fejezet tárgyalja részletesen. Itt elég annyit mondani, hogy a referátumok azok az adatelemek, amelyek a legtöbb intellektuális munkát követelik meg, hacsak nem a folyóiratból emelik ki őket. Érdemes azt is megjegyezni, hogy a referátum hossza nem feltétlenül jelzi a rekord informativitását, de az nyilvánvaló, hogy a hosszabb referátumok nagyobb valószínűséggel tartalmaznak több információt. Ezen felül az értéknövelő információt tartalmazó mezők jelenléte önmagában nem elegendő. A hozzáadott értékek akkor tölthetik be szerepüket, ha következetesen és pontosan alkalmazzák őket. Ezt tárgyaljuk meg a 7., a pontosságról, a 8., a következetességről és a 9., teljességről szóló fejezetben.

A primer dokumentumok teljes szövegének elérhetősége a rekordokban nyilvánvalóan nagy érték. Különböző módjai vannak azonban annak, hogy a teljes szöveget hogyan teszik elérhetővé. A sima ASCII formátum úgy tartalmazza a szöveget, hogy nem őrzi meg az eredeti tipográfiát. Azonban még ezen a szinten is különbségek vannak a szöveg olvashatóságában és az OCR (optikai karakterfelismerési) folyamat minőségében. A 6.10. és a 6.11. ábra ugyanakkor a teljes szövegű dokumentumnak ugyanazt a részletét mutatja.

The variety of databases, modestly listed under the subcategory "Forbes Lists" which doesn't reflect the fact that most are searchable files with dynamic output capabilities, cover a really impressive spectrum-200 Best Small Companies, Forbes Platinum 400, Forbes Annual Report on American Industry, Top 800 CEOs (by compensation), the International 800, the 400 Richest People in America, The World Richest People, the Forbes Celebrity 100, the Top 40 Entertainers, and Athletes, among others. Not all have the three-year retrospective coverage and the dynamic search and output generation capabilities of the 500 lists, but most do, with appropriate criteria. The International 800 database, for example, offers sorting by country, and two years of retrospective data.
(Illustration Omitted)
Captioned as: The usercustomizable Forbes 500 list on the Forbes Digital Toolbox site

6.10. ábra: Részlet az ABI/INFORM adatbázis egy rekordjából a DIALOG-nál.

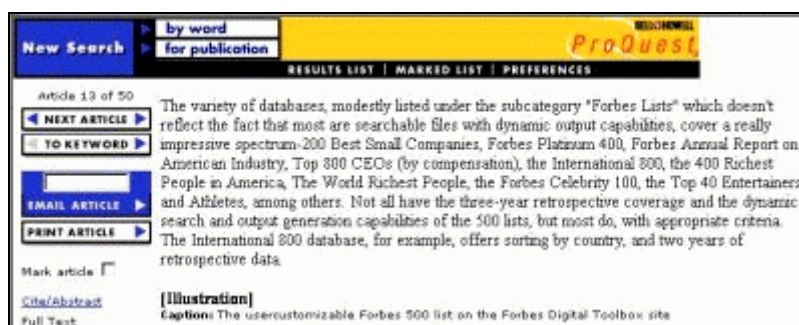
The variety of databases, modestly listed under the subcategory "**Forbes** Lists" which doesn't reflect the fact that most are searchable files with dynamic output capabilities, cover a really impressive spectrum-200 Best Small Companies, **Forbes** Platinum 400, **Forbes** Annual Report on American Industry, Top 800 CEOs (by compensation), the International 800, the 400 Richest People in America, The World Richest People, the **Forbes** Celebrity 100, the Top 40 Entertainers, and Athletes, among others. Not all have the three-year retrospective coverage and the dynamic search and output generation capabilities of the 500 lists, but most do, with appropriate criteria. The International 800 database, for example, offers sorting by country, and two years of retrospective data.
(Illustration Omitted)
Captioned as: The usercustomizable **Forbes** 500 list on the **Forbes** Digital Toolbox site

6.11. ábra: Részlet az ABI/INFORM adatbázis egy rekordjából az Ovidnál.

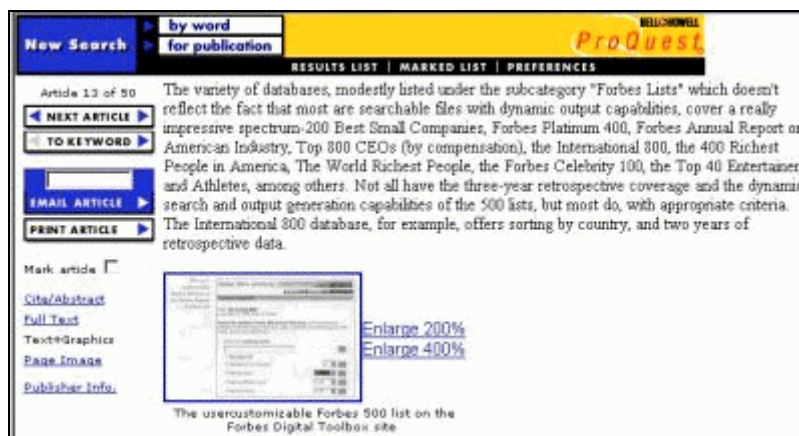
A teljes szöveg kiegészülhet ábrákkal, grafikonokkal, táblázatokkal és fotókkal, amelyeket nem reprodukálnak a cikk teljes szövegű változatában. Az ABI/INFORM megadja az ábrák címét, s jelzi, hogy honnan hagytak ki ábrát a szövegből. A Wilson Business Abstracts and

Full Text adatbázis nem említi meg a kihagyást a megfelelő helyen, s csak a cikk legvégén adja meg az ábrák címét. A legegyszerűbb a sima szöveges formátum, amely tartalmazza a képaláírásokat és jelzi az illusztráció helyét (6.12. ábra). A szöveg plusz grafika formátum tartalmazza az illusztrációnak a kisméretű változatát (6.13. ábra). Ez kinagyítható kétszeresére vagy négyszeresére és így jól olvashatóvá válik (6.14. ábra). Az eredeti cikk szöveges része tipográfiájának megtartása kétségtelenül érték.

Sok újságnak és folyóiratnak van jellegzetes betűtípusa és elrendezése, szedéstükre, az eredeti szöveg vastagított és dőltbetűs írásmódjának megőrzése nagy mértékben javíthatja a megértést. A ProQuest harmadik formátuma az eredeti oldal teljes képét visszaadja. A szöveg általában olyan jó, mintha fénymásolóval készült volna (6.15. ábra). Úgy tűnik, hogy az ábrák minősége sokkal inkább az eredetitől függ, mint a szöveges rész minőségétől. Bár több nagyítási lehetőség van ebben a formátumban, mint a szöveg plusz kép formátumban, a minőség nem olyan jó, és az egész oldalt kinagyítják, ez a navigációt valamennyire nehézkessé teszi, de még mindig megéri (6.16. ábra).

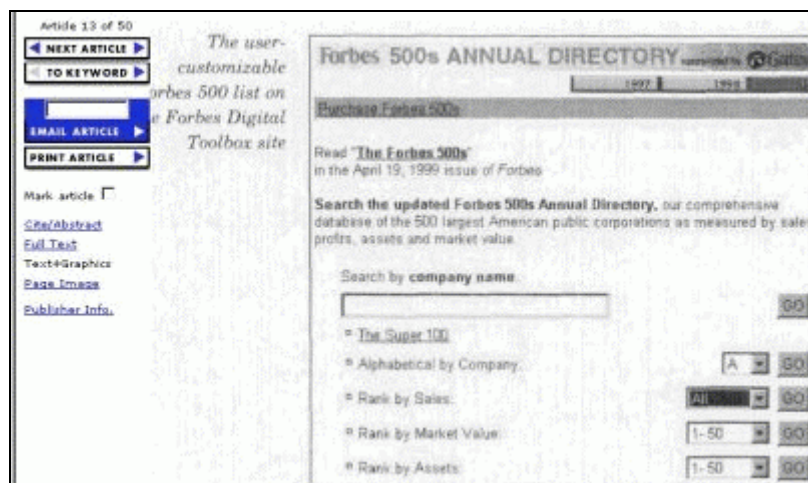


6.12. ábra: Részlet az ABI/INFORM adatbázis egy rekordjából a ProQuestnél.

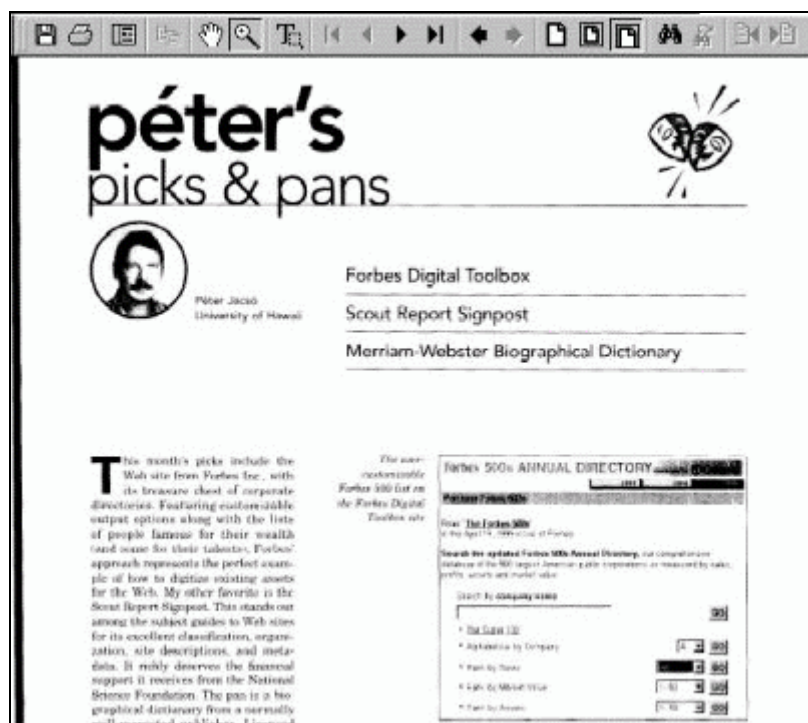


6.13. ábra: Szöveg plusz grafika formátum a ProQuestnél.

Egyes online szolgáltatók speciális szolgáltatást ajánlanak, amely különböző referáló és indexelő és teljes szövegű adatbázisok összekapcsolását használják ki. Az Ovid volt az úttörője ennek a technikának, s ők mutatták a legnagyobb fejlődést ezen a területen. Feltételezve, hogy a használóknak hozzáférésük van a megfelelő fájlokhoz, rákattinthatnak egy teljes szövegű adatbázisban a cikk végén a REFERENCES (hivatkozások) felíratra, s megjeleníthetik a teljes bibliográfiai rekordot, benne az indexelő és referáló adatbázisból származó referátummal.

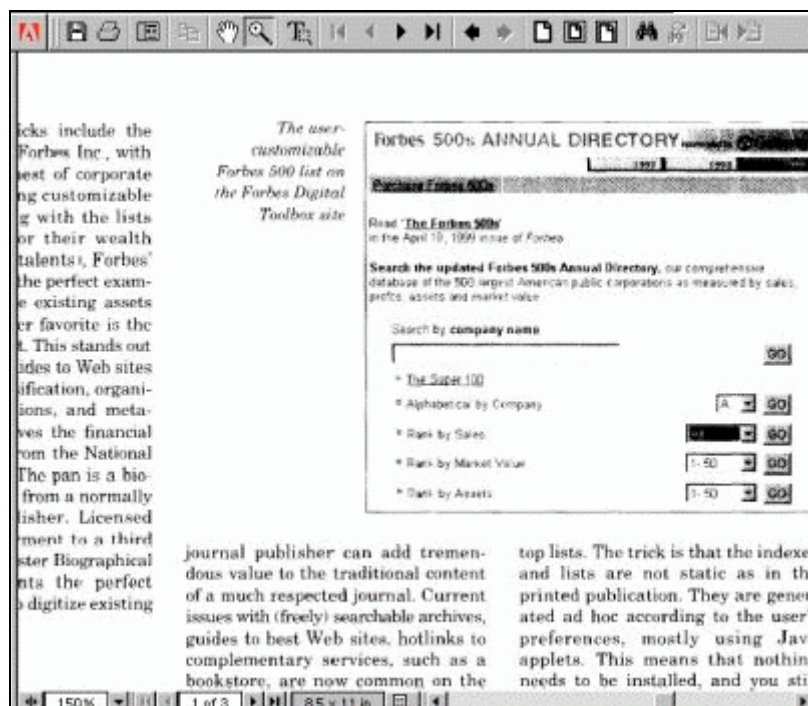


6.14. ábra: A kinagyított képi összetevő a ProQuestnél.

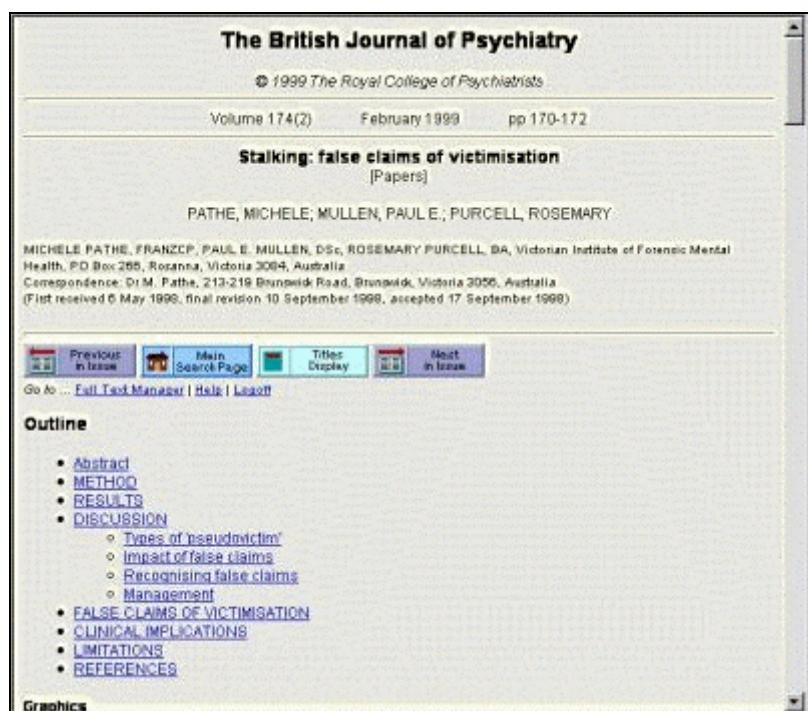


6.15. ábra: Részlet egy ABI/INFORM rekord képi változatából a ProQuestnél.

A 6.17. ábra a Mental Health Collection adatbázis Ovid változatának egy teljes szövegű rekordjából mutat egy részletet. A REFERENCES felíratra való kattintás megjeleníti a hivatkozások listáját. Ezek közül egyesek a PsycINFO, a MEDLINE vagy a BIOSIS adatbázishoz (vagy ezek kombinációjához) lesznek ugrópontok (6.18. ábra). Ha ezek valamelyikére rákattintunk, a kiválasztott adatbázisból a teljes bibliográfiai rekord meg fog jelenni, a referátummal együtt. A fordítottja is lehetséges - azaz egy indexelő és referáló adatbázisból átugorhatunk egy teljes szövegű adatbázis egy rekordjába (amelyhez a használatnak szintén hozzáférési joga van). Az ugrópontot olyan jól megtervezték, hogy még akkor is képes elérni a rekordot, ha elírások vannak a szerző nevében, mint a BIOSIS rekordja esetében, amelyben a *Kanin* név *Kianin* formában szerepel (nincs rá illusztráció).



6.16. ábra: A kinagyított kép a ProQuestnél.



6.17. ábra: Részlet az Ovid Mental Health Collectionének egy teljes szövegű rekordjából.



**6.18. ábra: Bibliográfiai adatok a referátumokkal
ellátott bibliográfiai rekordokhoz vezető
ugrópontokkal.**

Az életrajzi gyűjtemények kategóriájában mind a Gale Group Biography Resource Center (BRC), mind az H. W. Wilson cég Biographies Plus Illustrated adatbázisa teljes szövegű forrásnak minősíthetők alapos és kitűnően strukturált és szerkesztett életrajzaik miatt. Sok értéknövelt információt kínálnak, mint például nem, nemzetiség, születési és halálozási hely és dátum, amelyek a keresést nagyon hatékonyá teszik.

Más életrajzi forrásokban ezeket az információkat többnyire egybemossák az elég rövid leíró szövegrésszel, ami elég nehézé teszi azt, hogy úgy keressünk a korábbi információkra, hogy ne találjunk sok hamis rekordot is. A Gale Group (6.19. ábra) és az H. W. Wilson cég (6.20. ábra) életrajzi adatbázisai sok életrajzi tételhez kapcsolnak képet, ahogy azt a 6.3. és a 6.4. ábrák mutatták.

A Gale Group és az H. W. Wilson cég két életrajzi adatbázisa azzal is speciális hozzáadott értéket szolgáltat, hogy a teljes szövegű adatbázisaikban található szemlecekkel, kritikákkal és híryanagyokkal is összekapcsolják a rekordokat. (Az, hogy a Gale Group 1999-ben megvásárolta az IAC-t, jelentősen gazdagította referenz adatbázisainak gyűjteményét.)

A kereshetőség szükségessé teszi, hogy meghatározzuk azokat az információkat, amelyek a nyomtatott referáló és indexelő szolgáltatás olvasói számára feltehetőleg egyértelműek, amikor a nyomtatott kötet oldalait átnézik, mint például a nyelv vagy az eredeti dokumentum típusa (konferencia előadás, cikk vagy könyvkritika). Az ilyen jellemzőket egyértelművé kell tenni az eredmények hatékony szűrése érdekében. Még a bibliográfiai leírások strukturálása is fontossá válhat az indexelés és keresés érdekében. A National Criminal Justice Reference Service (NCJRS) előállítója nem különíti el a folyóirat címétől az egyes számok megjelenési idejét és számozását (6.21. ábra).

Ez visszafelé sül el, amikor a folyóiratcímek indexét akarjuk böngészni, s számos képernyőn kell lecsorognunk, amíg az egyik folyóirattól eljuthatunk a következőig, miközben ugyanazon folyóirat különböző köteteire és számaira vonatkozó tételek tucatjain (ha nem százain) kell átküzdenünk magunkat.

Az Adolescence (Kamaszkor) című lap összes tételén való átvergődés még a tinédzser gyerekek szüleit is türelmetlenné tenné. Hasonló módon a PaperChem adatbázisban az *Abbey Newsletter* tételeinek átnézése (6.22. ábra) egy szent türelmet is próbára tenné. Nem csak arról van szó, hogy ez a szakszerűtlen megoldás ellentétes a folyóiratcímek összehozásának céljával, de sokba is kerül a használatuk.

You have chosen Magazine and Newspaper Articles on *Stanley Kubrick*. Document(s) 1-10 of 34 displayed below.

Page#: **1** - 2 - 3 - 4

Subject / Citation

[Eyes of the Storm: STANLEY KUBRICK's final film, 'EYES WIDE SHUT,' kept stars TOM CRUISE and NICOLE KIDMAN on the set and out of sight \(not to mention underdressed\) for almost two years. Here's how he did it.](#) *Entertainment Weekly* Jul 23, 1999

[The Temptations: Tom Cruise and Nicole Kidman battle the demons of desire in Eyes Wide Shut, Stanley Kubrick's somber swan song.](#) (Movies)(Review) *Entertainment Weekly* Jul 23, 1999

[Dreaming With 'Eyes Wide Shut': As the hype dissipates, Stanley Kubrick's final film about sex and love steps quietly into the light.](#) (MOVIES)(Arts and Entertainment)(CE) *Newsweek* Jul 19, 1999

[Cinema: Kubrick's Dead, but His Projects Aren't.](#) (director Stanley Kubrick's ideas for film about artificial intelligence may be realized)(Notebook)(Brief Article) *Time* Jul 12, 1999

6.19. ábra: Ugrópont a kritika teljes szövegéhez a BRC adatbázisból.

Kennedy, John F.: 1960-

Miscellaneous

Below are citations and abstracts. To view the full text of the articles click on the titles of the articles.

1. Buckley, William F. [The honest agent](#), *National Review* v51 no11 p58 14 1999

The writer discusses the appearance of pornography magazine Larry Flynt at the White House Correspondents' annual dinner. He was there as the guest of John F. Kennedy Jr.

2. Kerwin, Ann Marie. [TFK Jr. keeps options open for 'George'](#), *Reader says he will turn to Condé Nast if Hachette wavers*, *Advertising Age* 970 no11 p4 15 1999

Although political monthly George is supported by Hachette Filipacchi Magazines, George founder John F. Kennedy Jr. has admitted that he is examining other options following the departure of David Perkin, the magazine's chief advocate at Hachette. Kennedy has been seen on several occasions with Steven T. Florio, CEO of Hachette rival Condé Nast Publications. However, Hachette insists that it remains committed to George.

6.20. ábra: Ugrópont a kritika teljes szövegéhez a BP adatbázisból.

E18	1	JN=ADOLESCENCE, V 12, N 46 (SUMMER 1977, P 175-1
E19	1	JN=ADOLESCENCE, V 12, N 48 (WINTER 1977), P 523-5
E20	1	JN=ADOLESCENCE, V 12, N 48 (WINTER 1977), P 559-5
E21	1	JN=ADOLESCENCE, V 13, N 49 (SPRING 1978), P 21-28
E22	1	JN=ADOLESCENCE, V 13, N 49 (SPRING 1978), P 29-43
E23	1	JN=ADOLESCENCE, V 13, N 50 (SUMMER 1978), P 209-2
E24	1	JN=ADOLESCENCE, V 13, N 50 (SUMMER 1978), P 339-3
E25	1	JN=ADOLESCENCE, V 13, N 50 (SUMMER 1978), P 365-3
E26	1	JN=ADOLESCENCE, V 13, N 51 (FALL 1978), P 375-392
E27	1	JN=ADOLESCENCE, V 13, N 51 (FALL 1978), P 401-410
E28	1	JN=ADOLESCENCE, V 13, N 51 (FALL 1978), P 461-482
E29	1	JN=ADOLESCENCE, V 13, N 51 (FALL 1978), P 483-493
E30	1	JN=ADOLESCENCE, V 13, N 52 (WINTER 1978), P 531-5
E31	1	JN=ADOLESCENCE, V 13, N 52 (WINTER 1978), P 643-6
E32	1	JN=ADOLESCENCE, V 13, N 52 (WINTER 1978), P 675-6
E33	1	JN=ADOLESCENCE, V 13, V 52 (WINTER 1978), P 543-5
E34	1	JN=ADOLESCENCE, V 14, N 53 (SPRING, 1979), P 65-7
E35	1	JN=ADOLESCENCE, V 14, N 55 (FALL 1979), P 451-464
E36	1	JN=ADOLESCENCE, V 14, N 56 (WINTER 1979), P 665-6
E37	1	JN=ADOLESCENCE, V 14, N 76 (WINTER 1984), P 817-8
E38	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 505-521
E39	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 549-555
E40	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 569-574
E41	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 595-621
E42	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 623-630
E43	1	JN=ADOLESCENCE, V 14, N 95 (FALL 1989), P 665-675
E44	1	JN=ADOLESCENCE, V 15, N 57 (SPRING 1980), P 13-22
E45	1	JN=ADOLESCENCE, V 15, N 57 (SPRING 1980), P 55-64
E46	1	JN=ADOLESCENCE, V 15, N 57 (SPRING 1980), P 65-74
E47	1	JN=ADOLESCENCE, V 15, N 58 (SUMMER (1980), P 257-

6.21. ábra: Tételek az évfolyamok, számok és oldalszámok adataival az NCJRS folyóiratindexében.

Az ISA tárgyi kód indexében való böngészés rendkívül zavaró a kódok rosszul szerkesztett szöveges megfelelői miatt (6.23. ábra). Például a 04. 08 kód mind az Information Generation and Promulgation, mind az Information Recognition and Description tárgyszóra vonatkozik, ahogyan azt az ábrán található első két tétel sugallná? A 04. 09 és a 04. 10 kódok szintén ehhez a két fogalomhoz kapcsolódnak? Miért van olyan sok különböző kódja az Information Processing and Control és az Information Systems and Applications fogalmaknak, anélkül, hogy bármilyen különbség lenne szöveges megfelelőjükben?

A válasz ezekre a kérdésekre az, hogy ezek főtárgyszavak, s megkülönböztető altárgyszavak nem láthatók az indexben - s nem is kereshetőek. A rekordokban láthatóak (6.24. ábra). Amikor megjelenítik őket egy, az információtudománnyal foglalkozó adatbázisban, egy kis információs technika és józan ész nem ártana. A tárgyszó szöveges részének két almezőre bontása (főosztály és alosztály) egyszerű módon megoldaná a problémát.

E19	1	JN=ABBAY NEWSL. 11, NO. 2
E20	1	JN=ABBAY NEWSL. 12, NO. 8
E21	1	JN=ABBAY NEWSL. 13, NO. 3
E22	1	JN=ABBAY NEWSL. 13, NO. 6
E23	1	JN=ABBAY NEWSL. 13, NO. 7
E24	1	JN=ABBAY NEWSL. 13, NO. 8
E25	1	JN=ABBAY NEWSL. 14, NO. 5
E26	3	JN=ABBAY NEWSL. 14, NO. 7
E27	1	JN=ABBAY NEWSL. 15, NO. 1
E28	3	JN=ABBAY NEWSL. 15, NO. 8
E29	1	JN=ABBAY NEWSL. 16, MONOGRAPH SUPPL. 2
E30	3	JN=ABBAY NEWSL. 16, NO. 2
E31	4	JN=ABBAY NEWSL. 16, NO. 3
E32	2	JN=ABBAY NEWSL. 16, NO. 4
E33	1	JN=ABBAY NEWSL. 16, NO. 7/8
E34	1	JN=ABBAY NEWSL. 17, NO. 1
E35	1	JN=ABBAY NEWSL. 17, NO. 4
E36	1	JN=ABBAY NEWSL. 17, NO. 6
E37	1	JN=ABBAY NEWSL. 17, NO. 7-8
E38	1	JN=ABBAY NEWSL. 18, NO. 3
E39	1	JN=ABBAY NEWSL. 18, NO. 8
E40	3	JN=ABBAY NEWSL. 19, NO. 3
E41	1	JN=ABBAY NEWSL. 19, NO. 5
E42	2	JN=ABBAY NEWSL. 19, NO. 6/7
E43	1	JN=ABBAY NEWSL. 20, NO. 2
E44	1	JN=ABBAY NEWSL. 20, NO. 3
E45	4	JN=ABBAY NEWSL. 20, NO. 4/5
E46	1	JN=ABBAY NEWSL. 20, NO. 8
E47	2	JN=ABBAY NEWSL. 21, NO. 1
E48	1	JN=ABBAY NEWSL. 21, NO. 3
E49	1	JN=ABBAY NEWSL. 7. NO. 3

6.22. ábra: Tételek kronológiai-numerikus adatokkal a PaperChem folyóiratindexében.

Ref	Items	Index-term
E1	6	SC=04.08 INFORMATION GENERATION AND PROMULGATION
E2	3075	SC=04.08 INFORMATION RECOGNITION AND DESCRIPTION
E3	10	SC=04.09 INFORMATION GENERATION AND PROMULGATION
E4	4897	SC=04.09 INFORMATION RECOGNITION AND DESCRIPTION
E5	3	SC=04.10 INFORMATION GENERATION AND PROMULGATION
E6	834	SC=04.10 INFORMATION RECOGNITION AND DESCRIPTION
E7	3838	SC=05.00 INFORMATION PROCESSING AND CONTROL
E8	1554	SC=05.01 INFORMATION PROCESSING AND CONTROL
E9	6938	SC=05.02 INFORMATION PROCESSING AND CONTROL
E10	102	SC=05.03 INFORMATION PROCESSING AND CONTROL
E11	2949	SC=05.04 INFORMATION PROCESSING AND CONTROL
E12	4619	SC=05.05 INFORMATION PROCESSING AND CONTROL
E13	10111	SC=05.06 INFORMATION PROCESSING AND CONTROL
E14	2651	SC=05.07 INFORMATION PROCESSING AND CONTROL
E15	3655	SC=05.08 INFORMATION PROCESSING AND CONTROL
E16	1397	SC=05.09 INFORMATION PROCESSING AND CONTROL
E17	1650	SC=05.10 INFORMATION PROCESSING AND CONTROL
E18	5401	SC=05.11 INFORMATION PROCESSING AND CONTROL
E19	612	SC=06.00 INFORMATION SYSTEMS AND APPLICATIONS
E20	6211	SC=06.01 INFORMATION SYSTEMS AND APPLICATIONS
E21	1	SC=06.02 BIBLIOGRAPHIC SEARCH SERVICES, DATABASE
E22	6023	SC=06.02 INFORMATION SYSTEMS AND APPLICATIONS
E23	1911	SC=06.03 INFORMATION SYSTEMS AND APPLICATIONS
E24	795	SC=06.04 INFORMATION SYSTEMS AND APPLICATIONS
E25	5602	SC=06.05 INFORMATION SYSTEMS AND APPLICATIONS
E26	6314	SC=06.06 INFORMATION SYSTEMS AND APPLICATIONS
E27	1573	SC=06.07 INFORMATION SYSTEMS AND APPLICATIONS
E28	3371	SC=06.08 INFORMATION SYSTEMS AND APPLICATIONS
E29	3513	SC=06.09 INFORMATION SYSTEMS AND APPLICATIONS

6.23. ábra: Különböző tárgyi kódok szemmel láthatóan ugyanahhoz a tárgyi kifejezéshez kapcsolódóan az ISA indexében.

1.	USC01	Subject Class Header (Number): Information Generation and Promulgation, Communications and Telecommunications Systems (03.11.) <small>Information Science Aba (Online® File 202) © Information Today, Inc. All rights reserved.</small>
7.	USC07	Subject Class Header (Number): Information Generation and Promulgation, Technology Transfer (03.06.) <small>Information Science Aba (Online® File 202) © Information Today, Inc. All rights reserved.</small>
9.	USC09	Subject Class Header (Number): Information Generation and Promulgation, Publicity (03.03.) <small>Information Science Aba (Online® File 202) © Information Today, Inc. All rights reserved.</small>

6.24. ábra: Tárgyi kódok és szöveges megfelelőjük a megjelenített rekordokban.

7. Pontosság

Az adatbázisból visszakeresett információk pontosságát gyakran természetesnek veszik. Sok használó számára az a tény, hogy az információt nem csak papíron szolgáltatják és prezentálják, hanem számítógépen is, extra hitelességet ad a CD-ROM-os és online adatbázisoknak. Az elmúlt 30 évben a nyomtatott output növekvő mértékben volt a számítógépes adatfeldolgozás mellékterméke. Az adatbázisok aligha lehetnek pontosabbak, mint a belőlük származó nyomtatott változatok. Sok más adatbázist pedig csak nyomtatott formában létező adatokból hoztak létre. A nyomtatott adatok számítógépes feldolgozása további hibákat hozhat és hoz is. Ezek közül van, amit technikai nehézség okoz, például amikor a szkennert/konverziós szoftvert nem ismer fel pontosan egy betűt vagy egy számot.

Amikor az adatokat manuálisan viszik be a nyomtatott szöveg, katalóguscédulák, űrlapok és kérdőívek alapján, nagy számban fordulnak elő további hibák. Bár igaz az, hogy megvan a lehetősége annak, hogy kifinomult adatbevitel-ellenőrző programokat használjanak, kevés fájllelőállító kíván erre költeni. Amikor visszamenőlegesen elvégzik az ellenőrzést, a megtalált és javított hibák mennyisége elképesztő.

Ez világosan látható olyan katalógusrekordoknál, amikor manuálisan vitték be az adatokat katalóguscéduláról vagy magából a primer forrásból. A probléma méretét jól illusztrálja az OCLC, amely napi 30000 javított hibát jelentett Online Közös Katalógusának takarításakor. Bár ez a szám mindenféle típusú hibát tartalmazott, így például kódolási hibákat és elírásokat is, nagysága még így is megdöbbentő.

A nyomtatott primer források felelősek sok hibaért. A közönség ritkán tud róluk, hacsak a hiba nem okoz kárt és nem teszi szükségessé egy könyv visszavonását. Az *American Libraries* (1996) jelentett számos ilyen esetet. Például ez történt Carol Walter *Great Cakes* (nagyszerű torták) című művével, amelyet kiadója, a Ballantine visszavont, „mert egy recept a hozzávalók között felsorolt egy mérgező virágot, a májusi gyöngyvirágot”. Valóban nem egészséges recept. „Csak” intellektuális kárt okozott volna egy történelmi tankönyvsorozat, amelynek több mint 20 millió dollár költséggel járó bevezetését már fontolgatták Texasban, amikor a könyv megvásárlása előtt egy jogvédő civil szervezet több mint 5200 hibát fedezett fel benne.

Dr. Ruth egyik könyvét azért vonták vissza, mert két olyan időszakot, amelyet a terhesség elkerülése szempontjából nem biztonságosként határozott meg, a végső változatban „biztonságos”-ra cseréltek. Pedersen (1992) a *Publishers Weekly* című lapban számolt be egy olyan tankönyv visszavonásáról, amely azt állította, hogy Truman elnök „könnyen befejezte” a koreai háborút az atombomba ledobásával.

Elég sokatmondó az, hogy az összes 3244 rekord közül, amelyet a *New York Times*ből a Newspaper Abstracts Daily adatbázishoz 1999 januárjában hozzáadtak, 121 volt egy korábbi számban megjelent információ korrekciója. Nem mindegyik volt világrengető hiba, de elég komoly volt ahhoz, hogy a *New York Times* szükségesnek tartsa javítását. Egy gyors keresés alapján a hibáknak az új rekordok számához viszonyított aránya a *Washington Post* (2114-ből 15), a *Wall Street Journal* (4117-ből 16) és a *Los Angeles Times* (1153-ből 10) esetében sokkal alacsonyabb volt, de ez kapcsolatban lehet az újságok eltérő korrekciós gyakorlatával is.

Meglepő, hogy milyen nyilvánvaló hibák kerülhetnek be vagy származhatnak még könyvtárosoktól is, akik pedig a pontosságot nagyra értékelik. Az *American Libraries* (1996) szerint egy, 1996-ban a Chicago Public Library által összeállított történelmi kalendárium első kiadásában olyan hibák voltak, mint Michael Jordon Michael Jordan helyett, Lord Tenneyson Lord Tennyson helyett, míg Jean Baptiste Point du Sable-nak mind a születési, mind a halálozási

dátumát 1818-ra tették. Keveseknek (ha egyáltalán vannak ilyenek) lehetnek olyan eredményeik újszülöttként, hogy kiérdemljenek egy helyet egy kalendáriumban.

A kalendárium két különböző évet is megadott Richard Daley első alkalommal való megválasztásaként. A politikában bármi lehetséges, de nem valószínű, hogy Richard Daley-t kétszer választották volna meg első alkalommal, még akkor sem, ha figyelembe vesszük Chicago változatos választási történetét. Hibák nagy számban fordulnak elő szépirodalmi és ismeretterjesztő művekben is, köztük a gyerekeknek szóló életrajzokban. Ahogy egyre több nyomtatott anyagból készül online adatbázis, a digitális változatok terjesztik a hibákat.

A helyesírási hibák és elírások az adatbázisok pontatlanságainak legáltalánosabb típusai. Online katalógusokat elemeztek a leggyakrabban, hogy felfedezzék a pontatlanságok változatait. O'Neill és Vizine-Goetz alapos elemzést adott erről a szakirodalomban 1988-ban. Az elírások problémájának nagysága szembeűnő. Klemmer és Lockhead (1962) úgy találta, hogy a hibák mértéke 2 és 6 között van 10000 leütésenként. A Pollock és Zamora (1975) által a Chemical Abstracts Service adatbázisában végzett mérés szerint a szavakban számított hibaarány relevánsabb lehet, mert egyetlen hiba egy szóban egy rekordot visszakereshetlenné tehet.

Minden 1000 szó között 2 elírtat találtak ebben a drága adatbázisban. Azt is megállapították, hogy a hibák 90-96%-a a négy legáltalánosabb hibatípusba tartozik: karakter(ek) kihagyása: 30-40 százalék; karakter(ek) beillesztése: 25-35 százalék; karakterek cseréje: 15-20 százalék; és felcserélése: 10-15 százalék.

Tudományos forrásokról van szó, amelyek legtöbbje átmegy lektoráláson és szerkesztésen, ezek feltételezhetően hozzájárulnak ezen folyóiratok magas költségeihez és szakmai nagybecsüléséhez. A nagyközönségnek szóló források, például a napilapok helyzete minden bizonnyal rosszabb.

A hibák legalaposabb, Yannakoudakis és Fawthrop (1983) által végzett elemzése feltárta, hogy sok hibát követnek el a magánhangzókkal, s azokkal a szavakkal, amelyekben a *w*, *y* és *h* betűk szerepelnek. A dittográfia és a haplográfia (olyan betűk megduplázása, amelyekből csak egynek kellene lennie és betűk kihagyása, amikor duplán kellene szerepelniük, például olyan hibák, mint *leter duplication*) nagyon általános. A hibatípusok és minták részletes elemzése nagyon hasznos lehet a defenzív keresési stratégiák kifejlesztéséhez.

Az 1990-es évek végén a problémát enyhítette sok online katalógusban az automatikus hiba ellenőrző és -korrigáló műveletek bevezetése, de a referáló és indexelő szolgáltatások nem alkalmazták kellő mértékben az eljárást. A legendásan tiszta Western Library Network (WLN); korábban Washington Library Network) bibliográfiai adatbázisának és az OCLC-nek az 1999-es összeolvasztása valószínűleg segíteni fogja az OCLC jelentős tisztítási próbálkozásait. Az is jó jel, hogy több adatbázist építenek úgy, hogy közvetlenül a fájl előállítójától veszik át a fájlokat. A közvetlen adatbevitel csökkenti a bibliográfiai hivatkozások, referátumok és a teljes szöveg újragépeléséből származó további hibák előfordulását. Ez különösen igaz a korábban idézett tanulmányok fényében, amelyek azt állítják, hogy a legtöbb gépelési hiba a QWERTY billentyűzet használatával függ össze. Valamennyire ironikus, hogy a sokkal jobb Dvorak billentyűzet, amely a tesztek szerint drasztikusan csökkenti a gépelési hibákat (és a carpal tunnel szindrómát), sohasem terjedt el.

A referáló és indexelő adatbázisokban előforduló helyesírási hibákat már nagyon korán, 1977-ben vizsgálta Charles Bourne az adatbázis szolgáltatójának szoftverének használatával. 3600 indexfogalmat (tárgyszavakat, valamint a referátumokból és a címekből származó szavakat) vizsgált 11 különböző adatbázisban. Az elírt indexfogalmak aránya a BIOSIS 0,4%-ától az ABI/INFORM 22,8%-áig terjedt, a továbbiak közül az ISMEC (Information Service in Mechanical Engineering) (0,6%); ERIC (4,2%); Social Science Citation Index (6,1%) és NTIS (6,5%) voltak az alacsony tartományban és a Predicast PATS adatbázisa (12,4%) és a Compendex (12,3%) a magasban.

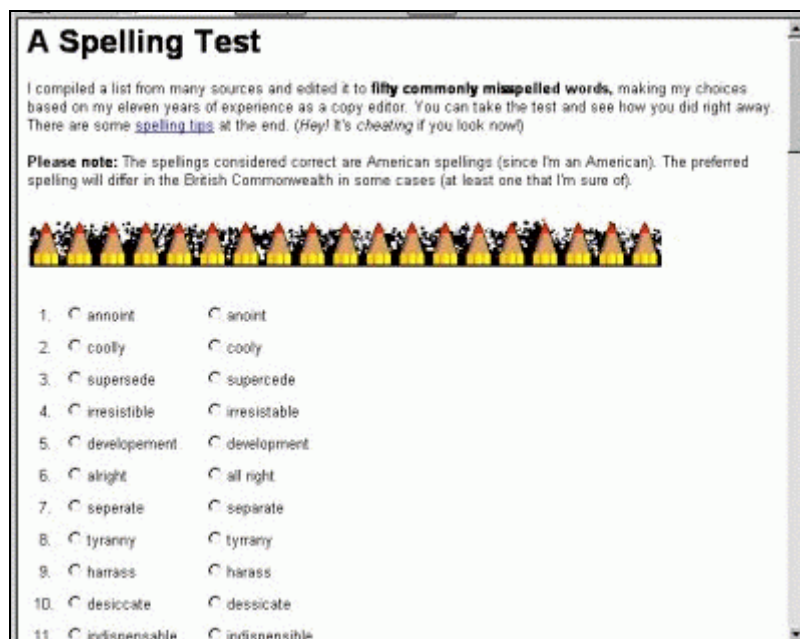
Kicsit ironikusnak találtam, hogy a ISMEC elírva, *ISMES* formában szerepel Bourne cikke (1977) egyik ábrájának feliratában, egy alaposan lektorált és szerkesztett folyóiratban, az *Information Processing & Management*-ben. Az ABI/INFORM később nagyarányú tisztítási folyamaton ment keresztül, amelynek keretében az elírások többségét megszüntették.

Jeffrey Beall a piszkos adatbázis teszteléséről talán kevésbé tudományos, de nagyon pragmatikus abból a szempontból, hogy képet kapjunk az adatbázisok helyesírásának pontosságáról, olyan gyakran elírt szavak használatával, mint a *Wensday*, *goverment*, *grammer* és egy tucat másik gyakran elírt szó. A *Database Searcher* szerint Harold Way (1988) számolta ki 14 adatbázis pontossági értékét Beall teszt-szavainak felhasználásával. Dwyer (1991) finomította Bourne és Beall módszereit, kiküszöbölve hiányosságait azzal, hogy kiszámolta ugyanazon szó elírt és helyesen írt változatainak arányát, ezt a módszert maga Bourne ajánlotta. Jacsó (1995) néhány szó pontos és elírt változatát ellenőrizte öt könyvtár- és információtudományi adatbázisban. Úgy találta, hogy a LibLit rendkívül tiszta, míg a legtöbb hiba az ISA-ban található, majd az ERIC-ben. Az ISA az *accommodat* szógyököt több mint 200 rekordban írta el; majdnem a helyesen írt változatok 20%-ában.

Cahn (1994) továbbfejlesztette ezeket módszereket - többek között - annak vizsgálatával, hogy a rekord tartalmazta-e az elírt szót helyesen írt formában is, azaz csökkentve vagy megszüntetve a keresésre gyakorolt hatását. Újra csak érdekes megjegyezni, hogy a John Wiley kiadó által megjelentetett *Harvard Business Review* (egy nagy befolyással rendelkező folyóirat és nagyon drága tudományos folyóiratok kiadója) utcahosszal megelőzte a 13 másik adatbázist - a legtöbb hibát produkálva Cahn tesztjeinek többségében.

Ballard és Lifshin (1992) végignézte egy egyetem online katalógusának kulcsszó indexében mind a 117000 fogalmát, s elemezte a hibákat az adatmező, szófaj és hibatípus szerint. 1082 hibás indexfogalmat találtak. Mivel a könyvkatalógusokban való kereséskor a címmezőnek kitüntetett szerepe van, elég hervasztó, hogy a címmezőből generált fogalmak tették ki a hibák 60%-át. Szerencsére több mint felük az alcímben volt, amely kevésbé kritikus a böngészés és a keresés szempontjából. A pozitív oldalhoz tartozik az, hogy nem voltak hibák a tárgyszavak mezőjében, s a szerzői mezőnek csak 2%-ában voltak elírások. (nem világos azonban a cikk alapján, hogy a szerzői nevek pontosságát hogyan ellenőrizték.) A hibák típusok szerint való megoszlása hasonló volt ahhoz, mint amit az előzőekben bemutattunk.

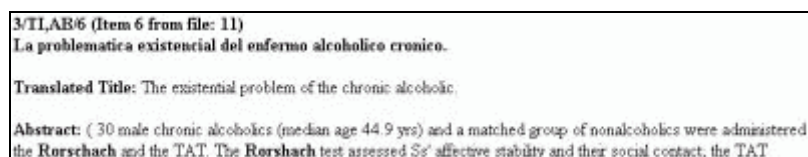
A hibának a cím mezőn belüli helyét is elemezték. Rossz hír az, hogy a hibák több mint 14%-a a cím első szavában fordult elő, s csaknem 41% az első három szóban. Az elírt kezdő szavakat tartalmazó tételek drámai módon csökkentik annak valószínűségét, hogy egy könyvet a cím alapján történő böngészéssel meg lehessen találni. Az elírt második és harmadik szavak szintén csökkentik egy ismert tétel megtalálásának valószínűségét, de kisebb mértékben, ami attól is függ, hány rövid tételt jelenítenek meg a képernyőn és milyen messze van az elírt tétel a helyesen írttól. A több rövid tétel növeli annak esélyét, hogy a használó észreveszi az elírt tételt, és a helyesen írt változattól való távolság kisebb lehet, ha az elírás a szó végén van. Az elírt *Jacsu* közelebb van a *Jacsó*hoz, mint a szintén elírt *Jascó*. Ironikus módon, amikor a katalógusban leggyakrabban elírt szavakra utalnak, köztük a *commerical*, *reseach* és *adminstration* változatokkal, a szerzők felsorolják a *government* szót is, azaz a helyesen írt változatot. Ez újra csak a nyomda ördöge lehetett. Az ilyen átfogó, a teljes populációra vonatkozó tesztek rendkívül sok időt követelnek meg. A problémás szavak egy vagy több leggyakrabban elírt formáját és a helyesen írt változatát használó módszer elegendő lehet ahhoz, hogy benyomást kapjunk az elírások mértékéről. Kiindulásul érdemes ellenőrizni egy weblapról (<http://www.sentex.net/~mmcadams/spelling.html>) a leggyakrabban elírt szavakat, a helyesen írt változattal és a helytelen változat gyakorisági mutatójával együtt egy olyan adatbázisban, amelyet gyakran használunk (7.1. ábra).



7.1. ábra: A leggyakrabban elírt szavak.

Különbséget kell tenni azok között a mezők között, amelyek egy keresési szempont kizárólagos forrásai (mint például a kiadás országa) és az olyanok között, amelyek más mezőkben is megjelenhetnek. Az elírás a referátum szövegében nem olyan kritikus, mert a helyesen írt változat szerepelhet a címben vagy a tárgyszavakban (vagy mindkettőben). Ez a megkülönböztetés nagyon hasonló ahhoz, ami a kontrollált szótárt használó mezők és a szabad szöveges mezők között van. A *Rorschach* név elírása a referátumban nem probléma a PsycINFO adatbázisban (7.2. ábra), mert pontosan is előfordul a referátum első mondatában, így a rekord visszakereshető. Az elírt változat az MHA cím mezőjében (7.3. ábra) nem tetszetős, de mivel helyesen is szerepel az eredeti, spanyol nyelvű címmezőben, ez csak a szemet bántja, amennyiben az eredeti címmezőt (azaz a spanyol címet) is felhasználják a kulcsszavak indexének vagy az alapszótárnak a létrehozásához.

Még abban az esetben is, ha az elírt változat az egyetlen előfordulás a referátumban (mint a második rekordban), ez nem súlyos probléma, amikor a fogalom helyesen megjelenik a tárgyszavak mezőjében. Amikor az elírt változat, *Rorshach* az egyetlen formátum, amelyben a név megjelenik, a rekord visszakereshetelenné válik a helyes névalak, Rorschach alapján (7.4. ábra). Olyan adatbázisok esetében, mint például az Inside Conferences, amelyekben nincs referátum s amelyek nagyon széles tárgyi kategóriákat használnak, a címben történő elírások végzetesek lehetnek. Öt rekordból kettőben a *toxoplasma* szót *taxoplasma* formában elírták a címben, ez látható a 7.5. ábrán. Az elírások hatalmas számának fényében üdvözölni kell azt, hogy egyes keresőprogramok okosabbá váltak, és vagy automatikusan korrigálják a használó által beírt szavakat, vagy felajánlják mind a betűrendben közel álló fogalmakat, mind a használó által beírthoz hasonlóan hangzó szavakat. Az előbbit a National Criminal Justice Reference Service adatbázisának a fájl előállítója által készített változatában használják; az utóbbit az *Encyclopedia Britannica* előfizetési változatában használták 1999-ig.



7.2. ábra: A *Rorschach* névnek egy helyesen írt és egy elírt változata van a referátumban.

3/8/19 (Item 9 from file: 86)
 0046269
 IPA1970-01382
Psychodiagnosis of Rorschach, and delinquency.
 Psicodiagnostico de Rorschach y delincuencia.
Descriptors: Behavioral Sciences; Crime and Delinquency
Identifiers: Journal Article; maturation
 Mental Health Abstracts (Dialog® File 86): (c) 1999 IP/Plenum Data Corp. All rights reserved.

7.3. ábra: Helyesen írt változat az eredeti (spanyol) címben és elírt változat az angol címben.

U/TLAB,DE,1D/8 (Item 3 from file: 86)
 Graphology -- an art, science, or fraud?
 Cases are made for and against the validity of graphology. Graphologists find considerable acceptance in industrial personnel departments. In courts of law, handwriting analysis generally is considered to be reliable with respect to the validation of a signature, but not with respect to personality analysis. A random survey of 10 police departments found that only one of them had ever used the services of a graphologist in compiling a personality profile, and that this analysis did not result in the apprehension of the writer. It is suggested that the law enforcement profession should regard graphology with skepticism, although graphologists insist on handwriting analysis as a valuable projective technique on a par with the Rorschach dream interpretation, and free association to reveal a person's character traits.
Descriptors: Crime; Personality; Psychometrics
Identifiers: Journal; Evaluation; Human; Overview
 Mental Health Abstracts (Dialog® File 86): (c) 1999 IP/Plenum Data Corp. All rights reserved.

7.4. ábra: A *Rorschach* név egyetlen előfordulása el van írva a referátumban.

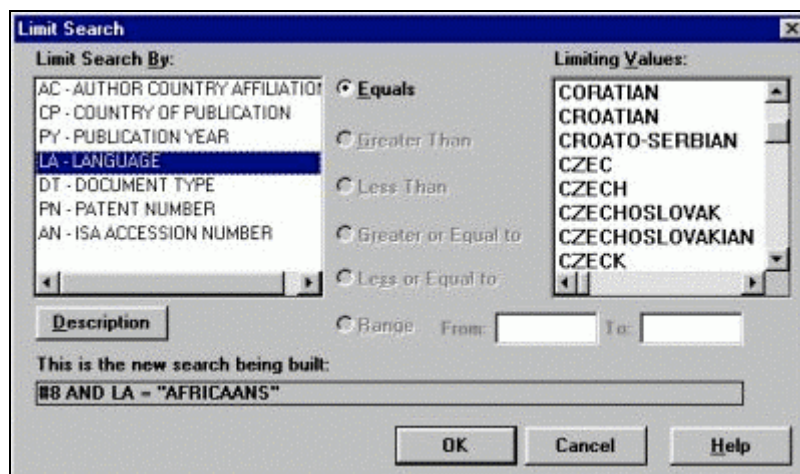
Azoknál a mezőknél, amelyekben a lehetséges értékeknek korlátozott a száma, mint a nyelv, a kiadás országa vagy a dokumentumtípus, elég egyszerű ellenőrizni és kiszámolni a pontosság mértékét - a helyesen írt és az elírt változatot tartalmazó rekordok arányát. Könnyű is kijavítani egy hozzáértő fájl előállító számára, aki kellően tiszteli a használatát. Egy adatbázis nyelvi indexe, amely viszonylag nem sok nyelven írt dokumentumokat tartalmaz, csak egyszeri áttekintést igényel, mivel a nyelvek száma általában egy-két tucat.

A hibás tételek nagy száma miatt a nyelvi index több képernyőre szóródhat szét, s emiatt a használó elveszíthet releváns rekordokat, ha nem veszi észre az elírt változatokat. Elég nyilvánvaló például az ISA nyelvi indexe alapján, hogy szükség lenne valakire, aki ellenőrizné, hogyan kell helyesen írni a Czech (cseh) szót. Ebben az esetben az elírások legalább közel vannak egymáshoz, de a helyes *Serbo-Croatian* (szerb-horvát) változat több képernyőnyire van a nem létező *Croato-Serbian* (horvát-szerb) nyelvtől (7.6. ábra).

5/TL,SO,PY/2
 Long-term immunity to shedding of *Taxoplasma gondii* oocysts by cats
 American Society of Parasitologists
 PROCEEDINGS- AMERICAN ASSOCIATION OF VETERINARY PARASITOLOGISTS, 1995, CONF 40, p. 178
 AAVP, 1995
Conference:
 American Society of Parasitologists
 Inada Conference (Dialog® File 65): (c) 1999 ELDSC all its. reserv. All rights reserved.

5/TL,SO,PY/5
 Role of .alpha./beta. and gamma/.delta. T cells in resistance against *Taxoplasma gondii*
 American Society of Parasitologists
 PROCEEDINGS- AMERICAN ASSOCIATION OF VETERINARY PARASITOLOGISTS, 1995, CONF 40, p. 54
 AAVP, 1995
Conference:
 American Society of Parasitologists

7.5. ábra: A *toxoplasma* szó egyetlen előfordulása el van írva a címben.



7.6. ábra: Elírások az ISA nyelvi mezőjében.

Míg a *Portuguese* (portugál) szó elírt változatai a helyesen írt mellett találhatóak a 7.7. ábrán, az a tény, hogy a pontatlan változatok száma eléri a helyesen írtak felét, nem növeli az ISA adatbázis vonzását. Mindezek ijesztővé válnak az adatbázis DOS-os CD-ROM változatában, mert ebben a változatban a nyelvi index nem böngészhető (csakúgy, mint a kiadás országa vagy a dokumentumtípus mező sem). Ez elkendőzi a fogyatékoságot a használó előtt, megfosztva őket portugál nyelvű dokumentumok tételeinek felétől, ha helyesen írják a nyelvet a kérdésben.

Egyes adatbázisok még ennél is rosszabbak, ott az elírt változatok száma ugyanennek a nyelvnek az esetében felülmúlja a helyesen írtakat. A 7.7. ábra mutatja a *Portuguese* elírt és helyesen írt változatainak megoszlását néhány adatbázisban 1999 februárjában. A FLUIDEX-ben (File 96) és a Pollution Abstractsben (File 41) közel háromszor annyi elírt változat van, mint helyesen írt. A PROMT adatbázisban (File 16) csupán 9-szer írták helyesen a *Portuguese* szót, 232 hibás mellett. Néha a helyesírási változatokat elfogadhatónak tekintik a szótárak, de a *Portuguese* változat úgy rossz, ahogy van.

HOME	DATABASES	COMMAND SEARCH	GUIDED SEARCH	COST
S1	LA=PORTUGUESE	638		Display
S1	LA=PORTUGUESE	147	202	Display
S1	LA=PORTUGUESE	25	96	Display
S1	LA=PORTUGUESE	95	41	Display
S1	LA=PORTUGUESE	139	28	Display
S1	LA=PORTUGUESE	232	16	Display
S2	LA=PORTUGUESE	500		Display
S2	LA=PORTUGUESE	290	202	Display
S2	LA=PORTUGUESE	71	96	Display
S2	LA=PORTUGUESE	37	41	Display
S2	LA=PORTUGUESE	93	28	Display
S2	LA=PORTUGUESE	9	16	Display

7.7. ábra: A *Portuguese* elírt és helyesen írt változatainak aránya a nyelvi indexben.

Miért keresné egy tipikus használó egy nyelv elírt változatait, ha egyes rekordokban valóban megtalálható a helyes változat? Ebből a szempontból a PsycLIT Compact Cambridge változata jobb, mert mind az 1375 rekordban elírták a *Portuguese* szót (a tény, hogy egyetlen rekordot sem talál, figyelmeztetheti a használót, hogy valami probléma lehet a nyelvi mezővel). Ugyanez érvényes az IAC két adatbázisára, az AeroSpace/Defense Markets Technology-

ra és a Marketing & Advertising Reference Servicesre amelyekben nincs egyetlen rekord sem, amelyben helyesen írják ezt a nyelvet.

Ugyanez igaz a Mental Health Abstracts adatbázisra is. Nincs egyetlen rekord sem, amelyben az *English* szó helyesen szerepelne a nyelvi mezőben. Érdekes, hogy az adatbázis dokumentációja arra utasítja az indexelőt, hogy ha „a dokumentum nyelve angol, nem kell feltüntetni”. Az ember kíváncsi, hogy vajon ezt a döntést az után hozták-e meg, miután csaknem 100 alkalommal nem sikerült ezt a nem túl nehéz szót helyesen beírni. Az a személy, aki a DIALOG adatbázis-leírását (bluesheetjét) írta az MHA adatbázisról az S LA=ENGLISH példát használva, minden bizonnyal nem tudott a döntésről, s nem tesztelte a mező használatára ajánlott példát. A használók, akik a nyelvi indexet böngésszik, ha hajlandók képernyőt képernyő után görgetni, legalább láthatják a küzdelmet, hányféle módon lehet elérni az *English* szót (7.8. ábra). 1992-től az ISA megtanulta, hogyan kell a *Portuguese* szót helyesen írni; így mostanra kétszer annyi helyesen írt eset van, mint elért.

Ha megvan a hozzáértés és a használó iránti tisztelet, akkor elérhető, hogy a nyelvi index teljesen tiszta legyen, ahogyan az a Library Literature-ben és az összes H. W. Wilson adatbázisban látható (7.9. ábra). Más kérdés, hogy miért található olyan sok rekordban az „undetermined” (meghatározatlan) változó a nyelvre vonatkozóan. Kiderül, hogy ezek mind könyvkritikákra vonatkozó rekordok, s a könyv eredeti nyelvére utalnak, nem pedig a kritikákéra. Az indexelő feltehetőleg kitalálhatná a könyv nyelvét a kritikából és megadhatná a megfelelő nyelvet a rekordban.

new search databases alerts cost logoff			
File	Items	Total	Index-term
86		1	LA=EMGLISH
86		1	LA=ENBLISH
86		5	*LA=ENG
86		1	LA=ENGIISH
86		24	LA=ENGLISH
86		1	LA=ENGLAIH
86		25	LA=ENGLAND
86		1	LA=ENGLIAH
86		2	LA=ENGLIDH
86		2	LA=ENGLIGH
86		1	LA=ENGLIISH
86		1	LA=ENGLILSH
86		3	LA=ENGLIS
86		5	LA=ENGLISH ENGLISH
86		1	LA=ENGLISH 12
86		8	LA=ENGLISH.
86		2	LA=ENGLISH/JAP.
86		1	LA=ENGLISH,
86		1	LA=ENGLISH'
86		3	LA=ENGLISN
86		1	LA=ENGLLISH
86		1	LA=ENGLSH
86		20	LA=ENGLSIH
86		1	LA=ENGLUSH
86		10	LA=ENLGISH

7.8. ábra: Az *English* szó nagyszámú elírása a Mental Health Abstracts adatbázis nyelvi mezőjében.

Dialog Response				
Ref	File	Items	Total	Index-term
E1	438		0	*LA=
E2	438		246	LA=AFRIKAANS
E3	438		60	LA=CHINESE
E4	438		980	LA=DANISH
E5	438		1013	LA=DUTCH
E6	438		143494	LA=ENGLISH
E7	438		768	LA=FINNISH
E8	438		2930	LA=FRENCH
E9	438		4054	LA=GERMAN
E10	438		191	LA=HUNGARIAN
E11	438		2	LA=IRISH
E12	438		532	LA=ITALIAN
E13	438		135	LA=JAPANESE
E14	438		6	LA=MULTILINGUAL
E15	438		1675	LA=NORWEGIAN
E16	438		1001	LA=POLISH
E17	438		508	LA=PORTUGUESE
E18	438		5285	LA=RUSSIAN
E19	438		132	LA=SPANISH
E20	438		281	LA=SWEDISH
E21	438		25880	LA=UNDETERMINED
E22	438		3	LA=URDU

7.9. ábra: A nyelvek teljesen hibátlan névalakjai a LibLit adatbázisban.

Nehezebb dolog szerzői nevek tízezreinek pontos helyesírását elérni, pedig ez sok adatbázis gondja. A pontos szerzői név alapvető fontosságú sok keresés esetében, és a legtöbb adatbázisban a pontos szerzői név nem vehető kieszpénznek (a névalakok következetessége még valószínűtlenebb). Az H. W. Wilson cég adatbázisai, a WLN bibliográfiai adatbázisa és az Internet Movie Database kiemelkednek hibátlan névrendszó-kezelésük (authority control) miatt, beleértve az egységes névalakok. Bár a Kongresszusi Könyvtár (LC) az elsők közé tartozott az egységesített névalakok kifejlesztésében és alkalmazásában, ezeknek is vannak hiányosságaik. Az LC authority fájljában előforduló hibák azért súlyosak, mert bekerülnek minden olyan katalógusba, amely az LC MARC rekordokat használja. Lehangelő látni, hogy F. W. Lancaster egyik keresztnéve hibás formában (*Wilfred*) szerepel az LC authority fájljában, a helyes *Wilfrid* mellett.

A szerzők vagy más személyek neve pontosságának tesztelése egy adatbázisban nem lehet mindenre kiterjedő, mert még a kis adatbázisokban is túl sok név szerepel. Nehezebb is mintát venni, mint a szabad szöveges mezők vagy a legtöbb kontrollált szótáros mező pontosságát ellenőrizni, hacsak nem rendelkezünk abszolút megbízható forrással a személynevekről. Jacsó (1989) Rolodexéből vett névjegykártyákat és a Columbia Egyetem Könyvtárosképzőjének (School of Library Services of Columbia University) katalógusában szereplő neveket használt, hogy ellenőrizze ismerősei és oktatókollégái nevét a könyvtárosok és információs szakemberek címjegyzékében (Directory of Library and Information Professionals).

Összesen 469 személyt keresett név alapján, de ezzel a megközelítéssel csak a könyvtárosok és információs szakemberek kevesebb mint felére vonatkozóan talált rekordokat. Egy más szempontok szerinti kiegészítő keresés valamivel növelte a találatok számát és feltárta, hogy az adatbázisban gyakori a nevek elírása. Olyan további keresési szempontok, mint a személy munkahelye, nem következetesen elérhetők és/vagy elavultak ebben az adatbázisban, ezért nem lehet azokra számítani tartalékként, ha a nevek szerinti keresés nem hoz eredményt.

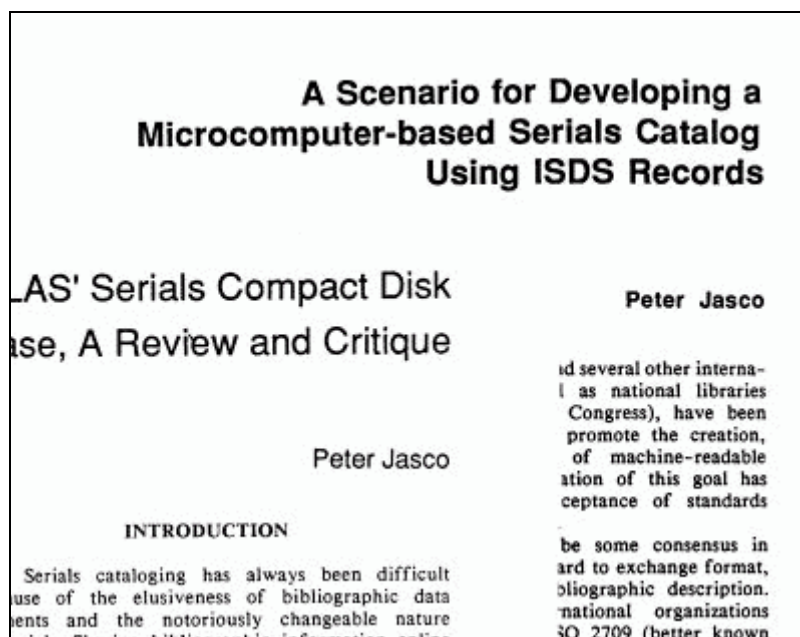
Ennek a címjegyzéknek a sok elírása még a szokottnál is fájdalmasabb, mivel az adatbázis az American Library Associationnel együttműködésben készült.

A Columbia Egyetem két oktatójának (tíz közül) az esetében az elírások végzetesek voltak, mert az első betűben követték el őket (*Fres*, *Beth Eres*, *Beth* helyett és *Tetherbridge*, *Guy Petherbridge*, *Guy* helyett.) Más pontatlanságok a név végénél voltak, így azokat a csonkolás alkalmazásával enyhíteni lehetett. Ebben az adatbázisban a személynév az egyedüli adatelem, amely minden rekordban jelen van; azaz bármilyen hiba elérhetetlenné teheti a rekordot.

Ebben a példában a fájl előállítója a felelős a pontosságért, de nem mindig ez a helyzet. A fájlok előállítói nem hibáztathatók, ha az eredeti forrásdokumentumban írtak el egy nevet. A *Serials Review*-nak sikerült kétszer a *Jacsó* nevet *Jascó* formában elírnia (7.10. ábra) annak ellenére, hogy a szerző nem éppen halk szavúan tiltakozott ez ellen az első eset után. Mivel a lektorált folyóiratok kiadói nem fizetnek a szerzőknek, az egyetlen ellenszolgáltatás (jutalom) az lehet, ha elismerik és idézik őket, de amikor a szerző nevét rosszul írják, még ez az ellenszolgáltatás is elvész.

A *Serials Review* lektorált folyóirat, de olvasószerkesztői munkája a szerzői kézirat munkafázisainál kevésbé rigorózus folyamat. Kivételes az, hogy az H. W. Wilson bibliográfiai rekordjában mindkét fenti esetben kijavította a szerző elírt nevét, s a LISA is egy esetben. Nagyon egyedi és könnyen megkülönböztethető nevek kis mintájának átnézése a szerzői indexben jól jelezheti a szerzői nevek pontosságát.

A több adatbázisban végzett keresés eredményei illusztrálják e szerző nevének leggyakrabban elírási formáit (7.11. ábra). Meglepő módon még olyan adatbázisokban is, amelyek törődnek a minőséggel, mint például a Microcomputer Abstracts, viszonylag magas az elírások száma ebben a mintában.



7.10. ábra: Pontatlan névalakok a forrásdokumentumban, a *Serials Review*-ban.

Your select statement is 'S AU=JASCO, P?'.

☒ Select All
 ☐ Clear Selections

File	Database Name	Hits
<input type="checkbox"/> 202:	Information Science Abstracts	18
<input type="checkbox"/> 233:	Microcomputer Abstracts(TM)	13
<input type="checkbox"/> 148:	IAC(SM) Trade & Industry Database(TM)	11
<input type="checkbox"/> 484:	Periodical Abstracts PlusText(TM)	5
<input type="checkbox"/> 15:	ABI/INFORM®	5
<input type="checkbox"/> 61:	LISA (Library & Information Science Abstracts)	4
<input type="checkbox"/> 47:	IAC(SM) Magazine Database(TM)	4
<input type="checkbox"/> 2:	INSPEC (1969- present)	2
<input type="checkbox"/> 239:	MathSci®	1
<input type="checkbox"/> 211:	IAC(SM) Newsearch(TM)	1
<input type="checkbox"/> 13:	Business & Management Practices®	1
<input type="checkbox"/> 1:	ERIC	1



There are 12 databases matching your statement 'S AU=JASCO, P?'.

7.11. ábra: A szerző neve leggyakrabban elírt változatának gyakorisága.

Az Institute for Scientific Information nem vádolható a hivatkozott szerzők nevének elírásával (és a más szempontból pontatlan hivatkozásokkal), ha az eredeti dokumentumban írták el a hivatkozási elemeket. Az előléptetésre váró oktatók aligha engedhetik meg maguknak, hogy nevük 15 rekordban el legyen írva (7.12. ábra) ebben az irigylet adatbázisban - különösen hivatkozott szerzőként, ami fontosabb, mint a közreműködő szerzői szerep. Ez csaknem a 20%-át reprezentálja azoknak a tételeknek, amelyek idézik a szerzőt. Az ember csak találgathatja, milyen arányban írhatják el a valóban hosszú thai vagy különösen a lengyel neveket, amelyekben sok a mássalhangzó.

Annak, hogy a szerző neve alapján nem találunk meg egy rekordot, sokkal súlyosabb következményei lehetnek annál, mint hogy egy tételhez nem jutunk hozzá. Pao (1989) azt tapasztalta, hogy a MEDLINE adatbázisban 4,5%-os a szerzői nevek elírásának (és következetlenségeinek) aránya. Az eredmény világosan bizonyította, hogy az elírások aránya olyan mértékben torzította a szerzők produktivitásának eloszlását, hogy az többé nem felelt meg Lotka eloszlási törvényének. Miután az elírt neveket nagy munkával kijavították, az eloszlás újra megfelelővé vált.

Nyilvánvaló, hogy a szerzők nevének elírásai jelentősen befolyásolhatják az előléptetéseket, kinevezéseket, pályázatokat és az oktatók produktivitás szerinti rangsorolását, ha az adatbázisokból kinyert adatokat ellenőrzés nélkül fogadják el. Azok a személyek, akiknek összetett nevük van, előtag van a nevükben vagy csak kis mértékben is szokatlan a nevük, különösen hátrányban vannak ebből a szempontból. A központosítás különbségei, a nagybetűs írásmód, az előtagok hátravetése szétszórja nevüket az indexben, azokon a problémákon felül, amelyeket az egyszerű nevekkkel is előforduló felcserélések okoznak.

		
Command Search		new search
Database Details		
Set	Term Searched	Items
S1	AU=JASCO P?	7
S2	AU=JACSO P?	105
S3	CA=JASCO P?	15
S4	CA=JACSO P?	71
Show Database Details for: 7: Social SciSearch®  Bluesheet		

7.12. ábra: A Jacsó név elírt és helyesen írt változatai szerzőként és hivatkozott szerzőként a Social SciSearch adatbázisban.

Míg a folyóiratcímek elírása nagyon gyakori, helyesírásuk és rövidítéseik következetlensége még gyakoribb, ezért ezzel a kérdéssel a következő fejezetben foglalkozunk.

Egy terület érdemel különös figyelmet: az eredeti dokumentumok hivatkozásainak pontatlansága. Ennek nyilvánvalóan komoly következménye van minden hivatkozási index számára, s kisebb mértékben minden teljes szövegű adatbázisra, amelyekben szerepelnek bibliográfiák és a jegyzetek. Smith (1981) a hivatkozási hibákkal foglalkozó gondolatébresztő tanulmányában majdnem 20 évvel ezelőtt figyelmeztetett arra, hogy a pontatlan hivatkozások eltorzíthatják a bibliometriai vizsgálatokból levont következtetéseket.

Pandit (1993) a hivatkozások pontosságát vizsgálta a könyvtár- és információtudomány olyan legmagasabban rangsorolt tudományos folyóirataiban, mint például a *College & Research Libraries*, *Library Resources & Technical Services*, *Library Quarterly*, *Library Trends* és a *Journal of Academic Librarianship*. Összesen 131 cikk 1094 hivatkozását vizsgálta. 193 hivatkozásban 223 hibát talált. A *Library Resources & Technical Services* és a *College & Research Libraries* című lapokban volt a legtöbb hiba (31,6% és 27,2%), míg a *Library Trends*-ben csak 3,8% volt a hibaarány. Ironikus, hogy a *Library Trends* az egyetlen nem lektorált folyóirat (bár ez olyan időszaki kiadvány, amelyben csak felkérésre lehet szerepelni, s a legrangosabbak közé tartozik). A *Library Trends* a szokásos hibaarányal kap kéziratokat, de kivételesen gondos szerkesztéssel kijavítják a hibák többségét a végső nyomtatás előtt, eszerint a vizsgálat szerint 92%-ukat.

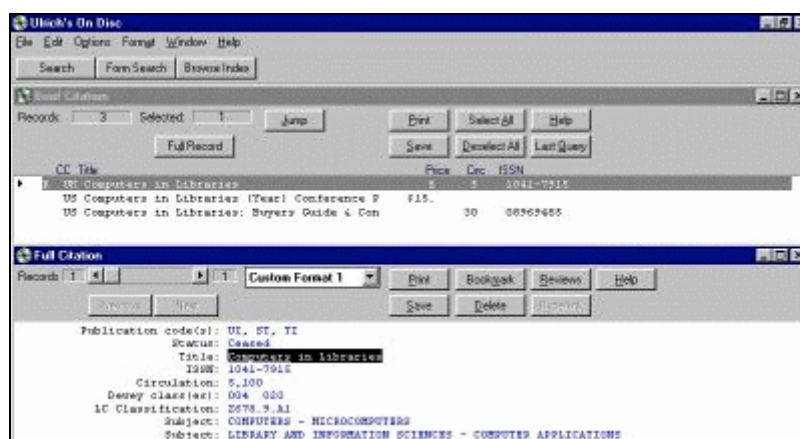
Jelen könyv szerzőjének azt a hipotézisét, hogy a lektorált folyóiratok jobban figyelnek arra, hogy a szerzők igazodjanak lapjuk hivatkozási stílusához, mint hogy elérjék a hivatkozások tartalmának pontosságát, Pandit kutatásai megerősíteni látszanak. Természetesen a hivatkozási hibák egyes típusainak különböző súlya van. Ez a vigasz számunkra, könyvtárosok és információs szakemberek számára, akik láthatólag ugyanynyi hibát követünk el a hivatkozásokban, mint az orvostudományi szerzők. A mi hibáink azonban Benning és Speer (1993)

tanulmánya szerint valamivel kevésbé súlyosak. 555 hivatkozást vizsgáltak a *Library Trends*, *Library Resources & Technical Services* és a *Bulletin of the Medical Library Association* cikkeiben. Az 555 hivatkozás közül 525 elemzésének eredményeit összehasonlították az orvosi folyóiratok hivatkozási hibáival, és a hibaarány csaknem azonos volt a két csoportban (28% és 29%). A jelentős különbség az volt, hogy a hivatkozási hibáknak csak 2%-a tartozott a súlyos hibák közé a könyvtár- és információtudományi folyóiratokban, míg 7% volt az orvostudományi folyóiratok esetében.

Moed és Vriens (1989) öt orvosi folyóirat 4500 cikkének 25000 hivatkozását vizsgálta meg, s úgy találta, hogy a hivatkozások csaknem 10%-a tartalmazott legalább egy hibát. Az ilyen méretű elemzés túl van az egyszerű halandók kapacitásán, de a tanulság figyelmet érdemel. Érdekes, hogy mind Sweetland (1989) (aki gazdagon illusztrált szemlét készített a hivatkozási hibák irodalmáról), mind Moed és Vriens arra a következtetésre jutott, hogy a pontatlan hivatkozások gyakran arra a tényre vezethetők vissza, hogy a hivatkozó dokumentum szerzője nem látta a hivatkozott dokumentumokat.

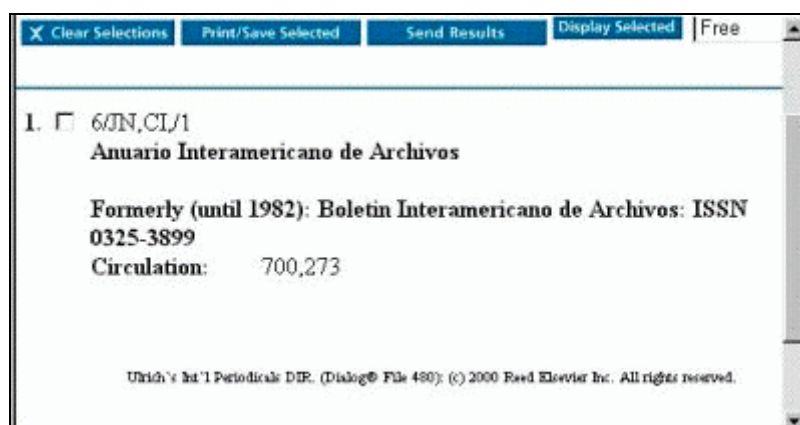
A pontatlanságnak egy formája az is, ha nem igazodnak a nemzeti és nemzetközi katalogizálási szabályokhoz. Boissonnas (1979) 151 LC MARC rekordot és 150 OCLC rekordnak a tagok által beküldött inputját elemezte, amelyeket a Cornell Law Library katalogizálóinak módosítani kellett. Az LC rekordok csupán 29%-át és az OCLC tagok rekordjainak mindössze 1,3%-át találták olyannak, amelyek megfeleltek az AACR2, az ISBD és a publikált LC szabályok értelmezésének.

A címjegyzék adatbázisokban a hibák megtalálása nehezebb, ha nem ismerjük a szakterületet, hacsak nem nyilvánvalóak a hibák első pillantásra vagy az adatok nem tartalmaznak önellentmondásokat. Ez a helyzet a G. K. Saur cég World Databases adatbázisával, amelyben még a dokumentáció mintarekordjai is nonszensz adatokat tartalmaznak, s ugyanez érvényes az összes többi rekordra is, amelyeket e könyv szerzője megvizsgált (Jacsó 1998e) a könyvben említett adatbázisokra vonatkozóan. A ténybeli hibák mennyisége ebben az adatbázisban megdöbbentő. Az még megdöbbentőbb, hogy a jónevű fájl előállító és online szolgáltató továbbra is meglehetősen súlyos árat fizetett a gyanútlan használókkal ilyen hibás adatokért. Az egyetlen vigasz az, hogy - annak ellenére, hogy évenkénti aktualizálást ígértek, 1997 vége, az adatbázis első betöltése óta új rekordokat (és új hibákat) nem adtak hozzá. Általunk jól ismert folyóiratok rekordjainak megnézése komoly pontatlanságokat tárhat fel az időszaki kiadványok címjegyzékeiben. Az Ulrich's például halottnak (megszűntnek) nyilvánította a *Computers in Librariest* és a *Multimedia Schoolst* (az Information Today Inc. két folyóiratát) az adatbázis 1998 őszi kiadásában (7. 13. ábra). 1999 téli kiadásában az Ulrich's mindkettőt visszahozta az életbe, miután rájött arra, hogy azok a téves információkat tartalmazó hírek, amelyek a két folyóirat haláláról szóltak, nemcsak koraiak voltak, de alaptalanok is. A folyóiratok előfizetőire mindez nem volt hatással, de a potenciális előfizetőkre talán igen. Az ember arra is kíváncsi, honnan származhattak a folyóiratok megszűnésére vonatkozó információk. Biztosan nem a kiadótól, amelynek feltételezhetően az információt kellett volna szolgáltatnia a Bowker számára. Nem tudhatjuk, hogy az egész adatbázisban ilyen mértékben vannak-e hibák, de biztosan érdemes a keresőknek mintákat venniük saját szakterületükön.



7.13. ábra: Hibás státusz információ az Ulrich's 1998 őszi kiadásában.

Szisztematikus hibafelderítés lehetséges a címjegyzékekben, legalábbis bizonyos adatelemekre. Az Egyesült Államok államainak kódjai nyilvánvalóan nem kereshetőek vissza egyetlen olyan kóddal sem, amely B, E, J, Q, Z, X vagy Y betűvel kezdődik. Hasonló módon cégek címtárában nem hozhat semmilyen eredményt a U. S. Standard Industry Classification 18-as kódja, mivel ezt a kódot nem adták ki. Az 1400-nál korábbi kiadási évvel rendelkező rekordokra vonatkozó keresés nem hozhat eredményt, kivéve, ha az adatbázisban két számjegyet használnak a kiadási évre vagy hibás adatok vannak ebben a mezőben. A PY>2001 kérdés sem eredményezhet rekordokat, kivéve *Books in Print*et és néhány más adatbázist, amelyben szerepelnek a jövőben megjelenő könyvek és jövőbeni események. Ha indextételek hosszú listáját nézzük meg annak érdekében, hogy megtaláljuk a szerkezeti ellentmondásokat, ez felszínre hozhat problémákat olyan mezőkben, mint az ISSN, ISBN, körzeti hívószámok és amerikai telefon és faxszámok, amelyeknek rögzített hosszúsága vagy mintája van (vagy mindkettő). Extrém adatok keresése a címjegyzékekben mindig jó módszer arra, hogy pontatlanságokat találjunk. A legnagyobb példányszámú könyvtári folyóiratokra való keresés olyan címeket hív elő, amelyeknek a példányszám-adatai nyilvánvalóan hibásak. Nehéz elhinni például, hogy a megszűnt argentin levéltári folyóirat több mint 700000 példányban jelent volna meg (7.14. ábra), pedig az Ulrich's ezt állítja. Ha úgy lett volna, biztosan nem kellett volna beszüntetni a megjelentetését.



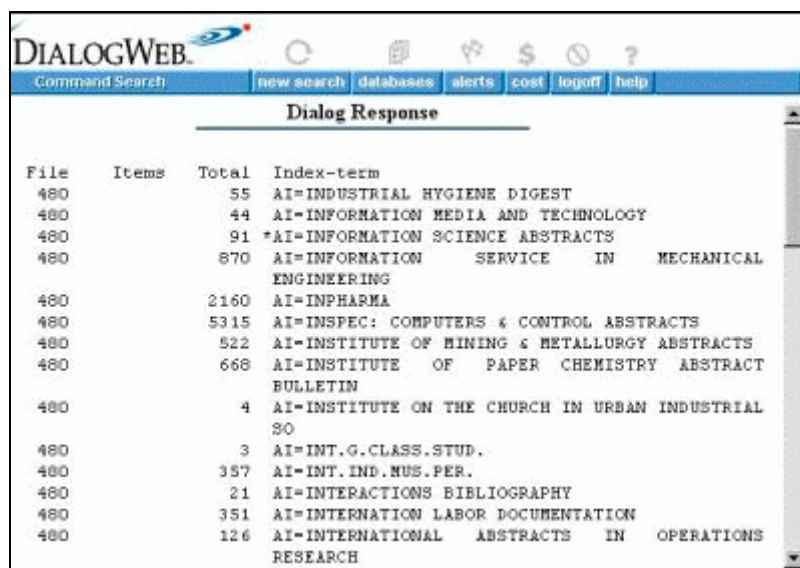
7.14. ábra: Egy argentin levéltári folyóirat valótlanul magas példányszám adatokkal.

Tények nagyarányú és szisztematikus ellenőrzésének másik módja két címjegyzék összehasonlítása, hogy megtudjuk, hogy az eredményeik megegyeznek-e. Jacsó (1991a) 19 referáló és indexelő szolgáltatás feldolgozottságát vizsgálta, ahogy azok az Ulrich'sban és az EBSCO

The Serials Directory című adatbázisban szerepeltek, s tízszeres különbségeket talált a két forrás között. Két évvel később Eldredge (1993) végzett hasonló szélesebb alapú össze-hasonlítást, amelyben szerepelt a National Library of Medicine SERLINE címjegyzéke is, hogy összehasonlítsa a feltárt információ indexelésének pontosságát. Eldredge-nek az a következtetése, hogy „a könyvtárosoknak szkeptikusan kell nézniük ennek a három folyóiratforrásnak az indexelési feltártságra vonatkozó információkat”, nagyon mértéktartó. Még ha az adatok csak egyetlen forrásból jönnek is, a téves információk nyilvánvalóak lehetnek.

Az EBSCO-nak semmilyen információja nincs az ISA-ról, sem rövidített, sem teljes formájában. Az Ulrich'sban van információ az ISA referáló és indexelő kiadványról, és azt állítja, hogy 91 folyóiratot dolgoz fel (7.15. ábra). Az ISA folyóiratbázisa csökkent, de nem ennyire. Az ISA dokumentációja azt állítja, hogy több mint 300 folyóiratot referálnak és indexelnek. Egyik adat sem pontos. Több mint 3500 időszaki kiadvány cím van az ISA-ra vonatkozóan a Journal Name Finder Database-ben (benne sok következtetlen változattal és elírással), amelyeket valamikor 35 éves történetük során indexeltek és referáltak. Az elmúlt néhány évben a feldolgozott folyóiratok száma körülbelül 230 volt.

A két időszaki kiadványokat feldolgozó címjegyzék közül egyikben sincs információ a Mental Health Abstractsról, és csak az Ulrich's tartalmazott információt a Microcomputer Abstractsról mint referáló és indexelő folyóiratról, az is régi címe, a Microcomputer Index alatt, amely 1994-ig volt érvényes. A cím 2000 elején újra megváltozott, Microcomputer Abstracts helyett Internet & Personal Computing Abstracts lett, de ez nem tükröződik az Ulrich's referáló és indexelő szolgáltatásokat tartalmazó mutatójában, s az EBSCO-t sem ösztönözte arra, hogy szerepeljen referáló és indexelő forrásokat feltáró eszközében. Vajon milyen lehet a minőség-ellenőrzés az időszaki kiadványok címjegyzékeinek aktualizálásakor?



File	Items	Total	Index-term
480		55	AI=INDUSTRIAL HYGIENE DIGEST
480		44	AI=INFORMATION MEDIA AND TECHNOLOGY
480		91	*AI=INFORMATION SCIENCE ABSTRACTS
480		870	AI=INFORMATION SERVICE IN MECHANICAL ENGINEERING
480		2160	AI=INPHARMA
480		5315	AI=INSPEC: COMPUTERS & CONTROL ABSTRACTS
480		522	AI=INSTITUTE OF MINING & METALLURGY ABSTRACTS
480		668	AI=INSTITUTE OF PAPER CHEMISTRY ABSTRACT BULLETIN
480		4	AI=INSTITUTE ON THE CHURCH IN URBAN INDUSTRIAL SO
480		3	AI=INT.G.CLASS.STUD.
480		357	AI=INT.IND.MUS.PER.
480		21	AI=INTERACTIONS BIBLIOGRAPHY
480		351	AI=INTERNATIONAL LABOR DOCUMENTATION
480		126	AI=INTERNATIONAL ABSTRACTS IN OPERATIONS RESEARCH

7.15. ábra: Az ISA-ban indexelt folyóiratok száma az Ulrich's szerint.

Hasonlóan árulkodó az a tény, hogy a két kereskedelmi forgalomban elérhető időszaki kiadvány címjegyzék nem tud egyetérteni olyan fontos kérdésekben, mint hogy melyek a lektorált folyóiratok egy adott diszciplínában. A zavar első jele az, hogy a címjegyzékekben feltűnő különbségek vannak a lektorált folyóiratok számában bizonyos szakterületeken. Kisebb különbségek adódhatnak az eltérő osztályozás miatt (például egy lektorált folyóiratot az egyik címjegyzékben a szociológiába, a másikban pedig a pszichológiába sorolhatnak. A különbségek azonban hatalmasak olyan diszciplínákban is, ahol a besorolás egyszerű és egyértelmű lehet, s mindkét címjegyzék azonos vagy majdnem azonos tárgyi fogalmakat használ (7.16. ábra).

Első pillantásra úgy látszik, hogy különbség a lektoráltnak jelzett folyóiratok száma közötti különbség (11407 a The Serials Directoryban és 13124 az Ulrich'sban) összhangban áll a folyóiratok teljes száma közötti különbséggel. Az azonos kategóriákban található számok azonban furcsa különbségeket tárnak fel. Az Ulrich's kétszer annyi lektorált folyóiratot jelez az Oceanography (óceánográfia) és majdnem háromszor annyit Water Resources (vízi erőforrások) kategóriában, mint az EBSCO, míg az utóbbiban az Anthropology (antropológia) területén van 30%-kal több lektorált folyóirat, mint az Ulrich'sban.

Ezeket az adatokat a két CD-ROM-os címjegyzék 1997 őszi kiadása alapján sikerült megállapítani, s a számok folyamatosan változnak. Két évvel később az EBSCO-ban 106 lektorált folyóirat volt (a korábbi 127-ről lecsökkent) a könyvtár- és információtudomány területéről, míg az Ulrich's 194 címet jelentett ilyenként (ennyire nőtt a korábbi 131-ről). A lektorált folyóiratok teljes száma az Ulrich'sban az 1997-es 13124-ről 20000 fölé emelkedett 2000 végére, - ami valószínűtlen növekedés. Hasonló különbségek mutatkoznak a lektorált folyóiratok számában, amikor országok (7.17. ábra) vagy kiadók (7.18. ábra) szerint korlátozzuk őket. Az adatok Ausztriára, Svédországra, Dániára, Norvégiára és Ausztráliára vonatkozóan a legfurcsábbak az országok szerinti összehasonlításban, a kiadók között pedig a John Wiley, Allerton Press és az Elsevier esetében.

	TSD	TSD-r	ULR	ULR-r	TSD/ULR	TSD-r/ULR-r
All records	187,161	11,407	218,002	13,124	0.86	0.87
LIS	2,514	127	2,633	131	0.95	0.97
Anthropology	737	197	1,118	151	0.66	1.30
Arch(a)eology	1,090	84	1,636	135	0.67	0.62
Oceanography	409	41	603	87	0.68	0.47
Geograph\$	1,037	90	1,236	112	0.84	0.80
Linguistic\$	2,388	200	2,799	276	0.85	0.72
Music\$	2,545	78	3,316	106	0.77	0.74
Pharmac\$	1,505	180	1,762	291	0.85	0.62
Publishing	1,348	21	1,583	29	0.85	0.72
Veterinary\$	734	105	619	105	1.19	1.00
Water Res\$	644	24	783	66	0.82	0.37

7.16. ábra: Jelentős különbségek vannak a lektorált folyóiratok szakterületenkénti számában.

	TSD	TSD-r	ULR	ULR-r	TSD/ULR	TSD-r/ULR-r
Hungary	611	25	763	29	0.81	0.86
Portugal	284	3	602	8	0.47	0.38
Austria	789	39	1,826	17	0.43	2.29
Sweden	921	46	2,233	52	0.41	0.88
Denmark	793	74	2,700	108	0.29	0.69
Finland	786	20	870	24	0.90	0.83
Norway	425	35	1,047	26	0.41	1.35
Australia	4,350	337	6,228	257	0.70	1.31
Canada	17,912	504	9,852	379	1.82	1.33
UK	15,263	1,647	21,995	2,045	0.69	0.81
USA	88,956	5,456	81,562	6,045	1.09	0.90

7.17. ábra: Jelentős különbségek vannak a lektorált folyóiratok országonkénti számában.

	TSD	TSD-r	ULR	ULR-r	TSD/ULR	TSD-r/ULR-r
Wiley	803	158	573	295	1.40	0.54
McGraw	671	5	558	4	1.20	1.25
Haworth	212	135	171	152	1.24	0.89
Ablex	64	8	94	13	0.68	0.62
Academic	553	157	465	222	1.19	0.71
Allerton	136	2	102	11	1.33	0.18
AMA	53	10	49	10	1.08	1.00
Blackwell	505	177	471	258	1.07	0.69
Bowker	190	3	207	1	0.92	3.00
Elsevier	1,734	927	1,748	1,364	0.99	0.68

7.18. ábra: Jelentős különbségek vannak a lektorált folyóiratok kiadónkénti számában.

Még abban az esetben is, ha a két címjegyzék majdnem azonos adatokat mond a lektorált folyóiratok számáról egy adott diszciplínában, az átfedés a kettő között elszomorítóan alacsony, figyelembe véve azt a fontosságot, amit a folyóiratok lektorált státuszához kapcsolnak sokan az egyetemi körökben. A könyvtár- és információtudományi kategóriában például az Ulrich's még mindig nem sorol a lektorált címek közé olyanokat, mint a *College & Research Libraries* (7.19. ábra), a *The American Archivist* vagy a *Library Resources & Technical Services*. Mindkét cég azt állítja, hogy adataikat a kiadóktól nyerik. Valószínűtlen azonban, hogy a kiadók a lektoráltság státuszát különböző módon jelenítsék a két kérdőívben.

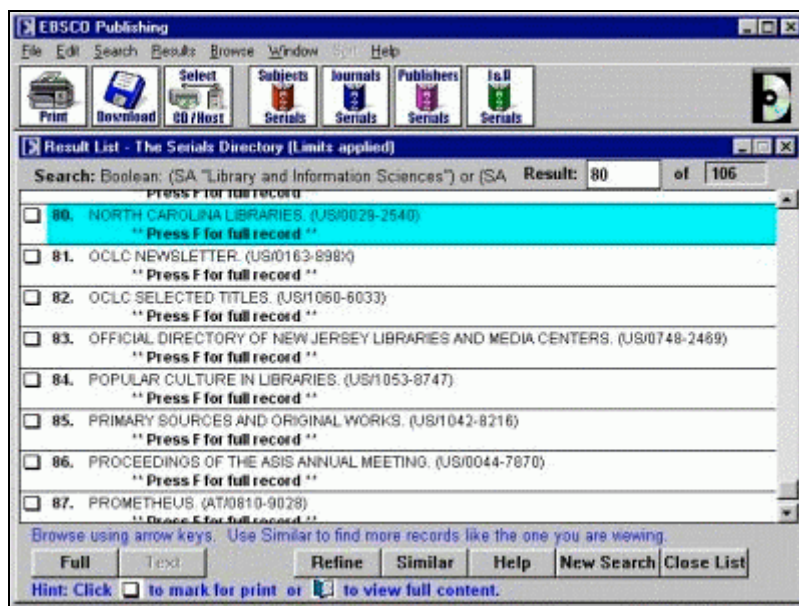
Jól kell ismerniük a szakterületet ahhoz, hogy észrevegyék egyes címek jelenlétének és mások hiányának abszurditását a lektorált folyóiratok listájáról. Magától értetődően nem lehet lektorált folyóirat a *Montana Library Directory* és az *OCLC Selected Titles*, ahogy azt a *Serials Directory* állítja (7.20. ábra). Hasonló módon megdöbbentő egy könyvtáros számára, hogy az EBSCO nem lektorált periodikaként határozott meg olyan címeket, mint az *Information Processing & Management*, az *Information Services & Use*, a *Journal of Government Information*, a *Knowledge Organization* és a *Serials Review*.

Results of your search: from 3 keep 1-194
Citations available: 194
Citations displayed: 31-40
<input checked="" type="checkbox"/> Citations in "Titles Display" format
Go to ... Help Logout
<input type="checkbox"/> 31. Bokasafnid [Academic/Scholarly Publication] Complete Reference
<input type="checkbox"/> 32. Books at Iowa [Academic/Scholarly Publication] Complete Reference
<input type="checkbox"/> 33. Books in Library and Information Science Series Complete Reference
<input type="checkbox"/> 34. Byblios [Academic/Scholarly Publication; Trade Publication] Abstract Complete Reference
<input type="checkbox"/> 35. Canadian Library Journal Abstract Complete Reference
<input type="checkbox"/> 36. Cataloging & Classification Quarterly [Academic/Scholarly Publication] Abstract Complete Reference
<input type="checkbox"/> 37. Catholic Library World [Academic/Scholarly Publication] Complete Reference
<input type="checkbox"/> 38. Choice (Middletown) [Academic/Scholarly Publication] Abstract Complete Reference
<input type="checkbox"/> 39. Collection Management [Trade Publication] Abstract Complete Reference
<input type="checkbox"/> 40. College & Undergraduate Libraries [Academic/Scholarly Publication] Abstract Complete Reference

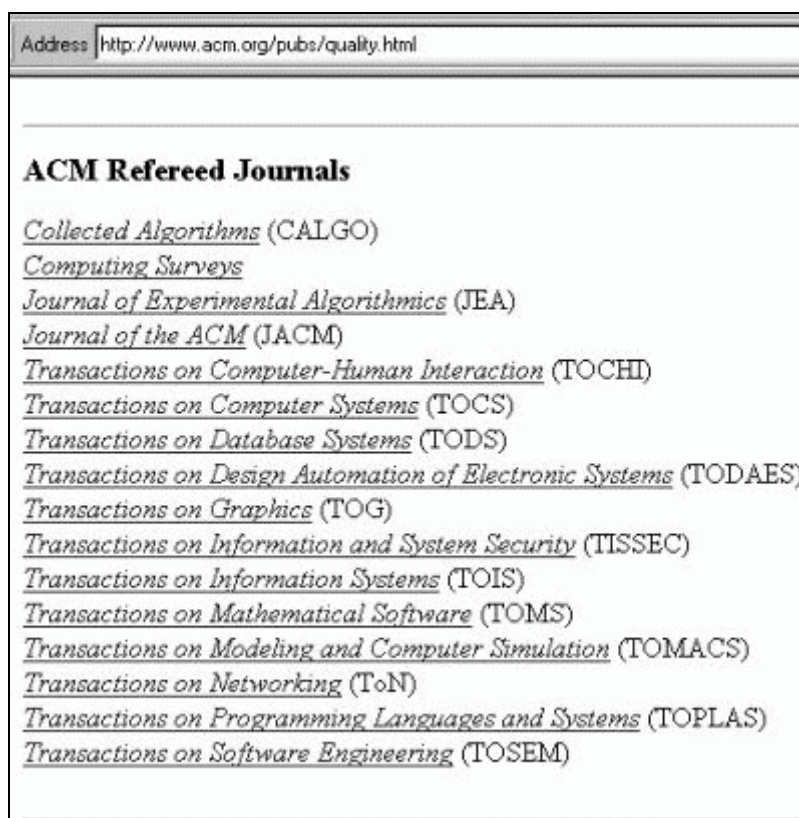
7.19. ábra: Lektorált könyvtár- és információtudományi folyóiratok listája az Ulrich's-ban (részlet)

Néha az ilyen listák összehasonlíthatók a legautentikusabb listákkal; azokkal, amelyeket a maguk a kiadók állítanak össze. Az Association for Computing Machinery (ACM) honlapján 16 folyóiratot határozott meg lektoráltként (7.21. ábra). Az Ulrich's ezek közül csak négyet

tart lektoráltnak (7.22. ábra). Ez feltétlenül fejlődés ahhoz a két évvel korábbi helyzethez képest, amikor csak egyetlen ACM folyóiratot minősítettek lektoráltnak. A Serials Directory csak nyolc ACM címet határozott meg pontosan lektoráltként, de ugyanakkor két címet tévesen tart annak: a *Communication of ACM*-et és a *Data Base*-t (7.23. ábra).



7.20. ábra: Lektorált könyvtár- és információtudományi folyóiratok listája a The Serials Directoryban (részlet)



7.21. ábra: Az ACM lektorált folyóiratai, ahogy azokat az ACM meghatározta.

REFEREED JOURNALS OF ACM ULRICH'S DECEMBER, 1999		

RANK: S3/1-4 Field: JN= File(s): 480		
(Rank fields found in 4 records -- 5 unique terms)		
RANK No.	Items	Term
1	1	A C M TRANSACTIONS ON GRAPHICS
2	1	A C M TRANSACTIONS ON MATHEMATICAL SOFTWARE
3	1	SHORT TITLE: T O C S
4	1	TRANSACTIONS ON COMPUTER SYSTEMS
5	1	TRANSACTIONS ON DESIGN AUTOMATION OF ELECTRONI

7.22. ábra: Az ACM lektorált folyóiratai, ahogy azokat az Ulrich's meghatározta.

Print	Full Text	Select Result For More Detail
<input type="checkbox"/>	<input type="checkbox"/>	ACM COMPUTING SURVEYS (US/03600300)
<input type="checkbox"/>	<input type="checkbox"/>	ACM JOURNAL OF EXPERIMENTAL ALGORITHMS [COMPUTER FILE] ACM JOURNAL OF EXPERIMENTAL ALGORITHMS (US/10846654)
<input type="checkbox"/>	<input type="checkbox"/>	ACM TRANSACTIONS ON COMPUTER SYSTEMS (US/07342071)
<input type="checkbox"/>	<input type="checkbox"/>	ACM TRANSACTIONS ON DATABASE SYSTEMS (US/03625915)
<input type="checkbox"/>	<input type="checkbox"/>	ACM TRANSACTIONS ON GRAPHICS (US/07300301)
<input type="checkbox"/>	<input type="checkbox"/>	ACM TRANSACTIONS ON MATHEMATICAL SOFTWARE (US/00983500)
<input type="checkbox"/>	<input type="checkbox"/>	ACM TRANSACTIONS ON PROGRAMMING LANGUAGES AND SYSTEMS (US/01640925)
<input type="checkbox"/>	<input type="checkbox"/>	COMMUNICATIONS OF THE ACM (US/00010782)
<input type="checkbox"/>	<input type="checkbox"/>	DATA BASE (US/00950033)
<input type="checkbox"/>	<input type="checkbox"/>	JOURNAL OF THE ASSOCIATION FOR COMPUTING MACHINERY (US/00045411)

7.23. ábra: Az ACM lektorált folyóiratai, ahogy azokat a The Serials Directory meghatározta

A két címjegyzéket összehasonlító keresés végrehajtása valószínűleg hasonló különbségeket tárna fel más diszciplínákban is. Az ilyen összehasonlítások fényt derítenek a folyóiratoknak arra a két jellemzőjére is, amelyeket értéknövelt információkként határoz meg a két címjegyzék: a folyóirat példányszámára és az előfizetési árra. Ezeket az információkat feltételezhetően azért gyűjtötték és jelentették, mert a tájékoztatásban, gyűjteményfejlesztésben és bizonyos üzleti döntésekben használják őket, mint például a hirdetési díjak meghatározására.

Nyilvánvaló, hogy az nem szolgálja jól ezeket a célokat, ha jelentős eltérések vannak a két címjegyzék között. A különbségek magukért beszélnek. Egyeseket könnyű ellenőrizni a nyomtatott időszaki kiadványban szereplő vagy a kiadók katalógusaiban hozzáférhetővé tett információk alapján. Az ilyen összehasonlítás kiterjeszthető más adatelemekre is, de feltételezhetően ezek a legfontosabbak.

Időnként az adatbázisok kiadói okozzák az adatbázisok pontatlanságait. Ez az oka az egyik legfurcsább, legnagyobb mértékű hibának, amellyel az Ulrich's DIALOG-os változatában találkozhatunk. Ha az *Excerpta Medica* (az egyik legnagyobb referáló és indexelő vállalkozás) kezdetű folyóiratcímeket keressük, 49 folyóiratot találunk, mint például az *Excerpta Medica*, Section 16: Cancer. Az *Excerpta Medica* által indexelt és referált folyóiratokra vonatkozó keresés azonban nulla találatot ad (7.24. ábra). Az input feldolgozásakor az adatbázis tervezője minden bizonnyal tévesen, az *Exerta Medicaként* elírt formában adta meg az eredeti kód (mondjuk EXM) feloldását. Ezért van az, hogy 5654 rekord van az elírt névalak alatt (7.25. ábra). Az Ulrich's Ovidos változata 5630 rekordban helyesen határozza meg a referáló és indexelő szolgáltatás nevét (7.26. ábra).

DIALOGWEB
Command Search new search databases

Search History
[Database Details](#)

Set	Term Searched	Items
S1	JN=EXCERPTA MEDICA?	49
S2	AI=EXCERPTA MEDICA?	0

Show Database Details for:

7.24. ábra: Olyan folyóiratok keresése, amelyeket az *Excerpta Medica* referál és indexel a DIALOG-nál.

E5	5654	AI=EXERTA MEDICA
E6	1	AI=EXERTA MEDICA (UNTIL 1983)
E7	20	AI=EXERTA MEDICA (UNTIL 1992)
E8	1	AI=EXERTA MEDICA (UNTIL 1992; 1994-)
E9	2	AI=EXERTA MEDICA (UNTIL 1992; 1995-)
E10	50	AI=EXERTA MEDICA (UNTIL 1993)
E11	2	AI=EXERTA MEDICA (UNTIL 1993; 1996-)
E12	2	AI=EXERTA MEDICA (UNTIL 1994)
E13	205	AI=EXERTA MEDICA (1993-)
E14	124	AI=EXERTA MEDICA (1994-)
E15	120	AI=EXERTA MEDICA (1995-)
E16	106	AI=EXERTA MEDICA (1996-)
E17	150	AI=EXPLORATION AND PRODUCTION HEALTH, SAFETY AND
E18	127	AI=EXTRAMED
E19	120	AI=FABABEAN ABSTRACTS
E20	1	AI=FAM.IND..
E21	780	AI=FAMILY INDEX
E22	45	AI=FAMILY MEDICINE LITERATURE INDEX
E23	1	AI=FAMILY MEDICINE LITERATURE INDEX (1980-1992)

7.25. ábra: Referáló és indexelő szolgáltatás hibásan kódolt neve az Ulrich's DIALOG-os változatában.

8 "excerpta medica".ai. 5630 [Display](#)

[Run Saved Search](#) [Save Search History](#) [Delete All Searches](#)

Enter **Keyword** or phrase:

[Perform Search](#)

Limit to:
☐ Active

[Ask A Librarian](#)

Results of your search: "excerpta medica".ai.
 Citations available: **5630**
 Citations displayed: **1-10**
☒ Citations in "Titles Display" format

7.26. ábra: Pontosán kódolt *Excerpta Medica* az Ulrich's Ovidos változatában.

Az adatbázisok címjegyzékeiben egyes pontatlanságok azonnal nyilvánvalóvá válnak a gyakorló szakember számára, ilyen a Social SciSearch adatbázis rekordja a World Databases címjegyzékben. Ha valóban minden héten 780000 rekorddal frissítenék, mérete 15 hét alatt megduplázódna. Ilyen fajta hiba rengeteg van a kivételesen silány World Databases címjegyzékben.

A kevésbé tapasztalt használók számára is egyszerűen megállapíthatóak a pontatlanságok, amikor olyan alapvető információkat hasonlítanak össze a két adatbázisban, mint az adatbázis mérete, az aktualizálás gyakorisága, az aktualizálás mérete, az óradíj, a nyomtatás ára vagy az adatbázis licenc díja (7.27. ábra).

☐ Document 1 of 3

World Databases (WDBS)

Title
 Science Citation Index
 (Alternative title(s): SCISEANEW).

Database name
 File 34.

Coverage
 Year started: 1988
 Coverage: International.

Updating
 Weekly, 780,000 records per update.

COPYRIGHT BY Bowker-Saur Abstr. & Indexes, E. Grinstead, UK

7.27. ábra: Nyilvánvaló pontatlanság a G. K. Saur World Databases adatbázisában.

Néha nem lehet a címjegyzék kiadóját hibáztatni a pontatlanságok miatt, mivel azok a primér forrásokból származnak. Ez okozott némi problémát a könyvtárosoknak, amikor a Sage Publications a *Journal of Contemporary Ethnography* 1999 februári számát a kötet első számaként határozta meg, pedig maga a folyóirat azt állítja, hogy áprilisban, júliusban, októberben és januárban jelennek meg (7.28. ábra). Ez csak kiegészítette azt a zavart, amit a Sage-nek az a terve okozott, hogy évente hat számot ad ki, de mivel ennek a megjelenési gyakoriságnak az elhatározása nem végleges, a könyvtárakat nem tájékoztatták róla (7.29. ábra).

Tárgy: Journal of Contemporary Ethnography

Nincs jó napom.

Ennek a lapnak a februári száma egy volt a problémák halmazában, egy üzenettel egy hallgatómtól, hogy a hónap megjelölése nem felel meg a szokásos megjelenési időnek. Így hát megnéztem belül, s azt találtam: „Megjelenik évente négyszer - áprilisban, júliusban, októberben és januárban.”

Mit csinálnak ezzel a többiek, s ha nem verem tovább a fejem a falba, elmúlik a fájdalom?

From: Jeanette Skwor **To:** SERIALST@LIST.UVM.EDU
Subject: Journal of Contemporary Ethnography
 Subject: Journal of Contemporary Ethnography

I am not having a good day.

The February issue of this title was on my Problem Pile with a note from my student, that the month does not match the pattern. So I looked inside, and found: "Published four times annually -- in April, July, October and January."

What are others doing with it, and if I stop banging my head against the wall now will the pain go away?

Jeanette Skwor
 Cofrin Library
 UW-Green Bay
SKWORJ@GBMS01.UWGB.EDU

7.28. ábra: E-mail a zavaró kronológiai jelzéssel kapcsolatban.

Tárgy: Re: Journal of Contemporary Ethnography

Felhívtam a Sage Publicationst, s azt mondták nekem, hogy a „Journal of Contemporary Ethnography” 1999-től kéthavi folyóirat lett. Így a februári szám az első ebben az évben. Azonban amikor megkérdeztem, hogy a Sage kiküldene-e levelet, informálva előfizetőit erről a változásról, az operátor nem tudta. Miközben segített nekem, hallottam mormogását, hogy a Sage számítógépeit nem aktualizálták teljesen, hogy az új hónapokat megmutathassa. Remélem, hogy ez segíteni fog. Kérlek, ne verd tovább a falba a fejed. Helyette írd fel a különböző kiadók nevét egy céltáblára, amelyek ilyen csinálnak, s azután lödd ki nyilaidat.

From: Sarah Peadar, Serials **To:** SERIALST@LIST.UVM.EDU
Subject: Re: Journal of Contemporary Ethnography

I called Sage Publications and was told that "Journal of Contemporary Ethnography" had been upgraded to a bimonthly publication as 1999. So, February is the first issue for this year. However, when I asked if Sage would be sending out a letter informing its subscribers of this change, the operator did not know. While she was helping me, I heard her murmur that Sage's computers had not been completely updated to show the new months. I hope this helps. Please stop banging your head against the wall. Instead, make a bulls eye for the different publishers that do this and throw darts at it. ;)

7.29. ábra: E-mail a Sage hibás információiról egyik kiadványuk megjelenési gyakoriságával kapcsolatban.

A könyvtárosok és az információs szakemberek érzik annak a terhét, hogy felelősek az ügyfelek számára adott információkért. Felelőtlenség hivatkozni akár az Ulrich's International Periodicals Directory, akár a Serials Directory adatára egy folyóirat lektoráltsági státuszával kapcsolatban egy egyetemi pozícióról folyó vitában. Törlési döntések meghozása a két adatbázis által nyújtott információk alapján arról, hogy melyik referáló és indexelő folyóirat tartalmaz meghatározott periodikumokat, hasonló felelőtlenség. Ha nem lehet megtalálni egy adatbázisban egy állásra jelentkező írásait amiatt, mert nevét gyakrabban írják ott hibásan, mint helyesen, eldöntheti, hogy felveszi-e vagy sem. Nyilvánvalóan hibás adatok

szolgáltatása adatbázisok ezreiről a World Databases ügyfelei számára hasonló az orvosi műhibához. Ilyen adatokért rendkívül magas árakat fizettetni erkölestelen dolog.

Sok olyan könyvtáros van, aki erőteljesen kritizálja az ingyenes webes adatbázisokat hiányosságaik miatt. Igazuk van abban, hogy ezt teszik, de tudniuk kell, hogy a nagyhírű hagyományos információszolgáltatók előkelő székházaikból gyakran sokkal hibásabb adatbázisokat szolgáltatnak csinos kis summáért. Pontos információkat tartalmazó adatbázisok összeállítása sokba kerül, ennek az árát kevés tartalomszolgáltató akarja megfizetni.

Az H. W. Wilson azon ritka cégek közé tartozik az információiparban, amelyek különösen kényesek arra, hogy ellenőrizzék információikat és visszamenőlegesen is korrigálják a téves adatokat, ha a hibákra vagy változásokra rájönnek. Ez biztosan sokkal többbe kerül, mint PR szakembereket vagy dumamatyikat alkalmazni, akik hamis portékát árulnak, kifogásokat és magyarázatokat találnak adatbázisaik sok hiányosságára.

Ezek a praktikák nem segítenek azoknak a könyvtárosoknak, akik néha maguk is végfelhasználók, vagy pedig közvetlenül szembe kell nézniük a végfelhasználókkal. Drága adatbázisok nagyfokú és súlyos pontatlanságai a könyvtáros legnagyobb ellenségei közé tartoznak. Először meg kell találni ezeket, majd a szolgáltatókat szembesíteni kell velük.

8. A formátum és a tartalom következetessége

Egy tökéletes világban szinte minden adatelemet úgy vinnének be egy adatbázisba, hogy kontrollált szótárt vagy valamilyen szabványosított név - vagy rövidítés - listát használnak. Még akkor is lennének különbségek az adatbázisok között a személyek és cégek neveiben, a folyóiratok címének vagy a deskriptoroknak a formájában, de legalább egy adatbázison *belül* a szavak helyesírása, a folyóiratcímek és a cégnevek rövidítései megjósolhatóak lennének. Cserében ez megkönnyítené a használóknak azt, hogy megtaláljanak minden olyan cikket, amelyet egy adott folyóiratról származik, minden hivatkozást, amely egy cégre vonatkozik, és a hírekbe a legkülönbözőbben transliterált formában bekerülő híres vagy hírhedt személyek minden említését. Gondoljunk csak Líbia vezetője, Thaiföld miniszterelnöke vagy Észak-Korea fővárosa nevének változataira.

Ezeknek a változatoknak némelyike a transliterációs szabályok változásainak köszönhető. Ezek kereszthivatkozások segítségével megtalálhatóak egy böngészhető listában. Az ilyen megoldások valóban jól működnek az H. W. Wilson adatbáziscsalád CD-ROM-os változataiban vagy a PAIS-ban, s kisebb mértékben az EBSCO adatbázisaiban. Az online világban azonban általában nincsenek kereszthivatkozások egy név egyik formájáról a másikra - azon néhány adatbázis deskriptorainak kivételével, amelyeknek van online tezaurusza. (Az indexelés minőségét, benne az indexelés következetességét a 10. fejezetben tárgyaljuk. A folyóiratok és más dokumentumtípusok feldolgozásának következetességét az 5., a feltárt forrásokkal foglalkozó fejezetben tárgyaljuk.

A következetesség megvalósítása különösen szükséges olyan adatbázisokban, amelyek sok különböző forrásdokumentumot használnak, s amelyeknek gyakran megvannak a maguk házi szabványai. Feltételezhető, hogy a *Washington Post*nak van ilyen szabványa; ezért egyetlen online fájlban belül az országok, személyek, mozgalmak és cégek neve következetes formában jelenik meg. Azonban azok, akik több indexelő, referáló és teljes szövegű adatbázist készítenek (mint például a Gale Group, Bell Howell és EBSCO), szabványosítaniuk kell ezeket a neveket, s ez alapján következetesen alkalmazniuk az újságok, folyóiratok és magazinok ezreiből alkotott rekordokhoz. Sok adatfájl előállítónak megvan papíron a maga egységesített névalak listája, de az adatok bevitelekor láthatólag nem vetik össze a tételeket ezekkel a listákkal.

Az H. W. Wilson cég egyike annak a kevés tartalomszolgáltatónak, amely elérte a következetesség szinte kifogástalan szintjét. Érdekes módon egy ingyenes, weben született adatbázis, az Internet Movie Database egy másik teljesen tiszta és következetes forrás, amely szintén szolgáltató kereszthivatkozásokat a címváltozatokból és a művészi álnevekből. Ezt olyan módon teszik, amit tanítani kellene a legjobb iskolákban is.

Az elírásokhoz és az egyéb pontatlan adatelemekhez hasonlóan, a következetlenség is azt okozhatja, hogy igazán releváns tételeket veszünk el egy adatbázisban történő keresés során. Azok a használók, akik defenzív keresési stratégiákat alkalmaznak, minimalizálhatják a következetlenség hatását, de sok idő kell a különböző változatok böngészéséhez és kereséséhez, ez pedig pénzbe kerül, különösen olyan rendszerekben, amelyek díjainak még mindig jelentős részét képezi a kapcsolati idő. Sok információkereső rendszer nem nyújt megfelelő böngészési lehetőséget, amely pedig csökkenteni tudná a következetlenség néhány problémáját.

Mezőspecifikus indexek (szerzői index, kiadói index, folyóiratcím index) böngészése és a bennük való keresés fényt derít a következetlenségekre. Azok a rendszerek, amelyek csak nagy, mindent együtt tartalmazó indexet szolgálnak a böngészéshez, mint a SilverPlatter

DOS-os változata, elrejtik a következetlenségeket. Nem nyújt segítséget az, ha olyan indextételeink vannak, amelyeket számos mezőből generálnak anélkül, hogy lennének megkülönböztető prefixeik (előtagjaik), amelyek meghatároznák, hogy a tételt melyik adat-elemből vették. Nem segít, ha nem böngészhetünk a nyelv, dokumentumtípus vagy a kiadás országa mezőjében.

Például a Sociological Abstractsben a román nyelv *Romanian* formában szerepel (8. 1a. ábra), míg a kiadás országa *Rumania* (8. 1b. ábra), az országra vonatkozó deskriptor pedig *Romania* (1985-től) és *Rumania* (1984-ig) formában. Egyik sem pontatlan, de a tételek kétségtelenül következetlenek, és a SilverPlatter DOS-os változatában soha nem szerzünk tudomást róluk. Szerencsére az adatbázisok DIALOG-os változatában (és sok másokban is) a nyelvek és az országok indexe böngészhető. Néhány online teaurusz is böngészhető, hogy megtaláljuk a kívánt tárgyi kifejezéseket és a szélesebb, szűkebb és kapcsolódó kifejezéseket (8. 1c. ábra).

new search databases alerts cost logoff help					
Dialog Response					
Ref	File	Items	Total	Index-term	
E1	37		2606	LA=PORTUGUESE	
E2	37		2952	LA=ROMAN	
E3	37		608	*LA=ROMANIAN	
E4	37		7216	LA=RUSSIAN	
E5	37		1105	LA=SERBIAN	
E6	37		4057	LA=SERBO	
E7	37		1105	LA=SERBO-CROATIAN WITH CYRILLIC ALPHABET/SERBIAN	
E8	37		2952	LA=SERBO-CROATIAN WITH ROMAN ALPHABET/CROATIAN	

8. 1a. ábra: A nyelvi index kifejezései a Sociological Abstractsben.

Dialog Response					
Ref	File	Items	Total	Index-term	
E1	37		238	CP=RICA	
E2	37		248	CP=RICO	
E3	37		530	*CP=RUMANIA	
E4	37		1946	CP=RUSSIA	
E5	37		55	CP=SALVADOR	

8. 1b. ábra: A kiadás országa indexkifejezései a Sociological Abstractsben

Dialog Response					
Total	Type	RT	Index-term		
535		5	*ROMANIA		
0	S		(FORMERLY (1964-1985)	DC	397650,
			RUMANIA/R...)		
333	F	1	RUMANIA		
116	B	8	BALKAN STATES		
1166	B	12	EASTERN EUROPE		
12	N	2	BUCHAREST, ROMANIA		

8. 1c. ábra: Deskriptorok a Sociological Abstracts teauruszában.

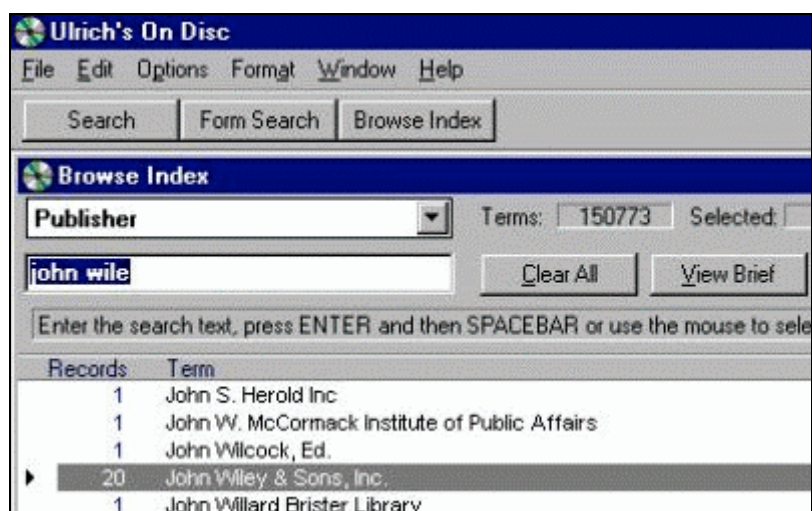
Bár a DIALOG sok mezőspecifikus böngészhető indexet ajánl, nincs külön böngészhető deskriptor index, például DE=, amely segítene a használóknak, hogy kiválasszák a legmegfelelőbb fogalmakat (különösen akkor, ha a teaurusz online változatát nem készítették el, hogy útmutatást adhasson), és megismerjék a pontatlanságokat és következetlenségeket. Másik oldalról az Ovid minden lehetséges indexelt adatelemet böngészhetővé tesz, s egyedi módon, lehetővé teszi a használóknak, hogy használat közben kombinált ad-hoc indexeket alkossanak. Például a Sociological Abstracts Ovid-os változatában a használó létrehozhatja a nyelv, a kiadás országa és a deskriptorok kombinált indexét. Az eredmények listájában a nyelv, a kiadás országa és a deskriptor mező különböző kifejezései egy közös fájlba kerülnek.

Minden adatbázis szerzői indexének böngészésekor megérezhetjük a következetlenség szintjét. Az elírások és a következetlenül írt formák gyakran egymáshoz közel találhatók. A Bowker cég Ulrich's adatbázisában számos következetlen formát használ a saját nevére (8.2. ábra). Ezek közül egyesek szabályosak, de sok csak a testületi nevek egységesített fájljának - és némi odafigyelésnek - a hiányát mutatja. Megjegyzendő, hogy sok további változat volt évekkel ezelőtt (mint például R. R. Bowker, RR Bowker és R R Bowker), amelyeket nagyrészt megszüntettek.

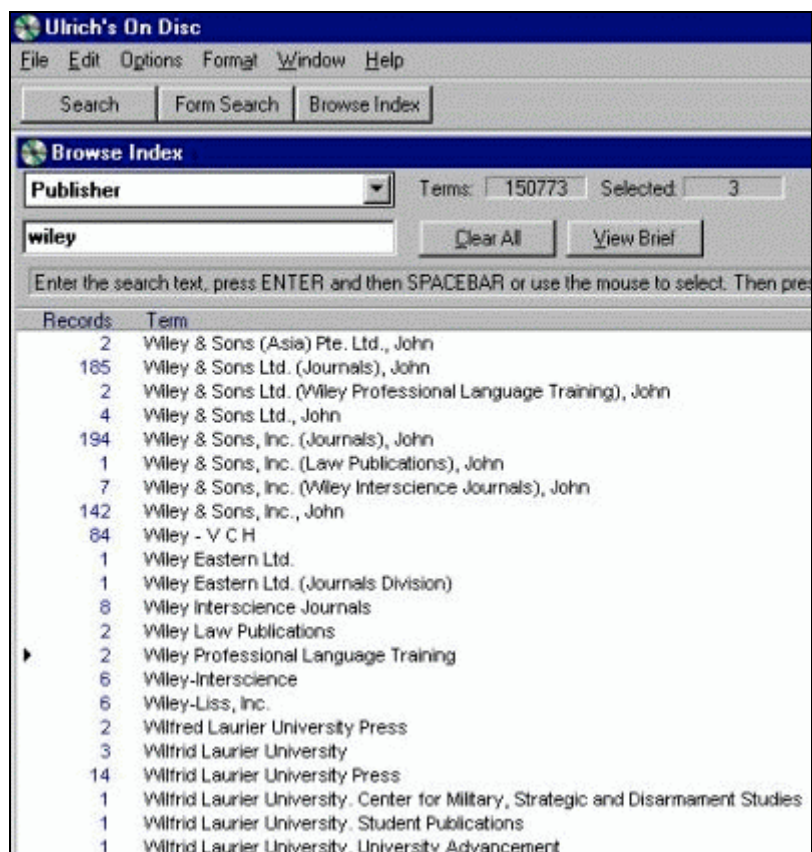
Records	Term
2	Bowker (A Division of Cahnerns Business Information), R.R.
19	Bowker (A Division of Reed Elsevier Inc.), R.R.
53	Bowker (A Reed Reference Publishing Company), R.R.
49	Bowker (A Unit of Cahnerns Business Information), R.R.
9	Bowker (Bowker A & I Publishing), R.R.
5	Bowker - Saur
2	Bowker - Saur (A member of Reed Elsevier plc group)
29	Bowker - Saur (A member of the Reed Elsevier plc group)
1	Bowker - Saur (A Reed Reference Publishing Company)
2	Bowker - Saur Ltd.
20	Bowker - Saur Ltd. (A member of the Reed Elsevier plc group)
4	Bowker - Saur Ltd. (A part of Reed Reference Publishing)
4	Bowker - Saur Ltd. (A Reed Reference Publishing Company)
9	Bowker A & I Publishing
1	Bowker Magazine Group, Cahnerns Magazine Division
1	Bowker Publishing Ltd.
1	Bowker Publishing Ltd. (Electronic Publishing Division)
3	Bowker, R.R.
8	Bowker-Saur
1	Bowker-Saur Ltd.-Butterworths
1	Bowling de Paris

8.2. ábra: A Bowker név variációinak tömege egy Bowker adatbázisban.

Amikor a következetlen tételek nincsenek egymáshoz közel, a használó elvesztheti a rekordok többségét, amelyek tartalmazzák azt a nevet, amelyikre keresett. Amikor egy használó megtalálja a John Wiley & Sons, Inc. alakot az Ulrich's kiadói indexében (8.3. ábra), jó esélye van annak, hogy nem fogja megpróbálni a Wiley formával kezdődő részben is ellenőrizni, azaz olyan rekordok százait vesztí el, amelyek eme konglomerátum brit, ausztrál és amerikai részlegének valamelyik névváltozata alatt jelennek meg (8.4. ábra).



8.3. ábra: Tételek az Ulrich'sban a John Wiley & Sons, Inc. név alatt.



8.4. ábra: További tételek százai az Ulrich'sban a Wiley & Sons, Inc. név alatt.

A folyóiratcím mező elszórt tételeinek keresését nagyban megkönnyítette a DIALOG a Journal Name Finder adatbázissal, amely mind szavanként, mind kifejezésként indexelt tételeket alkot az adatbázisok folyóiratcím mezőiből. A hozzáértő keresők alkalmazhatnak olyan keresést, mint SELECT news AND world AND report, hogy megtalálják a népszerű magazin címének (U. S. News & World Report) variációit a különböző adatbázisokban. Lehangelő, hogy egyedül az ERIC adatbázis három különböző formát használ az ebből a hetilapból felvett 37 cikkhez. Más adatbázisoknak szintén megvannak a maguk variációi, mint például a Transportation Information Services adatbázisnak (File 63 a DIALOG-nál) (8.5. ábra).

A következetlenségek abból is származhatnak, hogy a különböző primer forrásoknak eltérő a gyakorlatuk - például a szerzők második vagy harmadik keresztnévének rövidítésére. Maguk a szerzők is következtelenül használhatták őket a különböző publikációkban. Az H. W. Wilson cég egységesített névalak fájljának szépsége az, hogy egy tétel alatt összehozzák egy szerző műveit még akkor is, ha a primer források nem írják nevüket következetesen vagy pontosan.

A probléma mérésének egy viszonylag szisztematikus módja az, ha böngészünk a szerzők nevének indexében, és olyan előtaggal rendelkező holland vagy német családneveket keresünk, mint például Van Brakel vagy Von Seggern, azután megpróbáljuk megtalálni az olyan változatokat, amelyek az előtagból utótagot csinálnak. Egy gyors ellenőrzés sokat elmond arról, hogy a fájl előállítója milyen következetesen kezeli a személyneveket.

FILE NUMBER	RECORD COUNT	JOURNAL NAME
-----	-----	-----
18	3975	US NEWS & WORLD REPORT
18	275	U.S. NEWS & WORLD REPORT
63	23	US NEWS AND WORLD REPORT
63	1	US NEWS & WORLD REPORT
63	1	U.S. NEWS AND WORLD REPORT,
63	1	U.S. NEWS & WORLD REPORT.
149	2082	U.S. NEWS & WORLD REPORT
485	432	US NEWS & WORLD REPORT
485	379	U.S. NEWS & WORLD REPORT
570	101	US NEWS & WORLD REPORT
570	727	U.S. NEWS & WORLD REPORT
1	16	U.S. NEWS AND WORLD REPORT
1	17	U.S. NEWS & WORLD REPORT
1	2	U. S. NEWS & WORLD REPORT

8.5. ábra: A *U. S. News & World Report* címének különböző formái.

Nagyon lehangoló, hogy a LISA adatbázisban olyan sok különbözően írt változat van a személynevekre. Vegyük például Pieter van Brakelt, aki a V betű alatt 41 tételben szerepel négy különböző módon írt változatban, s a B betű alatt is 30-szor három változatban (8.6. ábra). Az ISA-ban a név 31 tételben található meg két formában s csak egyszer a B betű alatt. A LibLitben - nem meglepő módon - Pieter van Brakel mind az 58 rekordja a B betű alatt található egyetlen formában. Hasonló eredményre jutunk sok más tesztkeresés során is.

A Getty egységesített névalak fájlja (Getty Name Authority File) megmutatja, hogy a különböző források sajátos formái hogyan kezelhetők egy kontrollált szótár segítségével, kereszthivatkozásokkal a különböző névalakoktól a Getty adatbázisok által választott, szabványosított formához (8.7. ábra).

DIALOGWEB		
Command Search		
Set	Term Searched	Items
S1	AU="VAN BRAKEL, P."	4
S2	AU="VAN BRAKEL, P. A."	18
S3	AU="VAN BRAKEL, P. A."	14
S4	AU="VAN BRAKEL, PIETER A."	5
S5	AU="BRAKEL, P. A. V."	20
S6	AU="BRAKEL, P. A. VAN"	9
S7	AU="BRAKEL, P. V."	1

Format
Free
Number of Records
10

Show Database Details for:
61: LISA (Library & Information Science Abstracts)

8.6. ábra: Súlyos következtelenség a LISA szerzői indexében.

GETTY VOCABULARY PROGRAM	
ULAN Name Record	
Bosch, Hieronymus (Hieronymus van Aken)	[BA,GC,PR]
(Netherlandish painter, ca.1450-1516)	[PR]
(Netherlands artist, op.1474-m.1516)	[WC]
(North Netherlandish painter, ca.1450-d.1516)	[GC]
(early Netherlandish painter, ca.1450-1516)	[BA]
Aeken, Hieronymus van	[BA]
Aken, Hieronymus van	[BA]
Ambrosius Bosch	[PR]
Bosch	[PR]
Bosch, Hieronymous	[PR]
Bosch, Hieronymous (Hieronymus van Aken)	[PR]
Bosch, Hieronymus	[GC,PR]
Bosch, Hieronymus (Hieronymus or Jerome van Aken)	[WC]
Bosch, Jerome	[PR]
Hieronymous Bosch	[PR]
Hieronymus Bosch	[PR]
Hieronymus van Aken	[PR]
J. Bos	[PR]
Jer. Bos	[PR]
Jeronimus Bosch	[PR]
Bibliography:	
*Encyc. world art; *Friedländer, Early Neth. ptg.;	
*Kindlers Malerei Lex.; *RILA/BHA; *Thieme-Becker	

8.7. ábra: Részlet a példás Getty Name Authority File-ből.

A folyóiratcímeket még a szerzői neveknél is következtelenebbül alkalmazzák a változatok sokféle lehetősége miatt. Ezek a következők: a kiadó testületek rövidítéseinek és teljesen kiírt változatainak használata, a címben szereplő szavak közül soknak a rövidítése, az alcímek ötletszerű felvétele és kihagyása, az *and* és *and of* következtelen kihagyása, helyettesítésük az & és a kötőjel (-) karakterekkel, nem említve a kiadás helyével való kiegészítést, hogy megkülönböztessék egymástól az egyébként teljesen azonos folyóiratcímeket. A *Database* látszólag teljesen egyszerű címe nagyon sok változatban található meg ugyanazon az adatbázison, a PASCAL-on belül (8.8. ábra). Nincs olyan adatbázis, amely a folyóiratcímek mezőjének következtelenségeiben és pontatlanságaiban vetekedhetne a PASCAL adat-

bázissal, amelyet a Francia Tudományos Kutatási Központ ad ki. A könyv szerzőjének becslése szerint átlagosan 4,3 címvariáció és elírás jut egy tételre (nem számítva a szabályos rövidítéseket). Ez az arány úgy érhető el, hogy a hosszabb címekre variációk tucatjai vannak, ahogyan azt a *Bulletin of the American Society for Information Science* (8.9. ábra) és a *Journal of the American Oil Chemists' Society* (8. 10a. és 8. 10b. ábra) példája illusztrálja. Az első példában a legtöbb tétel elég közel van egymáshoz, de a második cím esetében a tételek nagy mértékben szóródnak. Csak azért soroljuk fel itt őket egymás után folyamatosan, hogy a listát ennek a könyvnek egy oldalára korlátozzuk. (Még szerencse, hogy az *oil* szó nem rövidíthető, s így nem alkottak még további tucatnyi vagy még több variációt.)

Search History		
Database Details		
Set	Term Searched	Items
S1	JN="DATABASE"	38
S2	JN="DATABASE (U.S.)"	1
S3	JN="DATABASE (USA)"	3
S4	JN="DATABASE (WESTON)"	2
S5	JN="DATABASE (WESTON, CONN.)"	3
S6	JN="DATABASE : (WEST)"	1
S7	JN="DATABASE : (WESTON)"	1

8.8. ábra: A *Database* folyóirat címének változatai egyetlen adatbázisban, a PASCAL-ban.

DIALOG(R)File 414 :Dialog Journal Name Finder(TM)		
(c) 1999 The Dialog Corp. All rts. reserv.		
FILE NUMBER	RECORD COUNT	JOURNAL NAME
144	250	BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATI
144	1	BULL. AMER. SOC. INFORM. SOC.
144	130	BULL. AMER. SOC. INFORM. SCI.
144	244	BULL. AM. SOC. INF. SCI.

8.9. ábra: Egy hosszabb folyóiratcím változatai a PASCAL-ban (egymáshoz közeli tételek)


DIALOG(R)File 414 :Dialog Journal Name Finder(TM)		
(c) 1999 The Dialog Corp. All rts. reserv.		
FILE NUMBER	RECORD COUNT	JOURNAL NAME
144	1	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOCIETY.
144	2	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOCIETY.
144	1	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOCIETY.
144	49	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOCIETY.
144	3450	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOCIETY
144	1	JOURNAL OF THE AMERICAN OIL CHEMISTS' SOC.
144	379	JAOCs, J. AM. OIL CHEM. SOC.
144	547	JAOCs - J. AM. OIL CHEM. SOC.
144	2	J. AMER. OIL. CHEMISTS' SOC.
144	17	J. AMER. OIL. CHEMISTS' SOC.
144	1	J. AMER. OIL. CHEMISTS' SOC
144	75	J. AMER. OIL CHEMISTS' SOC.
144	2	J. AMER. OIL CHEMISTS' SOC

8. 10a. ábra: Egy hosszabb folyóiratcím változatai a PASCAL-ban (szétszóródott tételek).

144	2	J. AMER. OIL CHEMISTS' SOC
144	2	J. AMER. OIL CHEMISTS'S SOC.
144	1	J. AMER. OIL CHEMISTS'. SOC.
144	1	J. AMER. OIL CHEMISTS' SOC.,
144	1235	J. AMER. OIL CHEMISTS' SOC.
144	9	J. AMER. OIL CHEMISTS' SOC
144	1	J. AMER. OIL CHEMISTS/ SOC.
144	1	J. AMER. OIL CHEMISTS. SOC.
144	9	J. AMER. OIL CHEMISTS SOC.
144	2	J. AMER. OIL CHEMISTR'S SOC.
144	59	J. AMER. OIL CHEMIST'S SOC.
144	5	J. AMER. OIL CHEMIST' SOC.
144	5	J. AMER OIL CHEMISTS' SOC.
144	1	J. AMER OIL CHEMISTS SOC.
144	4	J. AM. OIL. CHEM. SOC.
144	1	J. AM. OIL CHEMISTS SOC.
144	3807	J. AM. OIL CHEM. SOC.
144	1	J. AM. OIL CHEM SOC.
144	1	J. AM OIL CHEM. SOC.
144	22	AM. OIL CHEM. SOC.

8. 10b. ábra: Egy hosszabb folyóiratcím változatai a PASCAL-ban [folytatás]

Még egy kisebb adatbázisban is, amely a PASCAL-nál sokkal kevesebb folyóiratot tartalmaz, könnyen megtörténhet, hogy a szóródás miatt a használók nem találnak meg különböző módon írt folyóiratcímeket és így fogalmuk sincs az ebből a folyóiratból származó olyan rekordok nagy százalékáról, ahol épp ezt a formát használják. Például az ISA-ban szerepel a *MIS Quarterly* és a *Management Information Systems Quarterly* folyóiratcím is (8.11. ábra). Az általánosabban használt névalakhoz sokkal kevesebb rekord tartozik, mint ahhoz a névhez, amit sokkal kisebb valószínűséggel keressük, s amelyet a legtöbb másik adatbázisban csak ritkán használnak.

 new search data		
Database Details		
Set	Term Searched	Items
S1	JN="MIS QUARTERLY"	12
S2	JN="MANAGEMENT INFORMATION SYSTEMS QUARTERLY"	253
Show Database Details for: 202: Information Science Abstracts		

8.11. ábra: A találatok számának megoszlása egy folyóirat címének két egymáshoz nem közeli formája között.

A következtetések olyan látszólag egyszerű mezőkben is megmutatkozhatnak, mint a dokumentumtípus. A Periodical Abstracts Plus Text nem az egyetlen adatbázis, amely különböző dokumentumtípusként határozta meg ugyanannak a rovatnak („Péter’s Picks and Pans”) az egyes részeit, ahogy azt a 8.12. ábra mutatja.

Dialog Response		
Adding title to results...		
DOCUMENT TYPES FOR PETER'S PICKS AND PANS COLUMN FILE 484		

RANK: S2/l-11 Field: DT= File(s): 484		
(Rank fields found in 11 records -- 4 unique terms)		
RANK No.	Items	Term
-----	-----	-----
3	3	COMMENTARY
2	6	FEATURE
1	11	PERIODICALS
4	2	PRODUCT REVIEW-COMPARATIVE
---end of results---		

8.12. ábra: Következetlenül meghatározott dokumentumtípusok ugyanannak a rovatnak a közleményeiben.

9. Teljesség

A következtetlenségek az adatbázisokban problémát jelenthetnek a keresők számára. A pontatlanságok komoly következményekkel járhatnak. A teljesség hiánya azonban végzetes lehet egy keresésnél. A következtetlenségek és pontatlanságok láthatók. Az adatbázis nem teljes volta, tehát amikor egy vagy több, a keresésnél szokásosan szűrésre használt adatelem a rekordok nagy számából hiányzik, magától értetődően nem látható. Néha a súgó (help file) figyelmezteti a használót a teljesség hiányára; néha a hozzáértő használó képes felderíteni egy adatbázis teljességének hiányát, de gyakran senki nem tud erről. Az elkövetett hibák szabad szemmel is láthatóak, a mulasztásból eredő hibák nyilvánvalóan nem. Azonban drasztikusan torzíthatják az eredményeket, anélkül, hogy a használó tudatában lenne ennek.

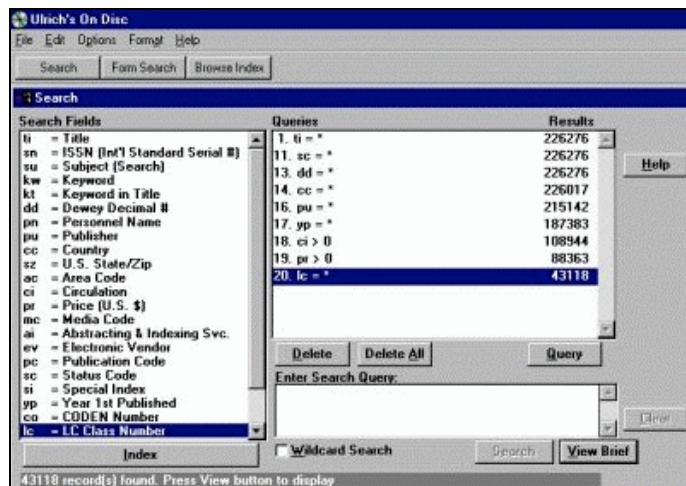
Súlyos következményeik ellenére a mulasztásokból eredő hibákat (kihagyásokat) nem tárgyalták olyan mértékben, mint az elkövetett hibákat. Basch (1990a) állapítja meg, hogy a kihagyások vagy az üresen hagyott mezők gyakran eredményezik azt, hogy a kereső nem jut hozzá releváns rekordokhoz. Quint (1989) állítása szerint félrevezető és drága outputot eredményez, amikor az eredményeket olyan adatelemek szerint rendezik sorba, amelyek nincsenek jelen minden rekordban. Ez még kevésbé lehet nyilvánvaló a használók számára, akik nem jönnek rá, hogy a kiadási év szerint való rendezés egyszerűen kirotálja azokat a rekordokat, amelyekben nem található érték a kiadási év mezőjében. Tenopir (1992, 1995) ismételten figyelmeztetett a hiányzó adatelemek veszélyeire. Jacobs (1993b) számos nagynevű adatbázisban mutatja ki a teljesség hiányának mértékét, továbbá azokat a módszereket, amelyekkel a kihagyás mérete vizsgálható.

A problémát súlyosbította az elektronikus adatbázisoknak a nyomtatott megfelelőikhez képest meglevő, széles körben hangoztatott előnye; vagyis az, hogy sok olyan adatelem szerint is kereshetők, amelyek szerint ezt a nyomtatott változatokban nem lehet megtenni. Ez igaz, de az is igaz, hogy a nagyméretű kihagyás esetén a keresés nem vezet eredményre. Például az időszaki kiadványok kereshetők az Ulrich's Plus adatbázisban a Kongresszusi Könyvtár osztályozási jelzete szerint. Ezt büszkén sorolják fel a hozzáférési pontok között a Bowker termék katalógusában, s a keresési űrlapon is előkelő helyen jelenítették meg. Amit a naiv használó nem tudhat, az az, hogy a 226276 rekord közül csak 43118-ban (19%) van meg ez a kód (9.1. ábra). Ezt nem lenne szabad hozzáférési pontként felajánlani, mivel ennek használata automatikusan az adatbázis ötödére korlátozza a keresést.

Amikor ezt a nagyon jelentős mértékű kihagyást a könyv szerzője egy konferencián tartott előadásán közzétette, a Bowker akkori képviselője azzal érvelt, hogy kevés használó keresne a kongresszusi könyvtár osztályozási jelzete szerint vagy használná azt szűrőként. Ez az érvelés azonban sántít. Minden egyetemi könyvtár és sok szakkönyvtár a Kongresszusi Könyvtár osztályozási jelzetét használja a Dewey-jelzet helyett (amely egyébként minden rekordban elérhető). Sokkal jobb lenne előzékenyen és egyértelműen jelezni a kereső képernyőn vagy legalább a súgóban, hogy a rekordoknak csupán 19%-ában található meg ez a jelzet. Az adatbázis Bowker által kiadott CD-ROM-os változatában legalább felfedezhető a kihagyás mértéke egy kereséssel, ahogy azt a 9.1. ábra mutatja.

Azokban a változatokban azonban, amelyek a két osztályozási jelzetet ugyanabba a mezőbe rakják vagy nem engedik meg a prefixes indexekben a teljes csonkolással való keresést, ez a hiányosság rejtve marad. A Kongresszusi Könyvtár osztályozási jelzetével végzett keresések eredményei durván félrevezetik a használót. A DIALOG-os változatban a Kongresszusi Könyvtár és a Dewey-féle osztályozás jelzeteit egy indexben olvastották össze, így még a hozzáértő használó sem képes arra, hogy kitalálja, milyen elfogadhatatlanul hiányosak a rekordok.

Az ár és a példányszám adatai hiányának mértéke az Ulrich'sban is figyelmeztető jel, hogy ne keressünk ezen adatelemek szerint még akkor sem, ha erre csábítanak minket a Bowker reklámanyagai. Az ár szerint nem lehet keresni a DIALOG-os változatban. Bár a kiadó mezője „csak” kevesebb mint 10000 rekordból hiányzik, ez furcsa, mert a feltételezések szerint az adatokat a folyóiratok kiadói szolgáltatják, így ennek az adatelemnek igazán elérhetőnek kellene lennie.



9.1. ábra: A Kongresszusi Könyvtár osztályozási jelzeteinek nagymértékű kihagyása az Ulrich's Plus adatbázisban.

A legtöbb adatbázis kihagy természetesnek tekintett adatelemeket, bár csak kevesen olyan mértékben, mint amit itt bemutattunk. Azonban még így is riasztó látni, hogy az adatbázis kiadója nem figyelmeztet arra, hogy az ERIC adatbázis rekordjainak 28%-ában nincs dokumentumtípus mező; ez a Compendexben 32%, az NTIS-ben 72% (9.2. ábra), a Packaging Science and Technologyban 78% és a GeoArchive-ban 90% (9.3. ábra). Még az adatelemek látszólag kismértékű kihagyása is nagyon jelentős lehet a legnagyobb adatbázisok esetében. Az AGRICOLA-ban a rekordok 5%-ában nincs nyelvi mező, ami nem hangzik rosszul, amíg rá nem jövünk, hogy egy 4 millió rekordos adatbázisról van szó, így a kihagyások csaknem 200000 rekordot zárnak ki, amikor valaki nyelv szerinti szűrést csinál.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	2081941
S2	LA=?	1403571
S3	DT=?	591031
Show Database Details for:		
6: NTIS - National Technical Information Service		
Bluesheet		

9.2. ábra: A dokumentumtípus mező kihagyása az NTIS adatbázisból minden magyarázat nélkül.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	644439
S2	DT=?	63372
Show Database Details for:		
58: GeoArchive		
Bluesheet Rates		

9.3. ábra: A dokumentumtípus mező kihagyása a GeoArchive adatbázisból minden magyarázat nélkül.

A három könyvtár- és információtudományi adatbázis (9.4. ábra) elég jó teljesítményt nyújt olyan gyakran használt keresési kritériumok körében, mint a kiadási év, nyelv és dokumentumtípus. (A LISA nem ígér dokumentumtípust, így a nulla találat elfogadható, bár ez elég kiábrándító tervezésre utal.) Ahogyan a 9.5. ábra adatai mutatják, a PsycINFO szintén jól teljesített, de a Mental Health Abstractsben 513017 rekord közül csak 82133-ban van dokumentumtípus (16%). A nyelvi mező ugyancsak hiányzik ezen adatbázis rekordjainak csaknem 90%-ából, de ez érthető, ha tudjuk, hogy házi gyakorlatuk szerint az angolt mint nyelvet nem kapcsolják a rekordokhoz. Más kérdés az, hogy az MHA DIALOG-os adatbázis-leírása az LA=ENGLISH formát használja keresési példaként (9.6. ábra), ami - nem meglepő módon - nulla rekordot eredményez (9.7. ábra). Bár együtt érezhetünk a fájl előállítójával, aki olyan szívósan küzdött ennek a szónak a helyes írásmódjával (9.8. ábra), a használókat figyelmeztetni kellene arra a gyakorlatra, hogy a sok próbálkozás után abbahagyták az angolnak a rekordokhoz nyelvként való hozzákapcsolását.

Set	Term Searched	Items	File
S1	UD=?	609082	
S1	UD=?	195544	438
S1	UD=?	207262	202
S1	UD=?	206276	61
S2	PY=?	600522	
S2	PY=?	195544	438
S2	PY=?	205814	202
S2	PY=?	199164	61
S3	LA=?	597128	
S3	LA=?	189176	438
S3	LA=?	207261	202
S3	LA=?	200691	61
S4	DT=?	394923	
S4	DT=?	187661	438
S4	DT=?	207262	202
S4	DT=?	0	61

9.4. ábra: A mezők jelenlétének ellenőrzése a könyvtár- és információtudományi adatbázisokban.

Search History			
Database Details			
Set	Term Searched	Items	File
S1	UD=?	2093197	
S1	UD=?	513017	86
S1	UD=?	1580180	11
S2	PY=?	2092586	
S2	PY=?	512457	86
S2	PY=?	1580129	11
S3	LA=?	1636034	
S3	LA=?	55854	86
S3	LA=?	1580180	11
S4	DT=?	1662313	
S4	DT=?	82133	86
S4	DT=?	1580180	11

9.5. ábra: A mezők jelenlétének ellenőrzése a PsycINFO és az MHA adatbázisokban.

ADDITIONAL INDEXES [return to top]				
SEARCH PREFIX	DISPLAY CODE	FIELD NAME	INDEXING	SELECT EXAMPLES
None	AN	DIALOG Accession Number		
AN=	AN	NDMH Accession Number ²	Phrase	S AN=APA1968-06726
AU=	AU	Author	Phrase	S AU=SCOTT, R. S.
BN=	BN	International Standard Book Number (ISBN) ²	Phrase	S BN=0-02-904330-1
CC=	None	Contractor Code ²	Phrase	S CC=HIS
None	CP	Country of Publication		
CS=	CS	Corporate Source	Word	S CS=(HOFFMAN(W)LA(W)ROCHE)
DT=	DT	Document Type ^{2,3}	Phrase	S DT=JOURNAL
JN=	JN	Journal Name ^{2,3}	Phrase	S JN=HUMAN RELATIONS?
LA=	LA	Language	Phrase	S LA=ENGLISH
PY=	None	Publication Year	Phrase	S PY=1997
None	RF	Number of References		
SN=	SN	International Standard Serial Number (ISSN) ²	Phrase	S SN=0018-7267
SO=	SO	Source Information ²	Word	S SO=(HUMAN(W)RELATIONS)
UD=	None	Update	Phrase	S UD=9999

² Only NDMH records (1969-1982).

³ Also searchable as /ID.

² Only ISI records (1983+).

9.6. ábra: Részlet az MHA adatbázis-leírásából.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	513017
S2	S1/500000.999999	82132
S3	S2 AND DT=?	82132
S4	LA=ENGLISH	0

Show Database Details for:











86: Mental Health Abstracts

9.7. ábra: Nincs rekord az English szóra az MHA nyelvi mezőjében.

A PsycINFO a legjobb minőségű adatbázisok közé tartozik, és nem meglepetés, hogy Ovidos változatában végzett tesztkeresés (9.9. ábra) azt mutatja, hogy mind az 1594013 rekordban volt dokumentumtípus mező. A populáció mezőjét szintén úgy alkalmazták, hogy vagy a *human* (emberre vonatkozó), vagy az *animal* (állatokra vonatkozó) fogalom szerepeljen benne, de egy friss teszt azt tárta fel, hogy a rekordok 20%-ában nincs meg ez az adatelem.

databases	alerts	cost	logoff	help
E5	86		1	LA=E NGLISH
E6	86		1	LA=EEN
E7	86		2	LA=EGLISH
E8	86		3	LA=EGNLISH
E9	86		2	LA=EIGLISH
E10	86		4	LA=EINGLISH
E11	86		1	LA=EMGLISH
E12	86		1	LA=ENBLISH
E13	86		5	LA=ENG
E14	86		1	LA=ENGIISH
E15	86		24	LA=ENGLISH
E16	86		1	LA=ENGLAIH
E17	86		25	LA=ENGLAND
E18	86		1	LA=ENGLIAH
E19	86		2	LA=ENGLIDH
E20	86		2	LA=ENGLIGH
E21	86		1	LA=ENGLIISH
E22	86		1	LA=ENGLILSH
E23	86		3	LA=ENGLIS
E24	86		5	LA=ENGLISH ENGLISH
E25	86		1	LA=ENGLISH 12
E26	86		8	LA=ENGLISH.
E27	86		2	LA=ENGLISH/JAP.
E28	86		1	LA=ENGLISH,
E29	86		1	LA=ENGLISH'
E30	86		3	LA=ENGLISN
E31	86		1	LA=ENGLLISH
E32	86		1	LA=ENGLSH
E33	86		20	LA=ENGLSIH
E34	86		1	LA=ENGLUSH
E35	86		10	LA=ENLGISH

9.8. ábra: Részlet az MHA adatbázis nyelvi indexéből.

PsycINFO 1887 to July 1999			
         			
#	Search History	Results	Display
1	19\$ up.	1594013	Display
2	("authored book" or "book" or "chapter" or "dissertation abstract" or "edited book" or "journal article" or "report" or "secondary publication") pt.	1594013	Display
3	("animal" or "human").po.	1286235	Display

9.9. ábra: A rekordok teljes számának és két mező jelenlétének meghatározása az Ovidban.

A kihagyást a következtetés speciális fajtájaként is meg lehetne határozni. Ám az alapvető adatelemek, mint például a nyelv, ország, kiadási év, illetve a dokumentum típusa megadásában megnyilvánuló következtetés olyan következményekkel jár, amely külön megközelítést igényel. Meg kell jegyezni, hogy nem minden adatelem esetében várható el, hogy minden rekordban jelen legyen. Nyilvánvaló, hogy a szerzői nevet nem lehet megadni anonim

közlemények esetén. Nem minden időszaki kiadványnak van ISSN száma, s nem mindegyik határozza meg a szerkesztőt adatelemként - ahogy például az Ulrich'sban sem.

Minden kiadványhoz tartoznia kell azonban nyelvnek és a dokumentumtípusnak. Gyakori eset, hogy az ilyen adatelemeket nem használták az adatbázis korai éveiben, később vezették csak be őket. Erre a tényre a használót figyelmeztetni kellene. Az ABC-Clio így tesz az America History & Life és a Historical Abstracts adatbázisok esetében, és felhívja a figyelmet arra, hogy a nyelvi mezőt csak 1980-tól adták meg az adatbázisokhoz (míg az adatbázisok 1963-ig, illetve 1973-ig mennek vissza).

Így tesz az Oceanic Abstracts is (9.10. ábra), megjegyezve, hogy a dokumentumtípust csak 1971 és 1973 között használták, ami megmagyarázza, hogy miért csak a rekordok 8%-ában található meg a dokumentumtípus mező. A Searchable Physics Information Notices (SPIN) adatbázis a használó tudomására hozza, hogy a dokumentumtípust csak akkor kapcsolták a rekordhoz, ha az nem folyóiratcikk volt. Néha az okot ki tudják következtetni a tapasztalt használok. Az Occupational Health and Safety adatbázisban is csak a rekordok 8%-ában van nyelvi mező. A nyelvi indexre való pillantás azonban indirekt módon magyarázattal szolgál: nincs angol nyelvű tétel, így nyilvánvalóan a feltételezett érték.

A *Books in Print*ben az angol nyelvű dokumentumok képtelenül kis eredményhalmaza sejtetheti meg a gyakorlott keresővel, hogy csak a két- és többnyelvű dokumentumok (például szótárak) esetében határozzák meg az angol nyelvet.

Egyes esetekben figyelmeztetnek egy adatelem nagymértékű kihagyására, de másakra nem. Az ERIC-ben nincs figyelmeztetés arra, hogy a dokumentumtípus mezőt a rekordok 28%-ában kihagyták, de van egy megjegyzés arról, hogy a nyelvet csak 1979-től adták meg. Ugyanez érvényes az NTIS-re is (9.11. ábra). A BIOSIS figyelmeztet arra, hogy a nyelvet csak 1978-tól adják hozzá a rekordokhoz, de nincs figyelmeztetés vagy magyarázat arra, hogy miért nincs a rekordok 65%-ában dokumentumtípus mező. Egy 12 millió rekordos adatbázis estében a kihagyásnak ez az aránya gigantikus.

ADDITIONAL INDEXES [return to top]				
SEARCH PREFIX	DISPLAY CODE	FIELD NAME	INDEXING	SELECT EXAMPLES
None	AN	DIALOG Accession Number		
AN=	AN	OA Accession Number ⁴	Phrase	S AN=71-1A-00061
AU=	AU	Author	Phrase	S AU=KING, K.
BN=	BN	International Standard Book Number (ISBN) ⁵	Phrase	S BN=0-930118-03-0
CL=	CL	Conference Location	Word	S CL=FLORIDA
CO=	CO	CODEN ¹	Phrase	S CO=MPNBAZ
CS=	CS	Corporate Source	Word	S CS=(USFWS)(SPATUXENT)
CT=	CT	Conference Title ¹	Word	S CT=(NORTH(W)SEA)
CY=	CY	Conference Year ¹	Phrase	S CY=1985
DT=	DT	Document Type ¹	Phrase	S DT=JOURNAL PAPER
JA=	JA	Journal Announcement ⁶	Phrase	S JA=V25N1
JN=	JN	Journal Name	Phrase	S JN=COLONIAL WATERBIRDS?
LA=	LA	Language	Phrase	S LA=JAPANESE
PU=	PU	Publisher ¹	Word	S PU=(PERGAMONWYPRESS)
PY=	PY	Publication Year	Phrase	S PY=1987
SN=	SN	International Standard Serial Number (ISSN) ⁵	Phrase	S SN=0093-3651
None	SO	Source Information ⁶		
TC=	TC	Treatment Code ¹	Phrase	S TC=1
UD=	None	Update	Phrase	S UD=9999

⁴ Present in records entered in the database from 1971 through 1973.

⁵ Present in records from January 1980 to the present.

⁶ Includes Journal Name, Corporate Source, Publication Year, and Pagination.

9.10. ábra: Az Oceanic Abstracts adatbázis-leírásának információi.

CP=	CP	Country of Publication ³	Word & Phrase	S CP=(UNITED(W)STATES) S CP=UNITED STATES
CS=	CS	Corporate Source	Word & Phrase	S CS=(DEPARTMENT(2(W)NAVY) S CS=DEPARTMENT OF THE NAVY, WASHINGTON? S CS=110050
DT=	DT	Document Type	Phrase	S DT=PATENT
JA=	JA	Journal Announcement	Phrase	S JA=QRA9801
LA=	LA	Language ²	Phrase	S LA=ENGLISH
NT=	NT	Note	Word	S NT=(FOREIGN(W)LICENSING)
None	PC	NTIS Price Code		
PY=	PY	Publication Year	Phrase	S PY=1997
RN=	RN	CAS(R) Registry Number ⁴	Phrase	S RN=8001-35-2
RN=	RN	Report Number	Word & Phrase	S RN=(AD(W)D018(W)520) S RN=PATAPPL8854511 S RN=PAT-APPL-8-854 511
SC=	SC	Section Heading Code ⁵	Phrase	S SC=QOF
SH=	SH	Section Heading	Phrase	S SH=ELECTROTECHNOLOGY--GENERAL S SH=40GE
None	SO	Source Information ⁶		
SP=	SP	Sponsoring Organization ⁷	Word & Phrase	S SP=(DEPARTMENT(1(W)ENERGY) S SP=DEPARTMENT OF ENERGY, WASHINGTON?
UD=	None	Update	Phrase	S UD=9712B1

³ Searchable in the Basic Index and in the Additional Indexes.

⁴ Also searchable as RN=

⁵ For best results, search without punctuation.

⁶ For records from 1980 forward.

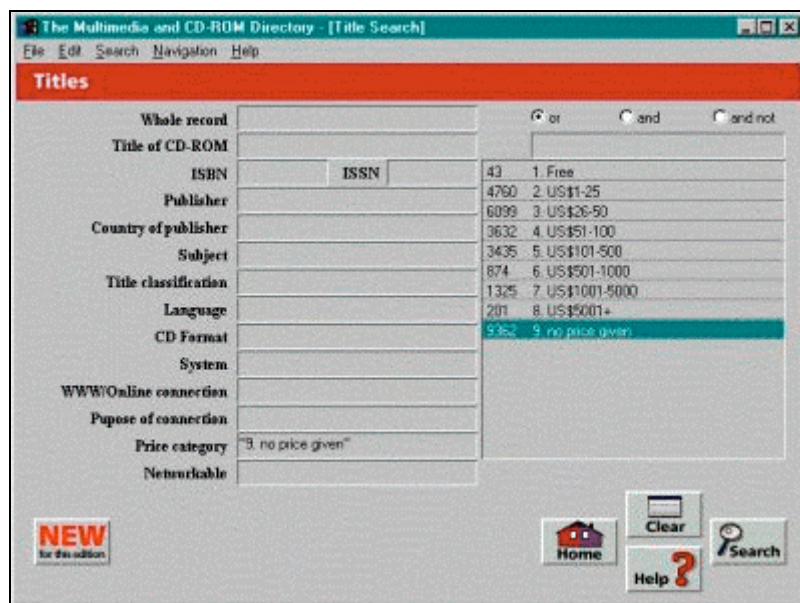
⁷ For records from 1970 forward.

9.11. ábra: Hiányzik a dokumentumtípus mező nagy mértékű kihagyására vonatkozó megjegyzés az NTIS-ben.

Az, ha nem figyelmeztetik a használókat az ilyen kihagyásra, olyan, mint amikor egy étterem vezetője nem ad vészjelzést, amikor a konyha lángra kap, mert arra van csak gondja, hogy minden vendég kifizesse a számlát, mielőtt elmenekül a vendéglőből. A keresők megégethetik ujjukat, ha nem figyelnek arra a súlyos fogyatékosagra, hogy a rekordok hiányosak. Az a hatékony eszköz, amellyel a keresés eredményét nyelv, dokumentumtípus és osztályozási jelzetek szerint szűrhetjük, könnyen veszélyes eszközzé válhat.

Nagyon kevés olyan adatbázis van, amely legalább világossá teszi az adatelemek kihagyását a keresési folyamat során. A Multimedia and CD-ROM Directory például világosan megmutatja, hogy hány rekordban nincs az árra vonatkozó információ (9.12. ábra).

A Directory of Library and Information Professionals, amely ugyanazt a szoftvert használja, nem használja ki ugyanezt a technikát arra, hogy figyelmeztesse a használót arra, hogy a keresésre felkínált adatelemeknek csak egy nagyon kis része található meg valóban a rekordokban. Az egyetlen kivétel az információk szakemberek neve, mivel a nyomtatott dokumentáció figyelmezteti rá a használót, hogy a rekordoknak csak kevesebb mint a felében van a nemre vonatkozó információ. (Az más kérdés, hogy hány használó fér hozzá a nyomtatott dokumentációhoz és/vagy akarná elolvasni azt.)



9.12. ábra: A hiányzó adatelemek egyértelmű meghatározása.

A legkönnyebb mód arra, hogy megismerjük az adatelemek kihagyásának mértékét az, amikor az adatbázis magától megadja ezt az információt, ahogy azt a 9.12. ábra mutatja. A hiányosságok megvizsgálásának másik módja az, hogy megnézzük az olyan indextételeket egy referáló és indexelő adatbázisban, mint a N/A, Not Available (nem elérhető) vagy Undetermined (Meghatározatlan). Nem csak rendkívül magasnak, de furcsának is tűnik, hogy a Library Literature-ben 25880 olyan rekord van, amelyben a nyelv meghatározatlan, amíg nem végzünk némi kutatást s rá nem jövünk, hogy az H. W. Wilson cég gyakorlata szerint a könyvkritikákat tartalmazó rekordoknál nem adják meg a nyelvet (9.13. ábra). Ez nem logikus, és a kritikából a kritizált mű nyelve biztosan meghatározható lenne.

Search History		
Database Details		
Set	Term Searched	Items
S1	LA="UNDETERMINED"	25880
S2	S1 AND DT=REVIEWS	25880
<div> Show Database Details for: 438: Library Literature Bluesheet Rates </div>		

9.13. ábra: Rekordok meghatározatlan nyelvvel a LibLit-ben.

Ezeket a meggyónt tételeket még nem mindig lehet egy az egyben elfogadni. Még akkor is, ha az *Undetermined* vagy az *Unavailable* szavakat tartalmazó rekordokat is beszámítjuk egy adott indexben, az összeg nem adja ki az adatbázisban meglevő rekordok teljes számát. Ez teljesen nyilvánvaló a Marquis Who's Who adatbázisban, ahol a nemek mezőjében négyféle érték van (9.14. ábra), de a hozzájuk tartozó rekordok mennyisége még mindig nem éri el a rekordok teljes számát. Ha csak azokat a rekordokat nézzük át, amelyekben a nem meghatározatlan, kíváncsiak lehetünk arra, miért olyan nehéz meghatározni olyan emberek nemét, akiknek a keresztnéve Tatiana vagy Charles, és még az életrajzban is lehetnek további nyomravezető jelek.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	965187
S2	SX="FEMALE"	281456
S3	SX="MALE"	681677
S4	SX="N"	2
S5	SX="UNAVAILABLE"	2026
S6	S1 NOT S2:S5	26
Show Database Details for:		
234: Marquis Who's Who®		
Bluesheet Rates		

9.14. ábra: Meghatározatlan nemre vonatkozó információk a Marquis Who's Who adatbázisban

A címjegyzék adatbázisokban a 9999 az általánosan használt kód annak jelzésére, hogy egy numerikus adatelem nem elérhető. Például, ha a Standard Industry Classification (SIC) kód nem határozható meg egy cég esetében, ezt a számot használják. A kód szöveges megfelelője nem állandó, változik. Egyes adatbázisokban ez nem osztályozott (nonclassified vagy unclassified) dolog, másokban osztályozhatatlan. Eltűnődhetünk, hogy ki által nem osztályozható. Ez a konvenció a fájlok készítőinek elég nagy mozgásteret ad, s úgy tűnik, ezzel vissza is élnek.

Az American Business Information (ABI) címjegyzékben több mint 100000 rekord van, ahol a SIC kód helyén 9999 szerepel. Felfoghatatlan, hogy olyan cégek, mint az Office Machines & Furniture és az Office Depot miért osztályozhatatlanok (9.16. ábra). Azt lehetne gondolni, hogy a SIC 502112 (OFFICE FURNITURE & EQUIP-DEALERS [WHOL], irodai bútor és berendezés-szállítók) vagy valamelyik variációja megfelelő lenne, ahogy 15000 cég valóban meg is kapta ebben az adatbázisban.

A Moody's Corporate News adatbázis a 9999-et adta a Walt Disney Company-nak (9.17. ábra), pedig gyerekjátéknak látszik ennek meghatározása. A Dun & Bradstreet valószínűleg nem kápráztatja el az ügyfeleket azzal, hogy a 9999-et vágja hozzá olyan cégekhez, mint az America Online, American Eagle Airlines és a Burger King Corporation (9.18. ábra). Ennyi pénzért, amit fizetünk azt lehetne várni, hogy a cég megengedheti magának, hogy olyan embereket alkalmazzon, akik képesek kitalálni, milyen SIC kódok illenének legjobban ezekre a cégekre.

Search History		
Database Details		
Set	Term Searched	Items
S1	SC=9999 (NONCLASSIFIABLE ESTABLISHMENTS)	115525
Show Database Details for:		
531: American Business Directory		
Bluesheet Rates		

9.15. ábra: Rengeteg olyan rekord van az ABI címjegyzékben, amely a 9999-es SIC kódot kapta.

50.	<input type="checkbox"/>	3/CO,PC,SC/50 OFFICE MACHINES & FURNITURE
Primary SIC and Yellow Page Product Line(s): 9999 (UNCLASSIFIED ESTABLISHMENTS) 999977 (NONCLASSIFIED ESTABLISHMENTS) <small>Amer. Bus. Directory (Dialog® File 331); (c) 1999 American Business Information.</small>		
81.	<input type="checkbox"/>	3/CO,PC,SC/81 OFFICE DEPOT
Primary SIC and Yellow Page Product Line(s): 9999 (UNCLASSIFIED ESTABLISHMENTS) 999977 (NONCLASSIFIED ESTABLISHMENTS) <small>Amer. Bus. Directory (Dialog® File 331); (c) 1999 American Business Information.</small>		

9.16. ábra: Az Office Depot mint osztályozhatatlan szervezet az ABI címjegyzékben.

3/CO,PC/15 (Item 1 from file: 556)	
DISNEY (WALT) CO. (THE)	
COMPANY INFORMATION:	
Primary SIC: 9999	Moody's(R)CorpNews

9.17. ábra: A Walt Disney Company a 9999-es kódot kapta a Moody'stól.

1.	<input type="checkbox"/>	3/CO,PC/1 AMERICA ONLINE INC
Primary SIC: 9999 Nonclassified establishment		D&B-Who Owns Whom
3.	<input type="checkbox"/>	3/CO,PC/3 AMERICAN EAGLE AIRLINES INC
AMR EAGLE-RAMCI		
Primary SIC: 9999 Nonclassified establishment		D&B-Who Owns Whom
4.	<input type="checkbox"/>	3/CO,PC/4 BURGER KING CORPORATION
BURGER KING MARKETING		
Primary SIC: 9999 Nonclassified establishment		D&B-Who Owns Whom

9.18. ábra: Az America Online, American Eagle Airlines és Burger King a 9999-es kódot kapta a Dun & Bradstreettől.

A Bowker cég akkor használja a 9999-es értéket, amikor a megjelenés éve nem ismert. Ez elfogadható megoldás lenne, de sok adatbázisban vannak problémái. Az ilyen kódokat vagy figyelmeztetéseket nagyon gyakran nem következetesen használják, s ezzel félreinformálják a használót. Ha ismerjük egy adatbázisban a rekordok teljes számát, megpróbálhatunk egy teljesen csonkolt keresést, mint például PY=?, amely azt jelenti, hogy „Keress meg minden rekordot, amelynek bármilyen értéke van a megjelenés éve mezőben.” Ha az ennek eredményeként kapott halmaz kisebb, mint a rekordok teljes száma, akkor tudjuk, hogy hány rekordnak nincs semmilyen értéke az adott mezőben - még figyelmeztető kódja vagy szövege sem. Ebben az esetben 66 rekordban található meg a speciális 9999-es kód a PY mezőben, de még ezekkel együtt is közel 80000 másiknak nincs semmilyen értéke a PY mezőben (9.19. ábra). Ez ellentétes azzal a céllal, ami miatt a 9999-es értéket használják a *Books in Print*-ben.

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	2359042
S2	PY=?	2271291
S3	PY=9999	66

9.19. ábra: A 9999-es kód félrevezető használata a *Books in Print*-ben

Honnan tudhatjuk meg egy adatbázis teljes méretét? Elfogadhatjuk azt, amit a tartalom-szolgáltató vagy az adatbázis kiadója mond (bár némi fenntartásokkal). Ennél jobb, ha keresést végzünk az UD (Update, aktualizálás) indexben - vagy teljesen csonkolt keresést használva, vagy aritmetikai művelettel, mert ezt a mezőt automatikusan generálják, amikor az adatbázist aktualizálják. A 9.20. ábra az ERIC adatbázis DIALOG-os változatában mutat be egy ilyen keresést, amely kihagyások jelentős mennyiségét tárja fel. Jegyezzük meg, hogy az UD mező néha az adatbázis újra betöltésének dátumára vonatkozik. Ez azért van, mert sok címjegyzéket nem aktualizálnak, hanem újra betöltenek (azaz a rekordokat törlik és módosítják aszerint, ahogy a címjegyzék tárgya változik). Ez történik például akkor, amikor egy cégnek a székháza elköltözik vagy addigi vezetőjét egy újabb váltja fel. Ha az UD mező nem elérhető, van néhány más adatelem, amely nagy valószínűséggel minden rekordban megtalálható és kereshető (és csak néhány lehetséges értéke van). Ezek további keresési eredményekkel is igazolhatók lehetnek.

A 9.21. ábra megmutat egy teljességet igazoló keresést a Bowker's Complete Video Directory-ban. Bár nincs benne UD mező, nagy biztonsággal kijelenthetjük, hogy 192166 rekord van az adatbázisban. A kulcsszavak indexében végzett teljesen csonkolt keresés (kw=\$) a legjobb igazolása az adatbázis méretének, mert megszámol minden rekordot, amelyben bármilyen információ van bármelyik mezőből, függetlenül a mező tartalmától. Nagyon tanulságos megtudni, hogy a rekordok csaknem egyharmadában nincs adat a PC (publication code, kiadási kód) mezőben, 43%-ban nincs benne a gyártás éve, s a rekordok több mint 75%-ában nincs kibocsátási év, s körülbelül 8000 rekordban nincs információ, arról, hogy mi a videó témája.

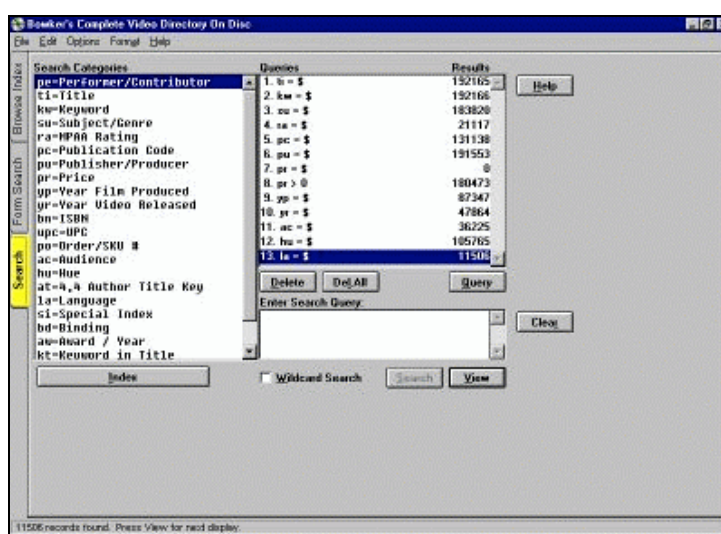
Hasonló keresés elvégezhető az Ulrich's Plus adatbázisnak a Bowker által kiadott CD-ROM változatában. A teljesen csonkolt és a több indexben végzett numerikus keresések világosan megmutatták, hogy 226276 rekord volt az adatbázisban. A számok azt is megmutatták, hogy mennyire hiányos sok rekord (9.22. ábra).

Search History		
Database Details		
Set	Term Searched	Items
S1	UD=?	990090
S2	PY=?	966799
S3	LA=?	656004
S4	PY>1978	617028
S5	DT=?	712423

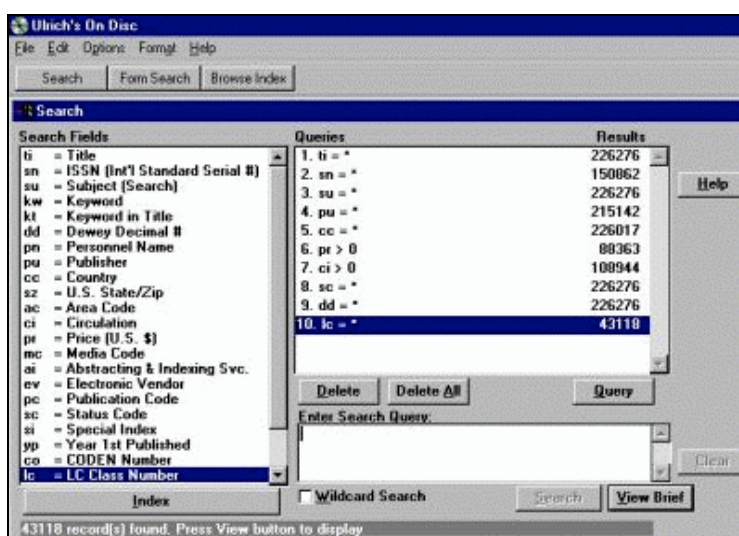
Show Database Details for:

1: ERIC Bluesheet Rates

9.20. ábra: A rekordok teljes számának keresése az ERIC DIALOG-os változatában.



9.21. ábra: A teljességre vonatkozó keresés a Bowker's Complete Video Directoryban.



9.22. ábra: A teljességre vonatkozó keresés az Ulrich's Bowker által kiadott CD-ROM változatában.

A teljességet megállapító keresések mindig adatbázis specifikusak, s feltételezik az adatbázisban alkalmazott hagyományok ismeretét. Egyes adatbázisokban a nyelvi mezőt csak a nem angol nyelvű dokumentumok esetében adják meg, feltételezve, hogy az angol az alapértelmezett érték. Ugyanez érvényes egyes adatbázisokban, ha a megjelenés országa az Amerikai Egyesült Államok. Hasonló módon az ERIC-ben a rekordok teljes száma könnyen meghatározható, bármilyen szoftvert használva, ha két keresést végzünk, és összehasonlítjuk ezek eredményeit másokéval.

Az ERIC-ben minden rekord vagy a RIE (Resources in Education), vagy a CIJE (Current Index to Journals in Education) részhalmazhoz tartozik. Néha a mezőspecifikus indexek átnézése azonnal jelzi, ha komoly problémák vannak egyes adatelemek nem teljes voltával. Az adatbázisok Ovidos változatának sok mezőspecifikus indexe jól illusztrálja ezt. Például a státusz kódok értékeinek átnézése az Ulrich'sban azonnal tájékoztat minket arról, hogy a rekordok teljes száma valahol 236000 közelében lehet ebben a fájlban (9.23. ábra).

Néhány mező jelenléte könnyen meghatározható úgy, ha rákattintunk a jelölőnégyzetekre vagy alkalmazzuk az előre meghatározott korlátokat, egy olyan keresés után, amely meghatározza a rekordok teljes számát. Az Ovidnál a rekordok teljes száma meghatározható a 19\$.ud parancs alkalmazásával, amely visszakeres minden rekordot, amely 1900 és 1999 között került be az adatbázisba. Amikor a PAIS Ovidos változatában az összes kiadványtípusra és az összes nyelvre vonatkozó keresés is ugyanannyi rekordot eredményez, az azt bizonyítja, hogy minden rekordban van mind kiadványtípus, mind nyelvi mező (9.24. ábra).

Ha a WebSPIRS-ben minden olyan rekordra keresünk, amelynek az aktualizálási dátuma nagyobb nullánál, megkapjuk a rekordok teljes számát. Ezt követően ez a halmaz korlátozható az angol majd a nem angol nyelvű dokumentumokra, hogy lássuk, van-e valami különbség. A DIALOG-ban ezt a műveletet a SELECT parancs *limit* szuffixumával végezzük, hogy megtudjuk például azt, hogy csak minden második rekordban van a példányszámra vonatkozó információ (9.25. ábra).

The screenshot shows a web interface with a header containing navigation links (O, V, I, D) and a 'Status Code' title. Below the header is a 'Perform Search' button and a 'Back in Index' button. The main content area is titled 'Choose from among the following' and contains a table with the following data:

Select Status Code	# of Citations
<input type="checkbox"/> active.st.	148198
<input type="checkbox"/> ceased.st.	54781
<input type="checkbox"/> unverified.st.	33139

At the bottom of the table, there is another 'Perform Search' button and a 'Back in Index' button.

9.23. ábra: Egy kevés értéket tartalmazó index átnézése az Ovidban.

PAIS International 1972 to November 1999			
Author Title Journal Search Fields Combine Limit Basic Change Database Logoff			
#	Search History	Results	Display
1	("book or monograph" or "journal article").pt.	467620	Display
2	("english" or "french" or "german" or "italian" or "portuguese" or "spanish").lg.	467620	Display
Run Saved Search Save Search History Delete All Searches			

9.24. ábra: A kiadványtípus és a nyelvi mező teljességének összehasonlítása a PAIS Ovidos változatában.

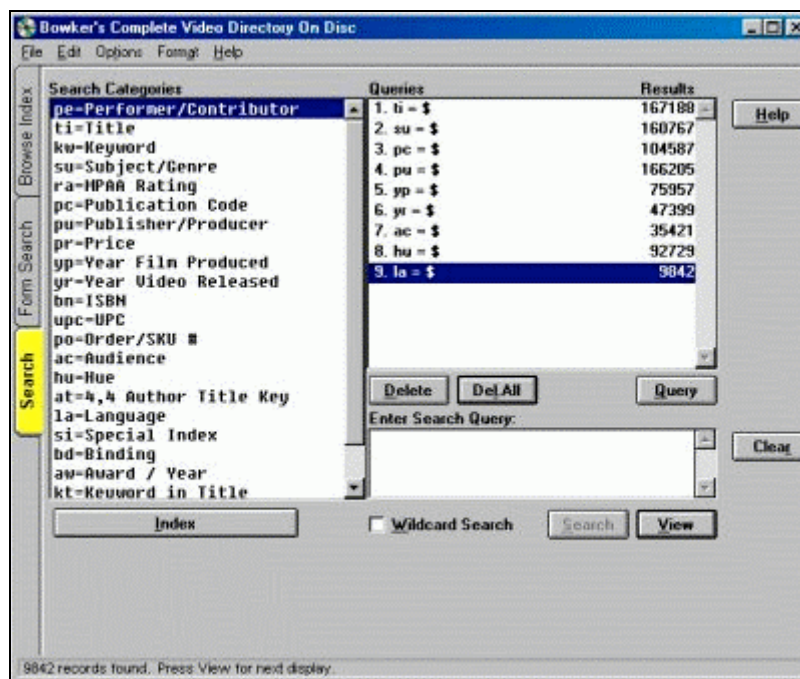
Set	Term Searched	Items
S1	UD=?	236155
S2	S1/NOCIRC	116992
Show Database Details for: 480: Ulrich's(TM) International Periodicals Directory		

9.25. ábra: A példányszám mező teljességének meghatározása az Ulrich's DIALOG-os változatában.

A hatékony kereséshez alapvető annak ismerete, hogy egy adatbázisban a mezők hiányosak. Ennek az adatnak azonnal elérhetőnek kellene lennie nemcsak a reklámanyagokban, de feltűnő helyen a képernyőn is az adatbázis indításakor. Az ilyen adatok figyelmeztetnék a használat például arra, hogy a legnagyobb példányszámú folyóiratok megtalálása az Ulrich's adatbázisban reménytelen, mivel a rekordok felében nincs ilyen adat, vagyis kizárnánk őket az ilyen keresésből. Legalább ennek az adatbázisnak az esetében a használat megtanulhatja, hány rekordban nem található meg egy bizonyos adatelem.

A The Serials Directory esetében a használatnak a semmi információja sincs erre vonatkozóan. A keresés korlátozására gyakran használt adatelemek nagymértékű hiánya drámai módon torzíthatja a keresési eredményeket. A hiányzó mezők hatásának tetőfoka az, amikor teljesen tönkreteszik a keresést - például akkor, amikor a megjelenési évet, a dokumentumtípust és a nyelvet használják a keresés korlátozására. Még tovább rontja a dolgot, ha a reklámanyag írója nincs tisztában ezekkel a korlátokkal, s olyan keresést ajánl, amely nulla rekordot eredményez. A Bowker's Complete Video Directory (BVCD) 2000 tavaszi kiadása illusztrálja ezt az esetet, s összegzi annak következményeit, ha több olyan mező hiányos, amelyet a keresés finomítására használnak.

A rekordok sok keresési szempontból hiányos voltának bénító hatása van a legtöbb keresésre. A rekordok teljességét igazolni akaró tesztkeresések (9.26. ábra) lehangoló eredményeket hoznak a legújabb kiadásban is. Bár tárgyi deskriptorokat a rekordok 95%-a kapott, a kiadás éve csak a rekordok 45%-ában van jelen, a videón való megjelentetés éve csak 28%-ban, a 21% a jelzi, hogy milyen nézőtípusnak szánják a filmet, a szín 63%-ban és a nyelv 6%-ban.



9.26. ábra: A teljesség tesztelése a Bowker's Complete Video Directory 2000 tavaszi kiadásában.

Ez a rendkívüli mértékű kihagyás nem gátolta meg a reklámszöveg íróját abban, hogy az adatbázis kereshetőségével próbálja eladni az adatbázist. Tényleg kereshetünk 19 keresési szempont szerint, de biztosan nem találjuk meg azt, amit szeretnénk, mivel az, hogy 19 keresési szempont közül sok hiányzik rekordok tízezreiből, drasztikusan leszűkíti a kereshető rekordok számát, amely tényről a használatnak a legcsekélyebb fogalma sincs.

A reklámszöveg írója az éjféli információs tévéprogramok túlradó stílusában próbálja rávenni a használatot arra, hogy állítson össze a videókról meghatározott feltételeknek megfelelő listát, például minden 1988-ban angol nyelven kiadott, PG minősítésű (csak szülő felügyelet mellett nézhető) vígjátékról. Kezdjük azzal, hogy a megjelenés éve csak a rekordok 28%-ában található meg, így a keresési tartomány egyből 47399 rekordra korlátozódik. Csak 3249 film felel meg az 1988-as megjelenés feltételének. A korhatár csak a rekordok 11%-ában található meg (nem is kell magas százalékot várni erre az adatelemre, mert sok filmnek nincs korhatár-jellemzője).

Ön 19 szempont szerint kereshet önmagában vagy kombinálva őket, vagy átnézheti a címjegyzék kilenc böngészhető indexét.

Megoldhatja akár a legnehezebb kérdéseket is, mint például „Tud olyan francia filmet találni, amelynek a címében szerepel a „blue” (kék) szó?

Megjelenítheti és kinyomtathatja több mint 5500 kritika teljes szövegét a Variety című folyóiratból, évente újabbak ezrei adódnak az adatbázishoz

Negyedévenkénti kumulatív aktualizálásokkal folyamatosan információt kaphat a legújabb videókról és kritikákról

Összeállíthatja videók olyan listáját, amely megfelel meghatározott szempontoknak, mint például „minden 1998-ban kiadott PG minősítésű angol nyelvű vígjáték”

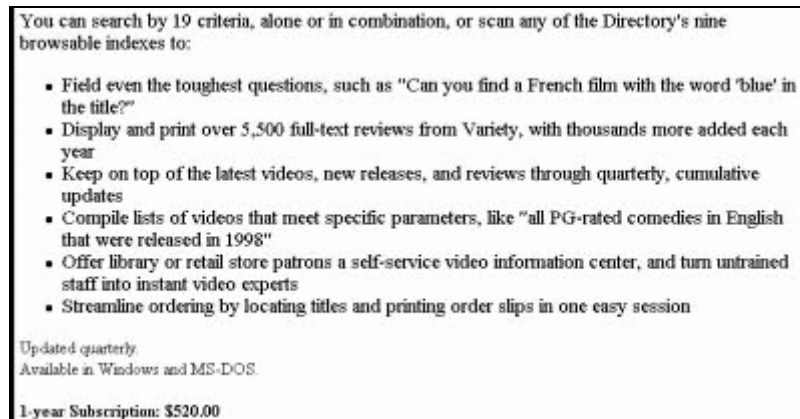
A könyvtárak és a kiskereskedelmi boltok ügyfeleinek önkiszolgáló videó információs központot ajánl, a képzetlen személyzetet rövid idő alatt videók szakértőjévé teheti

Megkönnyíti a rendelést, a címek megtalálásával és a rendelőlapok kinyomtatásával egyetlen könnyű folyamattá egyszerűsíti a megrendelést

Negyedévenként aktualizált.

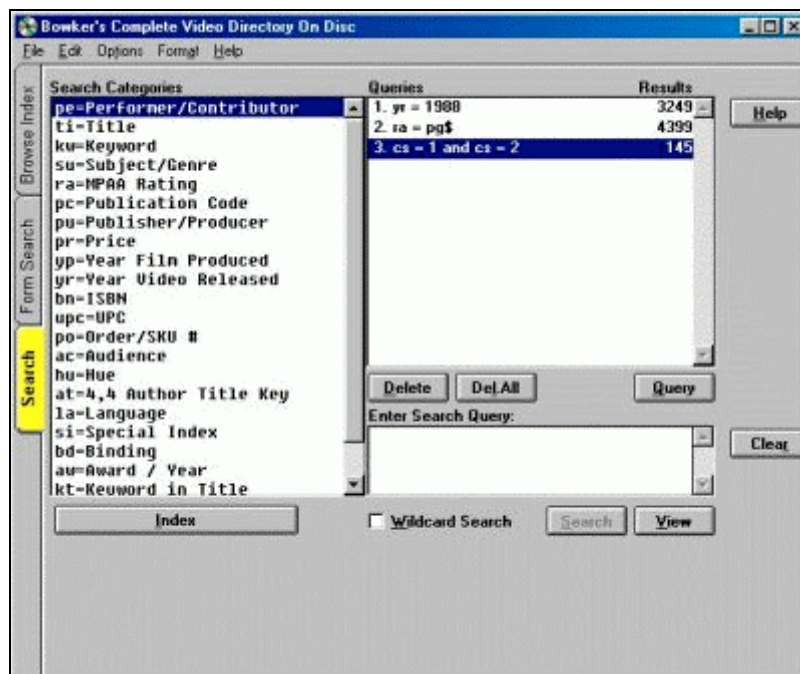
Windowsos és MS-DOS-os formában érhető el.

1 éves előfizetési díja 520 dollár.

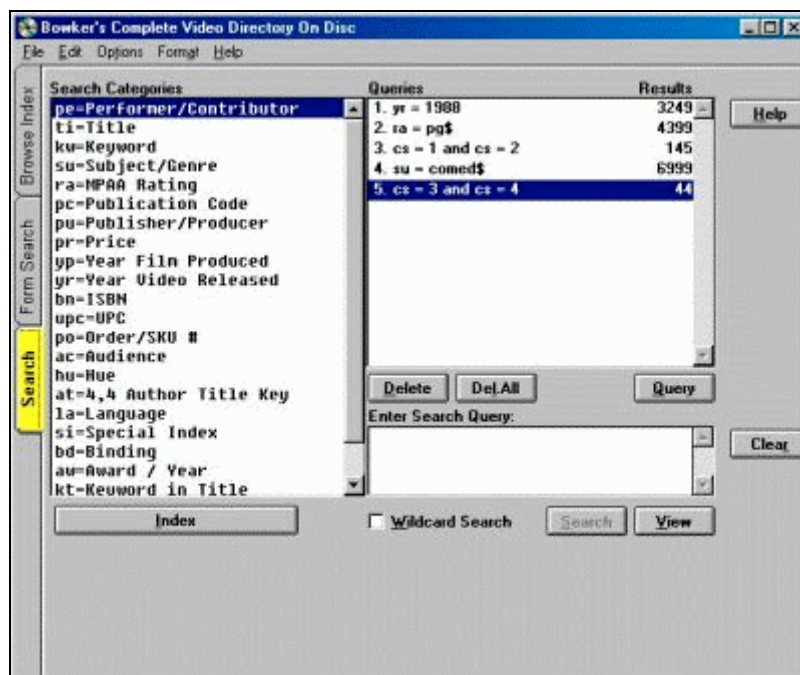


9.27. ábra: A katalógus kihívása és hallgatóságos ígérete - és a valóság.

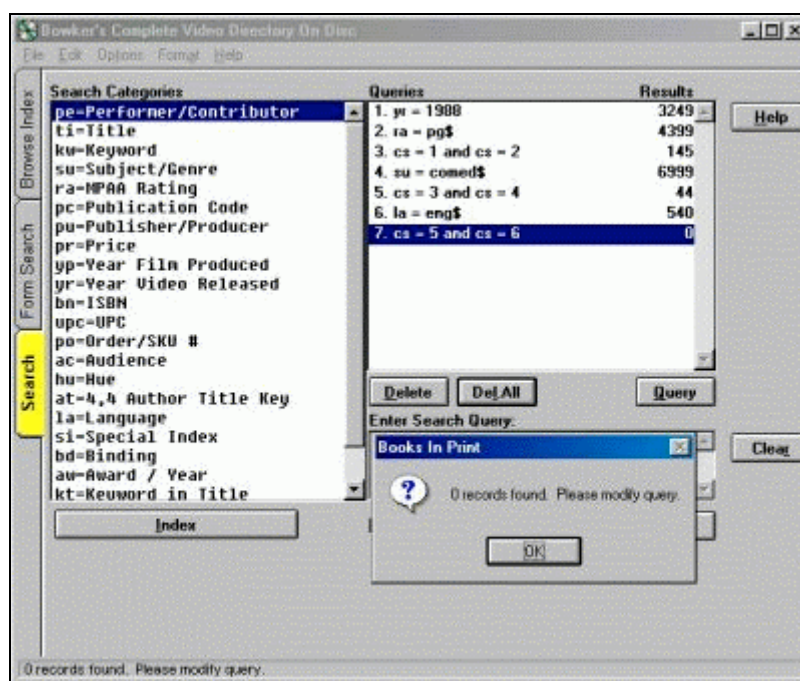
Ahogy az látható a keresés előrehaladásából (9.28. ábra), az első két kritérium szerinti kombináció a választékot 145 rekordra szűkítette. A vígjáték feltételének hozzáadása a keresést 44 filmre szűkíti (9.29. ábra). Az utolsó lépés teszi igazán vígjátékká a dolgot, amikor az angol nyelvre korlátozza a keresést, s nulla találatot kap (9.30. ábra). Az üzenet a *Books in Print* adatbázisra utal (ahogy a súgó nagy része), bár a videó címjegyzékben kerestünk, ahogy az a képernyő tetején látható.



9.28. ábra: A keresés első lépései.



9.29. ábra: A keresés szűkítése a vígjáték műfajára.



9.30. ábra: Nem kapunk találatokat, maga a keresés válik vígjátékká.

Az örömködő reklámszövegíró nem ellenőrizte a százalmas súgó fájlt (s attól félek, hogy a használók sem tennék ezt), amely azt mondja, hogy „az ENG (English, angol) kifejezés csak a többnyelvű könyvekre használatos” (9.31. ábra). Ha a *könyveket* filmekkel helyettesítjük, akkor ez többnyelvű filmeket jelentene. Csak 540 ilyen rekord van, és egyik sem 1988-ban megjelentetett vígjáték. Ezért nem kapunk egy találatot sem.

A **nyelvi kód** szerint való keresés lehetővé teszi önnek, hogy visszakeresse az összes nem angol nyelvű könyvre vonatkozó leírásokat.

ENG (English, angol) csak többnyelvű könyvek esetében használatos

Példák

La=fre or ger

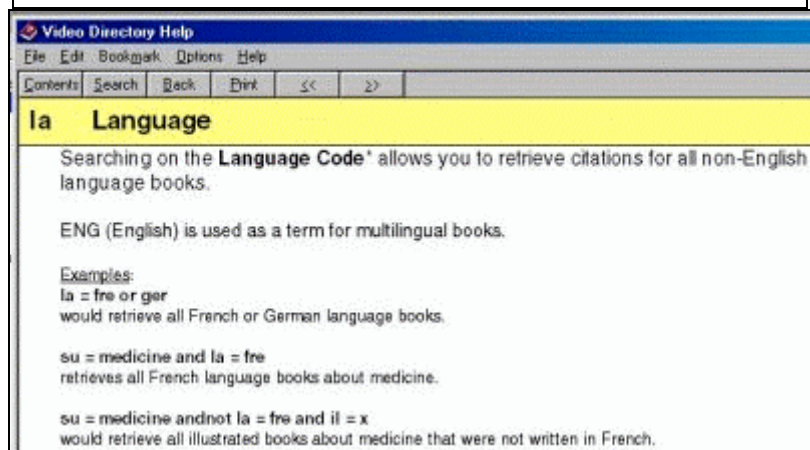
visszakeres minden francia vagy német nyelvű könyvet

su=medicine and la=fre

visszakeres minden francia nyelvű, az orvostudománnyal foglalkozó könyvet

su=medicine andnot la=fre and il=x

visszakeres minden olyan, az orvostudománnyal foglalkozó illusztrált könyvet, amely nem franciául jelent meg



9.31. ábra: A sűgő könyvekre utal, de azért ad némi útmutatást.

Egy dolog az, hogy az adatbázist ilyen szűrnyű állapotban tartják. Másik dolog a nonszensz reklámanyag, amelyet azonnal el kellene távolítani a webről. Nyilvánvaló, hogy a Bowkernek nem áll érdekében, hogy ezt megtegye, amíg elég könyvtár van, amely örömmel fizet 520 dollárt a BVCD-ért.

A sok információban gazdag, naprakész, pontos és ingyenes mozi adatbázis mellett a BVCD-nek nincs jövője. A könyvtárosok rá fognak jönni, hogy a Bowker jól ismert nevével egy szánalmas adatbázist ruházott fel.

10. A tárgyi indexelés minősége

A dokumentumok megfelelő tárgyi indexelése a hatékony visszakeresés előfeltétele. Ez egyértelművé vált a nyomtatott világban, ahol a tárgyi indexelés tételei voltak a referáló és indexelő folyóiratok leggyakrabban használt elérési pontjai az egyes füzetekben és a kumulált kötetekben is. A tárgyi indexelés fontossága valamennyivel csökkenni látszik az elektronikus adatbázisokban, mert a rekordok sok olyan tárgyi tartalmat hordozó adatelem szerint is elérhetők, mint a cím, a referátum és a tárgyi jellegű azonosítók mezője (subject identifier). A tárgyi indexelés minőségének mérése a legnagyobb súllyal szerepelt az adatbázisok minőségének értékelésében; ezért ez a fejezet többet foglalkozik a téma szakirodalmával, mint a többi.

Fogalmak és szakirodalom

Milstead (1994a) bizonyította az ARIST éves szemléinek tételei alapján, hogy az indexeléssel kapcsolatos kutatások mennyisége csökken. Míg az ARIST első évtizedében kötetenként 25-30 tétel volt az Indexing (indexelés) és rokon kifejezései alatt megtalálható, az elmúlt 20 év átlagában ez kötetenként 10-12-re csökkent. Meggyőzően érvel a mérőszámok olyan csoportja mellett, amely lehetővé tenné az indexelő rendszerek tesztelését és képes lenne kompenzálni az indexelés minőségének értékelésére elsődlegesen használt teljesség/pontosság módszertani hiányosságait. A tárgyi indexkifejezések minőségi értékelésének négy fő szempontja van:

- alaposság
- specifikusság
- pontosság
- következetesség

Soergel (1994) ezeknek az értékelési szempontoknak egy finomított definícióját adta, megkülönböztetve az entitás és a deskriptor nézőpontját az indexelés pontossága és következetessége szempontjából, továbbá az alaposság nézőpontját és fontosságát abból a szempontból, hogy a deskriptorok mennyire fedik le egy dokumentumban tárgyalt fogalmakat. (Vitatkozik „az indexelés mélysége kifejezés” használatával annak lehetséges többértelműsége miatt, de mivel sok szerző kedveli ezt a kifejezést, amikor a dokumentumhoz rendelt deskriptorok számára utalnak, ez a fejezet a kifejezést ennek a később tárgyalandó fogalomnak a jelölésére használja.) Soergel (1994) megtárgyalja mindezeknek az indexelési jellemzőknek a hatását a visszakeresési teljesítményre, s ugyanígy azokat az okokat, amelyek miatt tanulmányozni és értékelni kell az indexelésnek ezeket a mérőszámait.

A National Library of Medicine (az Egyesült Államok nemzeti orvosi könyvtára) végezte a kevés nagyszabású vizsgálatok egyikét, amelyben azt vizsgálták, hogy az operációs rendszerekben hogyan lehet számszerűsítve értékelni az indexelés minőségét. Griffith et al. (1986) a MEDLINE, a PsycINFO, a BIOSIS és az *Excerpta Medica* indexelésének minőségét hasonlította össze. Vizsgálatuk - amely egymáshoz szorosan kapcsolódó dokumentumok klasztereit használta - azt elemezte, hogy mennyire megfelelő módon kapcsolódnak az indexfogalmak a kapcsolódó dokumentumokhoz, milyen mértékben különböztetik meg általában az összekapcsolt dokumentumokat az egész adatbázison belül, és mennyire finoman az egyes dokumentumokat. Módszertanukat a különböző minőségű adatbázisok megkülönböztetési indexének kiszámítására Ajiferuke és Chu (1988) fejlesztette tovább. Átdolgozott mérési módszert használtak az indexelés minőségének értékelésére a Library Literature-ben, a LISA-ban és az ISA-ban (1989), de módszereik hasonló módon alkalmazhatók más adatbázisokra is.

Beaubien (1992) az IAC és az H.W. Wilson adatbázisai indexelésének minőségét tárgyalva azt állítja, hogy az IAC egyetlen kontrollált szótára sokkal „kontrolláltabbnak” tűnik, mint a Wilson próbálkozásai, de azt elismeri, hogy a Wilson hajlamos arra, hogy több pontos fogalmat használjon a következetesség rovására, és azt állítja, hogy a könyvtárnak el kell döntenie, vajon a használókat az IAC következetesebb indexelésével vagy a Wilson pontosabb indexelésével szolgálnák ki jobban.

Nincs szükség vizsgálatra ahhoz, hogy rájövünk, a pontatlan és következtelen tárgyi indexelésnek milyen hatása van a visszakeresési teljesítményre. Ezen két szempont ellenőrzésének eredményei az adatbázisok valódi kontrollált szótárainak indextételeiben önmagukért beszélnek, ahogy azt Jacsó illusztrálta (1998c). Martin és Bergerhoff (1991) a Chemical Abstracts Service adatbázis kontrollált kifejezés indexének A betűvel kezdődő kifejezéseit elemezte, s összehasonlította ezt a nyomtatott tárgyszólista érvényes kifejezéseivel. Az utóbbiban mindössze egy elírást találtak.

1713 kifejezés jelent meg azonban az online változatban, amely nem volt benne a nyomtatott kontrollált szótár mintatételeiben, az okok még elszomorítóbbak voltak. Az elírások 9,4%-ot adtak, az esetek 6,8%-ában a javasolt egyes vagy többes számú forma helyett a másikat használták, 61%-ot nem lehetett megtalálni a nyomtatott szójegyzékben. Ezek az adatok azt sugallják, hogy sok kifejezést használtak deskriptorként bármilyen ellenőrzés nélkül.

A MEDLINE tárgyi indexelésének jó minőségét bizonyította indirekt módon a Drexel projekt, amelyről McCain, White és Griffith (1987) számolt be. Azt tapasztalták, hogy a MEDLINE-ből vissza nem keresett rekordok esetében az ok legtöbbször az volt, hogy nem dolgozták fel az adott folyóiratot, egy adott számot vagy cikket, s nem a nem megfelelő vagy pontatlan indexelés.

A pontatlan indexelés néha nem az indexelő hibája, hanem a kontrollált szótár megalkotóié. Miller (1996) illusztrálja azt, hogy a tezaszuszkészítés alapvető szabályait hogyan sértették meg a Sports Thesaurus 1994-es kiadásában. Tipikus hiba például a deskriptorok közti kapcsolatok téves értelmezése, homonimák és szinonimák törlése, a földrajzi nevek jelentős történeti változásainak figyelmen kívül hagyása, ezek pontatlan indexeléshez vezettek, alacsony pontossághoz és alacsony teljességhez. A problémákat később súlyosbították a Sports Thesaurus nyomtatott és CD-ROM-változatának következtelenségei, eltérései.

Nehéz számszerűen meghatározni vagy objektíven megítélni a kontrollált szótárakban található kifejezések vagy a rekordokhoz való hozzákapcsolásuk megfelelőségét (vagy mindkettőt). Mindazonáltal egy Brooks által (1993) végzett, a korlátlan aliasing stratégiájának érvényességét igazolni akaró kísérlet járulékos eredménye figyelemre méltó megfigyeléseket szolgáltatott az indexelés minőségéről.

Ebben a kísérletben az ERIC, a LISA és az ISA adatbázisokból választották ki ugyanakkor a 21 dokumentumnak a rekordjait. A deskriptorokat összegyűjtötték, s a deskriptorok listáját különböző kombinációkban (mint például csak ERIC, ISA/LISA keverék és ERIC/ISA/LISA keverék) mutatták meg minden rekord esetében, s mellette az egyes rekordok címét és referátumát.

Mérték, hogy mennyi időre volt szüksége 40 keresőnek a legjobban összeillő deskriptorok megtalálására. Az eredmények, amelyek cáfolták a korlátlan aliasing előnyét, amelyet a korábbi vizsgálatok állítottak, azt is jelezték, hogy a rekordoknak az egyetlen ISA index-fogalommal való összepárosítása sokkal hosszabb átlagidőt, több válogatást tett szükségessé, és sokkal több hibát okozott, mint a többi szimpla típusú indexek. Ezek az eredmények azt sugallják, hogy „az ISA indexe rosszul teljesít,” és a keresőket „akadályozta az ISA indexe, amely kevés... egyszerű szót vagy kifejezést ajánlott.”

Sievert és Siebert (1991) a FRANCIS és a Philosophers's Index átfedésének összehasonlítása során azt tapasztalta, hogy a 104 vissza nem kereshető, de az utóbbiban benne levő rekord közül 49 azért volt visszakereshetetlen, mert a megfelelő fogalmat vagy nevet nem indexelték. Míg a tárgyi jellegű tárgyszó meghatározásának megfelelő volta szubjektív, az index-fogalomként hiányzó nevek egyértelműen indexelési hibák a Philosopher's Index tezaurusának világos útmutatója szerint. Ez kötelezővé teszi „az indexelést a személy neve szerint, ha két oldal vagy a cikk fele foglalkozik a nevezett személlyel.”

Egy vizsgálatban, amely nagyon egyedi volt a vizsgált témák szempontjából, Braam és Bruil (1992) azt vizsgálta, hogy a szerzők milyen mértékben voltak elégedettek a cikkeikhez a Chemical Abstracts Services (CAS) által megadott indexkifejezésekkel. A kérdőívre válaszoló 211 szerző közül csak 52% értett egyet azzal az állítással, hogy a CAS indexelői által a cikkeikhez megadott indexkifejezések megfelelő módon reprezentálták a témát. 48% szerette volna megváltoztatni az indexkifejezések listáját (hozzáadva, törölve vagy kicserélve). A CA szekciókódok megfelelőségével kapcsolatban a szerzők 19%-a gondolta úgy, hogy cikküket nem a megfelelő fő CA szekcióba sorolták be.

Giral és Taylor (1993) - más jellemzők mellett - összehasonlították a legáltalánosabb elérési pontokat (megnevezett személyek, földrajzi nevek, tárgyi deskriptorok stb.) az ugyanazokról a cikkekről készült rekordokban az Avery Indexben és az Architectural Periodicals Indexben. Azt állapították meg, hogy csak a személy- és cégnevek 55%-a, a földrajzi nevek 50%-a és az építészeti helyek általánosan használt nevének 40%-a egyezett meg a két forrásban, vagy - a földrajzi nevek esetében - legalább részben. (Ugyanazoknak az összetevőknek a használatát, függetlenül sorrendjüktől [azaz Portland (Maine) és Maine-Portland] megegyezőnek tekintették.)

A Chemical Abstracts Services Registry Number (CASRN, a Chemical Abstracts Services regisztrációs száma) helyes meghatározását tesztelte Buntrock (1994) hat adatbázisban, majd később újra Buntrock (1995) három adatbázisban: Azt tapasztalta, hogy azokban az adatbázisokban, amelyek egy algoritmust használnak a kémiai nevek kiválasztására a teljes szövegből vagy a referátumokból, és ezt párosítják egy CASRN számmal, a pontatlanságok eltérése sokkal magasabb, mint azokban az adatbázisokban, amelyeknek a készítői egy „élő” CAS regisztrációs rendszert használnak.

Még rosszabb a hatékony indexelés és keresés esélye, ha a kontrollált szótár nem hozzáférhető. A kontrollált szótár a minőségi indexelés alapfeltétele (egy, a szélesebb osztályozást szolgáló osztályozási rendszerrel együtt). Jacsó (1997b) azt állítja, hogy a kontrollált szótár hiánya az Information Science Abstracts korai éveiben rendkívül károsan hatott az ISA tárgyi indexelésének minőségére.

Hood és Wilson (1994) a LISA-nak a példástól messze álló indexelési politikáját és gyakorlatát annak a ténynek tulajdonítja, hogy „azt a tezauszt (vagy kontrollált szótárt), amelyet a LISA használ, nem a szokásos tezauszkészítési alapelveknek megfelelően állították össze.” Megállapítják, hogy „a LISA online használói kézikönyvének első kiadása, amely a tezauszt az előnyben részesített kifejezések listájának (Preferred Terms List) nevezi a LISA adatbázisban található DE kifejezések forrásának őszintébb értékelése.” A LISA adatbázis új tulajdonosa, a G. K. Saur 1992-ben abbahagyta a LISA szótárának használatát, s egy új kontrollált szótárt hozott létre. Bár ezt nem adták ki nyomtatásban, elérhető az adatbázisnak a messze legjobb CSA-féle változatában.

Még a szakszerűen összeállított tezauszok is akadálnak bizonyulhatnak a kezdők számára, akik nem képesek megérteni, még kevésbé keresésben alkalmazni egy olyan kifinomultan strukturált tárgyszó listát, mint amilyen például a MeSH. Killion (1995) magyarázza meg, hogy ez volt az egyik indoka annak, hogy az RNdex Top 100 adatbázis kifejlesztése többek

között épp abból a megfontolásból eredt, hogy az indexelők és keresők egy, a MeSH-nél vagy a CINAHL Thesaurusnál egyszerűbb tezaurszt használhassanak. További okok is voltak, mint a kifejezések nagyobb frissessége, jobb tárgyköri megjegyzések, kiterjedt utalórendszer és természetesebb tárgyszavak, amelyek az invertált formák helyett az eredeti szórendet alkalmazzák.

Az indexelés mélységével (alaposságával) kapcsolatban nincsenek általánosan érvényes bűvös számok. Sparck Jones első átfogó vizsgálata (1973) megmutatta, hogy az indexelés mélysége számít; ez növeli a teljességet (a Boole-operátorokat használó keresőrendszerekben), de egy bizonyos ponton túl drasztikusan csökkenti a pontosságot.

Bár az indexelés alaposságának mértéke általában arányban van a dokumentumhoz hozzárendelt deszkriptorok számával, a deszkriptorok magasabb száma nem garantálja a megfelelően alapos indexelést. Az indexelés mélysége pontos jellemzője lehet azonban az alaposságnak, különösen akkor, amikor két vagy több indexelő szolgáltatásnál hasonlítjuk össze az ugyanazokhoz a dokumentumokhoz hozzárendelt indexkifejezések számát.

Amikor a CINAHL és a MEDLINE adatbázisokban ugyanazoknak az ápolással kapcsolatos cikkeknek az indexelési gyakorlatát hasonlították össze, Brenner és McKinin (1989) úgy találták, hogy míg mindkét adatbázisban ugyanannyi (3) fő deszkriptor van, a MEDLINE háromszor annyi mellékdeszkriptort használt, mint a CINAHL. Fontosabb azonban az az eredmény, hogy a közös fő deszkriptorok átlagos száma kevesebb mint 0,5 volt, jelezve azt, hogy kismértékű volt a megegyezés a fő deszkriptorok kiválasztásában, annak ellenére, hogy a CINAHL deszkriptorainak 70%-a a MeSH-ből származik.

A közös alapon való feltárás hiányának súlyos hatása van a több adatbázisban való keresésre. Ez a tanulmány nem vizsgálta a fő deszkriptorok kiválasztásakor tapasztalható össze nem illés okait, de világosan illusztrálta, hogy a CINAHL használói, akik kettőnél több deszkriptort használnak keresőkérdésükben a Boole-algebra ÉS operátorával összekapcsolva, nagy mértékben csökkentik a visszakeresés lehetőségét.

Ahmad (1991) napilap indexek nemzetközi választékának indexelési gyakorlatát vizsgálta. Azt állítja, hogy a napilap indexelésének széleskörűsége (alapossága) és következetessége a híryananyagok tárgyi elemzésének hatékonyságától függ, míg a rekordokhoz adott indexkifejezések száma - amely egytől két tucattig terjedt - nem meghatározó tényező az indexelés minőségében. Jacsó (1992a) azt állítja, hogy a tárgyszavak számának egyre csökkentése a Magazin Article Summaries (MAS) adatbázisban (a korlátot 1993-ban megszüntették) nem biztosította az indexelés megfelelő mélységét. Azt is megjegyzi, hogy a deszkriptor kifejezések magas száma a LISA-ban (ez jellemezte a rekordok többségét 1994-ig) sem volt garancia az indexelés minőségére. Azt állítja, hogy valójában a rendkívüli redundancia volt a legfőbb hiba a LISA adatbázis indexelésében.

Ugyanazoknak a cikkeknek az Avery Indexben és az Architectural Periodicals Indexben (API) található rekordjait összehasonlítva Giral és Taylor (1993) azt elemezte, hány deszkriptort és személyre vonatkozó hozzáférési pontot adtak meg a két forrásban. Az Avery Index mintájából a mindkét forrásban szereplő 444 rekordhoz az Avery indexelői 687 név szerinti hozzáférési pontot határoztak meg, az API indexelői 688-at. Az API mintájából mindkettőben meglevő 671 rekordhoz az Avery indexelői 1192, az API indexelői pedig 1203 név szerinti hozzáférési pontot határoztak meg.

A számok nagyon impresszívek lehetnek, de önmagukban félrevezetőek. A neveknek majdnem egyharmada nem szerepelt mindkét forrásban. A nevek formájában az egyezés csak 55% volt (meg kell jegyezni persze, hogy a megegyezés szempontjai nagyon szigorúak voltak, csak a kis- és nagybetűs eltéréssel nem foglalkoztak). A tárgyi deszkriptorok

rekordonkénti átlagos száma nagyon közel volt egymáshoz, s az egyik forrásban megtalálható deszkriptorok első eleme az esetek 40%-ában jelen volt a másik forrás deszkriptorai között, annak ellenére, hogy a források különböző tezauruszokat használtak.

Turner (1995) összehasonlító vizsgálatot végzett mozgóképek tárolására és visszakeresésére a használók által, illetve az indexelők által adott kifejezésekkel kapcsolatban. A kanadai nemzeti filmes testület (National Film Board of Canada) felvétel gyűjteményének felvételenkénti indexelését használták arra, hogy megállapítsák, milyen mértékű a megegyezés azok között a kifejezések között, amelyekre a használók gondolnak, amikor film- és videofelvételeket keresnek, s amelyeket a hivatásos indexelők adnak meg. A 81 résztvevő 44 filmfelvételt nézett meg, s olyan szavakat és kifejezéseket adtak meg hozzájuk, amelyekről azt gondolták, hogy később kézenfekvőek lesznek a visszakeresésre önmaguk és mások számára. Az eredmények nagyfokú konzisztenciát mutattak az indexelők és a használók között, továbbá magasabb szintet a pre-ikonográfia (mihez kapcsolódik), mint az ikonográfiai (miről szól) szinten. Ez megerősítette Layne (1994) hipotézisét, hogy több megegyezés van egy képnek az alapvető és objektívebb aspektusaiban, s kevesebb a másodlagos és szubjektív aspektusokban.

Mind az osztályozási rendszerekben, mind a kontrollált szótárakban szükség van változásokra az idők folyamán, hogy specifikusabb terminológiát és neologizmusokat vezethessenek be. Ahogy Mintz (1990) megfigyelte, a keresők nem igazán örülnek, amikor tudomást szereznek ilyen változtatásról, mert egy új osztályozási rendszer bevezetése, illetve kifejezések deszkriptorként való elfogadása után a korábbi rekordokban visszamenőlegesen nem vezetik át ezeket a változásokat. Példásnak nevezhető, de ritka az a gyakorlat, amelyet a MEDLINE és a CINAHL folytat, hogy a tezauszt évenként aktualizálják, és az új tárgyszavakat visszamenőlegesen hozzárendelik a régiekhez. Az ilyen konverziók hiányában azonban Mintz (1995) figyelmeztetése szerint a használó felelőssége az, hogy a keresési stratégiát egy adatbázis teljes időtartama alatti terminológiai különbségekhez igazítsa. Amikor új deszkriptorokat vezetnek be, az H. W. Wilson cég azokat a kifejezéseket visszamenőleg hozzáadja a létező rekordokhoz. Ennek jelentős költségei vannak, de megkönnyítik a használók dolgát.

Az indexelés mélységét vagy alaposságát, illetve specifikusságát elsődlegesen a visszakeresésre gyakorolt hatással összefüggésben vizsgálták. Hogy mi az indexelés optimális mélysége és specifikussága, az mindig az egyedi dokumentumoktól és a célközönségtől függ; ezért nehéz egyetértésre jutni a deszkriptor kifejezések optimális számát és specifikusságát illetően. Svenonius és McGarry (1993) azonban igazolták, hogy a tárgyszavak objektív értékelése valóban megoldható.

Hipotézisük az volt, hogy legalább az esetek 80%-ában egyértelműen meg lehet határozni, hogy a Library of Congress Subject Headings (LCSH, a Kongresszusi Könyvtár tárgyszavai) hozzárendelése jó vagy rossz volt. Ezt megerősítette az, hogy a szerzők 100 tudományos témákkal foglalkozó könyvhöz választottak tárgyszavakat, annak tudatában, hogy a tárgyszót csak akkor lehet megadni, ha a könyvnek legalább a 20%-a a tárgyszóban jelzett témáról szól. 202 tárgyszó közül csak hat olyan volt, amelynek a kiválasztásában és formájában a szerzők nem értettek egyet.

Az elvárható deszkriptorok kihagyása jelentős mértékben csökkentheti az eredményeket (a találatok számát), ha a használók a keresésüket a deszkriptor mezőre szűkítik, hogy biztosítsák a pontosság megfelelő szintjét. Jacsó (1992a) a LISA-nak egy olyan részalmazát elemezte, amely a *CD-ROM* kifejezés minden lehetséges variációját tartalmazta (azaz kötőjellel vagy kötőjel nélkül, szóközzel vagy szóköz nélkül, dőlt vonallal vagy anélkül írva) a címekben, de a deszkriptorok mezőjében nem. Úgy találta, hogy a megfelelő deszkriptor elhagyása különösen hátrányos annak fényében, hogy a LISA rekordjaiban rengeteg a túl

széles és redundáns deszkriptor. Bár egyes indexelési szabályok azt tanácsolják az indexelőknek, hogy kerüljék a redundanciát azzal, hogy kihagyják azokat a deszkriptorokat, amelyek benne vannak a címben, ez megkérdőjelezhető és nem alkalmazható a LISA esetében, mert a nyomtatott változatnak hozzáférési pontokat kell adnia a tárgyszavakon keresztül, ezek pedig egybeeshetnek a címben szereplő kifejezésekkel.

Bár a túlindexelés nem olyan ártalmas, mint az alulindexelés, csökkentheti a keresés eredményének relevanciáját. A deszkriptorok rendkívüli redundanciája semmissé teheti az indexelés előnyeit. Jacsó (1995) azt állítja, hogy a LISA esetében a mennyiség nem csap át minőségbe, sőt még gyengíti is azt, ami jó indexelés lehetne. Feltételezése szerint az, hogy a LISA-ban rekordok tízezreiben kétszer, háromszor és négyszer fordul elő ugyanaz a deszkriptor, annak a következménye, hogy a deszkriptorokat automatikusan generálják a LISA-ban használt osztályozási jelzetekből, mert valószínűtlen, hogy indexelést végző emberek ilyen rendkívüli mértékben redundánsak lennének.

Az indexelés specifikusságával kapcsolatban Hernon és Metoyer-Duran (1992) érdekes eredményre jutottak. Öt felsőoktatási és szakkönyvtár 43 könyvtárosával folytatott fókusz-interjúban az interjúalanyok visszatérően kritizálták az H. W. Wilson cég Library Literature című adatbázisát amiatt, hogy a tárgyszavak túl általánosak és túl kevés van belőlük. Jelen szerző véleménye ennek ellenkezője, szerinte a tárgyszavak és altárgyszavak kombinációja túlságosan specifikus a Library Literature-ben. A kontrollált szótár egyszerű böngészése is egyértelművé teszi, hogy az indextételek többsége egyke (azaz csak egy rekordhoz kapcsolódik hozzá.) Ez a túlzottan specifikus indexelés tipikus szimptomája.

Hood és Wilson (1994) sokkal tovább ment annál, hogy mintát választanak ki a rekordok közül az alaposság és a specifikusság vizsgálatokhoz. Egyéb jellemzők mellett az indexelés alaposságát és specifikusságát a LISA rekordjainak teljes populációján vizsgálták, nem egy mintán. Megnézték a LISA minden rekordját, és - miután kiszűrték az ugyanabban a rekordban többször előforduló deszkriptorokat - 28191 egyedi deszkriptor 669403 előfordulását vizsgálták. A szerzők úgy találták, hogy 1979 óta a LISA következetesen átlag hat vagy hét deszkriptor kifejezést kapcsolt egy rekordhoz, de a deszkriptorok 64%-a csupán egyszer fordult elő az egész adatbázisban. Mérsékelt hangnemben azzal a következtetéssel összegezték munkájukat, hogy a LISA indexelési politikája és gyakorlata messze van a példástól.

A leíró katalogizálás és a bibliográfiai leírás megítélésével szemben a tárgyi indexkifejezések kiválasztásának értékelése rendkívül szubjektív. Még nehezebb a tárgyi indexelés következetességét egy adatbázison belül megítélni (sokkal inkább, mint a különböző adatbázisok indexelőinek következetességét.) Nem csoda, hogy viszonylag kevés cikk jelent ugyanazon indexelő és referáló szolgáltatás indexelési következetességéről.

Leininger (2000) vizsgálata, amelyben a PsycINFO adatbázisba véletlenül kétszer bekerült 60 duplikátum indexelésének következetességét tesztelte, kivétel a témával foglalkozó kisszámú kurrens publikáció között. Cooper gyakran idézett 1969-es kritikája (a következetes indexelés csak következetesen rossz lehet) elbátortalaníthatta a kutatókat, hogy a következetességgel kapcsolatos vizsgálatokba fogjanak bele. Fugmann (1992) azt állítja, hogy Cooper állítását sohasem cáfolták meg (s úgy tűnik, hogy egyet is ért vele).

Mindenesetre megállapítja, intuitívan érezhető, hogy a (tárgyi) indexelésnek valamilyen mindeddig feltáratlan módon kapcsolatban kell lennie az indexelés minőségével és a keresés hatékonyságával. Azután azzal érvel, hogy „a következetesség helyett a megjósolhatóságnak ... kell az indexelés céljának lennie mind a dokumentumok lényegének meghatározásában, mind ennek a lényegnek a leírásában.”

Ugyanazon dokumentum indexelése következetességének mérése olyankor, amikor ugyanaz a személy vagy ugyanaz az adatbázis végzi ezt különböző időpontokban (magához mért indexelői következetesség) vagy két vagy több személy vagy indexelő szolgáltatás (indexelők közötti következetesség), azonos vagy hasonló kontrollált szótárt és indexelési gyakorlatot használva, nagyon népszerű kutatási téma. Ennek egyik oka az, hogy az ilyen kutatás egy adott forrás következetességének mértékére is fényt deríthet. A sok indexelők közötti következetességet vizsgáló tanulmány - Leonard (1977) szemlélte őket az 1954 és 1975 közötti periódusra, Markey (1984) pedig 1984-ig - egyértelműen azt sugallja, hogy közvetlen korreláció van a visszakeresés hatékonysága és az indexelők közötti következetesség között.

Chan (1989) angol nyelvű munkáknak ugyanarra a kiadására vagy számára vonatkozó 100 rekordpárt hasonlított össze, a párok egyik tagját az OCLC tagkönyvtárai által beszállított rekordok szolgáltatták, a másikat a Kongresszusi Könyvtár rekordjai, olyan művekről volt szó, amelyekben nem volt Cataloging in Publication (CIP, katalogizálás a kiadványban) információ. A cél az volt, hogy meghatározzák Kongresszusi Könyvtár adott tárgyszavainak variációit. A kis minta azért volt elfogadható, mert a tanulmány elsődleges célja az volt, hogy teszteljék a következetesség értékelésének módszertanát.

A tanulmány azt állapította meg, hogy a rekordpárok 15%-a egyezett meg tökéletesen (azonos tárgyszavak, vagy hat pár esetén a tárgyszavak hiánya). A részleges egyezések 80%-ot tettek ki (egyes tárgyszavak nem egyeztek meg teljesen vagy részlegesen, például különbségek voltak az alosztályokban). Öt százalék volt teljesen különböző (két esetben az volt az ok, hogy a nem a Kongresszusi Könyvtárból származó rekordban nem volt tárgyszó).

Míg az ideális a teljes következetesség lenne, ezt csaknem lehetetlen elérni. A 80 részlegesen egyező pár elemzése megmutatta, hogy az ugyanazzal a témával foglalkozó rekordok többségében voltak közös tárgyi szavak, amelyek képesek garantálni a visszakeresés részleges sikerét, ha olyan rendszereket használnak, amelyek lehetővé teszik a kulcsszó szerinti keresést a tárgyszavak mezőjében. Ahol nem volt egyezés, annak az esetek többségében az volt az oka, hogy nem megfelelően kapcsolták össze a tárgyszavakat és altárgyszavakat. Megfontolandónak tartja, hogy az LCSH-nak a döntően prekoordinált rendszerről áttérjen a posztkoordinált rendszerre. Az, hogy nagyon alacsony volt az olyan rekordok száma, amelyek teljes következetlenséget mutattak, nem tesz lehetővé semmilyen általánosítást.

Reich és Biever (1991) az indexelők közötti következetességet tárgyalva azt mérték, hogy egymástól függetlenül hány teljesen azonos kifejezést adtak ugyanazokhoz a cikkekhez az AGRICOLA-ban és a CAB-ban, a CAB Theasurust használva. (A tezasurust kis mértékben módosította a National Agricultural Library (az USA országos mezőgazdasági könyvtára), amikor az AGRICOLA számára adoptálta, hogy kövesse az amerikai helyesírást és beiktassa CAB tezasurusa által le nem fedett szakterületek további kifejezéseit.) Két folyóiratból összesen 236 cikket választottak ki, s azok deszkriptorait hasonlították össze. Bár a két indexelő szolgáltatás által adott deszkriptorok száma nem mutatott lényeges különbséget (8,2 kontra 8,9), a cikkenként adott deszkriptorok száma kevés esetben mutatott azonos indexelési mélységet amely megmagyarázhatja az egyezés alacsony, 27%-os mértékét. A másik ok az, hogy ugyanarra a fogalomra szinonimákat lehet érvényes deszkriptorként megadni (pl. cold stress és winter hardiness vagy developmental stages és growth stages), s ez nem segíti elő az indexkifejezések következetes meghatározását.

Tonta (1991) 82 könyvtár- és információtudományi tétel indexelésének következetességét hasonlította össze a Library of Congress (LC, Kongresszusi Könyvtár) és a British Library (BL) között. Olyan művekről volt szó, amelyekhez a British Library a PRECIS kifejezések láncolata mellett a Kongresszusi Könyvtár tárgyszavait is megadta. A tételekhez adott kifejezések száma jelentősen különbözött (3,44 az LC-nél és 1,55 a BL-nél), mert a BL-nek

nem kellett kizárólagosan a Kongresszusi Könyvtár tárgyszavaira alapoznia. A különbségek ellenére a tételek 21%-a esetében ugyanannyi tárgyszót adtak, de 74%-ban a LC katalogizálói adtak több LCSH kifejezést. A BL által adott tárgyszavak majdnem 40%-a teljesen megegyezett az LC által adott tárgyszavakkal, s további 35% részlegesen egyezett meg (a fő tárgyszó megegyezett, az alosztások azonban nem.) Az indexelés következetességének mértékét a teljes és a részleges egyezésre 36%-osnak találták - elég alacsony érték, amely mindazonáltal igazolta a korábbi vizsgálatok eredményeit. Az eredménynek fontos következményei vannak a nem eredeti rekordot létrehozó katalogizálók számára, akik a másik ügynökség rekordjait használják az Atlanti-óceán két oldalán.

Sievert és Andrews (1991) adatbázison belüli következetességi vizsgálatot végzett az Information Science Abstracts használatával. 71 azonos duplikátum rekord-párt választottak ki négy folyóiratból, hogy a deszkriptorok és azonosítók (identifikerek) következetességét vizsgálják. A fő tárgyszavak 52% következetességet mutatott, az altárgyszavak következetessége 45,5% volt, így a következetesség átlagosan 48% volt. Az azonosítók következetességét 33%-osnak találták. Ezeket a számokat annak a ténynek a fényében kell értelmezni az ISA esetében, hogy nagyon kicsi a szójegyzék, amelyből az indexelők kifejezéseket választhatnak, s csak egy fő tárgyszót lehet megadni. Mindkét tényező növeli a megegyezés valószínűségét.

Minél kevesebb kifejezés van, annál valószínűbb, hogy a két indexelő ugyanazt a deszkriptort választja, ahogy az indexelés mélysége nő, következetessége csökken. Bár a minta viszonylag kicsi volt, az eredmények megerősítették annak a hasonló vizsgálatnak az eredményeit, amelynek során a MEDLINE-ban Funk, Reid és McGoogan (1983) alacsony indexelők közötti következetességet találtak olyan dokumentumokra vonatkozóan, amelyeknek tökéletes volt a szemantikai hasonlósága. Ezt használta az eredmények magyarázatának tényezőjeként Harter, Nisonger és Weng (1993) a hivatkozó és hivatkozott cikkek közötti szemantikai kapcsolatokra vonatkozó újszerű kutatásukban.

Az indexelési minőség kiértékelésének folyamata

Az elmélet gyakorlatba való ültetésének legjobb módja néhány próbakeresés lebonyolítása. Az indexelés minősége értékelésének első lépése azonban megelőzi a keresést. Mivel az indexelés a tezaurusz minőségétől függ, vagy legalább a használt kontrollált tárgyi szótártól, amelyek alapján indexfogalmakat keresnek és rendelnek a dokumentumokhoz az indexelők, ezért nagyon hasznos az adatbázis kontrollált tárgyi szótárának tanulmányozása. Sajnos sok adatbázis-készítő nem publikálja ezt nyomtatásban (például a mintacsoportban az H. W. Wilson), de legalább online vagy CD-ROM-on lehet tanulmányozni, ha az információszolgáltató elkészítette ezt. Bár ez a szoftver kérdése, jegyezzük meg, hogy jelentős különbségek vannak ugyanannak a tezaurusznak a különböző információszolgáltatók által készített változatai között

Az Ovidnál található tezauruszok megjelenési formája a legintuitívabb és leginformatívabb az online és CD-ROM-kiadók közül, ezt a CSA kiváló változata követi. A DIALOG adatbázis-gyűjteményéből elszomorítóan kis százalékban van meg a tezaurusz funkció, s gyakran ezek is az elvárt szint alatt maradnak. Példa erre az, ahogyan az ERIC tezaurusza tárgyi megjegyzéseit kezelik, s gyakran el is hagyják őket. A SilverPlatter-nél a Windows alatt működő CD-ROM és webes verziókban jó változatai vannak a különböző tezauruszoknak.

A legjobb az, ha előzetesen tanulmányozzuk a tárgyi szótár nyomtatott változatát, elolvassuk az előszót, megvizsgáljuk a tételeket és az utalókat annak érdekében, hogy benyomást szerezzünk arról, milyen az adott változat.

A kontrollált szótár mérete önmagában nem döntő tényező, bár egy adott diszciplínában a túlságosan vékony listák, mint a kifejezések listája az ISA-ban és az MHA-ban, azonnal kétséget ébreszthetnek. Multidiszciplináris adatbázisokban nehéz egy adott terület terminológiai lefedettségét megállapítani. Könyvtár- és információtudományi és -technikai fogalmak megtalálása az INSPEC-ben (amely alapvetően elektro- és elektronikai mérnöki adatbázis), az ERIC-ben (elsődlegesen oktatási adatbázis), vagy az ABI/INFORM-ban (üzleti adatbázis, amely az információtechnikát is nagyon jól feltárja) időigényes folyamat.

Azt várnánk, hogy a könyvtár- és információtudományi adatbázisok példás kontrollált szótárral rendelkeznek a szakterületre. A LISA és az ISA biztosan nem felel meg ennek a hipotézisnek. Egyikük sem szabványos a szótára, de legalább a LISA 1993-ban egy újra váltott. Az ISA sokkal hamarabb átválthatott volna az ASIS információtudományi tezaurusára (*Thesaurus of Information Science*).

A LISA-t 1992 végén szerezte meg a G. K. Saur a brit könyvtárosegyesülettől (Library Association, LA), s egy éves szünet után egy teljesen átdolgozott és sokkal jobb kontrollált szótárral jött elő. Am továbbra is ismerni kell a régit kontrollált szótárat, mert több mint 20 év anyagát azzal indexelték. Azoknak a használóknak, akik retrospektív keresést végeznek, az 1993 előtti időszakra kiterjesztve, együtt kell élniük ezzel a „múmia átkával”, ahogy Reva Basch (1990b) érzékletesen nevezte az adatbázisiparnak ezt az örökségét. A szótár utolsó nyomtatott változata, amelynek megmagyarázhatatlan módon a *LISA Thesaurus* címet adták, 6000 tételt tartalmaz. Ez valójában nem tezaurusz: maga a LISA online kézikönyve is megállapítja, hogy „ez nem kíván a LISA-ban 1969 óta használt kifejezések kimerítő listája lenni, csak a legfontosabb és legszélesebb körben használt kifejezések alapvető listáját szolgáltatja.”

Még így is többnek mutatja magát, mint ami. Kétségtelenül nem tezaurusz, amely sok és következetes utalót; szűkebb, bővebb és kapcsolódó fogalmak listáját; tárgyi és a használatra vonatkozó megjegyzéseket ad. A LISA alkalmanként ad valamilyen tárgyi megjegyzést és az ajánlott kifejezésekhez vezető utalásokat, de sok ilyen tétel elég abszurd. Egyetlen oldal (10.1.) is jól illusztrálja, milyen hibásak és zavaróak lehetnek a megjegyzések és az utalások. A LIBRARY SCHOOL STUDENTS*-re (könyvtároshallgatók) vonatkozó megjegyzésnek nincs sok értelme és redundáns is. A LIBRARY RELOCATION és a LIBRARY REMOVAL (a könyvtár áttelepítése és a könyvtár elköltöztetése) a ROMOVALS TO NEW BUILDINGS formát javasolja. (A REMOVAL TO NEW BUILDINGS, átköltöztetés új épületbe elírt változata.) Az elírások nem szépek a cikkekben, de egy tezaurusban bűnnek tekintendők.

* Az egyértelműség kedvéért az utalások leírásai nagybetűvel szerepelnek. A mintarekordokban néha kisbetűvel, néha nagybetűvel írják őket.

<i>Lib Online User Manual: Thesaurus</i>	
AREA MATERIALS (post-1978) and also under specific countries.	SEE LIBRARY MATERIALS
LIBRARY OF CONGRESS:CLASSIFICATION USE LIBRARY OF CONGRESS (Since references are indexed under both this term and: LIBRARY OF CONGRESS CLASSIFICATION For best results a search on both terms is advised)	LIBRARY SCHOOL LECTURERS SEE LECTURERS
LIBRARY OF CONGRESS:SUBJECT HEADINGS USE LIBRARY OF CONGRESS (combined with) SUBJECT HEADING SCHEMES	LIBRARY SCHOOL STUDENTS (Occasionally and incorrectly used instead of: LIBRARY STUDENTS)
LIBRARY ORGANISATIONS SEE LIBRARY ASSOCIATIONS	LIBRARY SCHOOLS UF SCHOOLS OF LIBRARIANSHIP GRADUATE LIBRARY SCHOOLS RT EDUCATION (PROFESSIONAL) LIBRARIANSHIP (pre-1978) PROFESSIONAL EDUCATION (post-1977)
LIBRARY ORIENTATION SEE USE INSTRUCTIONS	LIBRARY SCHOOLS:STUDENTS SEE LIBRARY STUDENTS
LIBRARY PERIODICALS UF LIBRARY PERIODICALS LIBRARIANSHIP:LIBRARY PERIODICALS RT ARCHIVES PERIODICALS PUBLISHED MATERIALS (Occasionally and incorrectly as: LIBRARY PERIODICALS)	LIBRARY SCIENCE SEE LIBRARIANSHIP
LIBRARY PROGRAMMES (Since references are indexed under both this term and: EXTENSION WORK For best results a search on both terms is advised)	LIBRARY SKILLS
LIBRARY PROJECTS	LIBRARY STAFF SEE STAFF (post-1977) STAFF (LIBRARY STAFF) (pre-1978)
LIBRARY PUBLICITY SEE PUBLICITY (post-1977) PUBLICITY (LIBRARY PUBLICITY) (pre-1978)	LIBRARY STANDARDS SEE STANDARDS (post-1977) STANDARDS (LIBRARY STANDARDS) (pre-1978)
LIBRARY RELOCATION SEE ROMOVALS TO NEW BUILDINGS	LIBRARY STATISTICS SEE STATISTICS (post-1977) STATISTICS (LIBRARY STATISTICS) (pre-1978)
LIBRARY REMOVALS SEE ROMOVALS TO NEW BUILDINGS	LIBRARY STOCK SEE LIBRARY MATERIALS STOCK (post-1977) STOCK (LIBRARY STOCK) (pre-1978)
LIBRARY RESEARCH SEE RESEARCH (post-1977) RESEARCH (LIBRARY RESEARCH) (pre-1978)	LIBRARY STUDENTS UF LIBRARY SCHOOLS:STUDENTS (Occasionally and incorrectly as: LIBRARY SCHOOL STUDENTS) (Since references are indexed under this term and: STUDENTS (post-1977) STUDENTS (ACADEMIC STUDENTS) (pre-1978) STUDENTS (LIBRARY STUDENTS) (pre-1978) For best results a search on all terms is advised)
LIBRARY RESOURCE CENTRES SEE RESOURCE CENTRES (post-1977) SCHOOL LIBRARIES (pre-1978)	LIBRARY SUPPLIERS (Since references are indexed under both this term and:
LIBRARY RESOURCES	

10.1. ábra: Zavaró és hibás hivatkozások a LIBRARY SCHOOL STUDENTS és LIBRARY RELOCATION alatt

Vannak UF (used for) (használguk helyette) utalások a SEAS-től (tengerek) az OCEANOGRAPHY-hoz (oceanográfia), de az OCEANS (óceánok) kifejezés csak magányosan lebeg. Az OLDER WOMEN (idősebb nők) olyan kifejezés, amely összesen egyszer fordul elő az egész adatbázisban mégis, van utalás róla a WOMEN (nők) szóra a kontrollált listán, az OLD PEOPLE (idős emberek) kifejezéshez azonban nem. Az is furcsa, hogy nem található meg az OLDER MEN (idősebb férfiak) kifejezés.

Az ember nem tud magához térni a csodálkozástól, mitől lesznek az olyan szavak, mint az „oceans” vagy „oil shales (palaolaj)” „a legfontosabb és legszélesebb körben használt” kifejezésekké egy könyvtár- és információtudományi adatbázisban. Az első 14 rekordban szerepel, a második egyáltalán nem található az egész adatbázisban. Nehéz elhinni, hogy a brit könyvtáros egyesület hogy tudott meglelni olyan hosszú ideig ezzel a rosszul szerkesztett kontrollált szótárral.

Az ISA esetében az a meglepő, hogy az IFI/Plenum továbbra is a kifejezéseknek azt a rosszul összeállított, bármilyen tárgyi megjegyzés és utaló nélküli listáját használta, amikor az ASIS jól felépített információtudományi teaurusza természetes választás lehetett volna egy ilyen adatbázis számára, miután első kiadása 1994-ben megjelent.

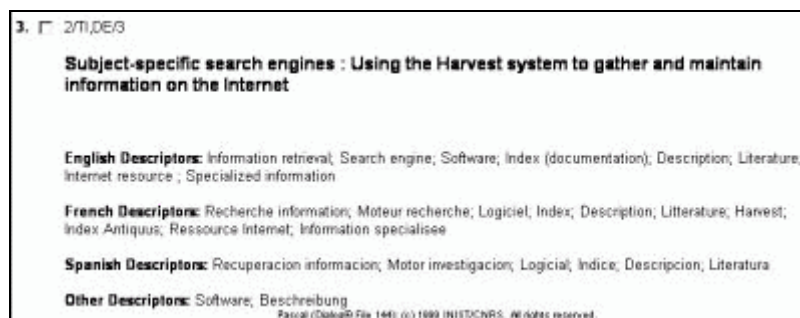
Egy másik probléma az, ha hiányzik a megfelelő kifejezés, ez a nyomtatott teaurusz átnézésével válhat egyértelművé. Az ERIC teauruszában sok nagyon specifikus és jól kiválasztott könyvtár- és információtudományi kifejezés található, de sem a legáltalánosabban használt kifejezés, a „CD-ROMs”, sem annak teljes formája nincs benne. A legközelebbi kifejezés az OPTICAL DATA DISKS (optikai adatlemezek). Nem megfelelő ennek a kifejezésnek a megadása, amikor egy fajt, nem pedig egy nemet tárgyalnak egy cikkben – túlságosan tág kifejezés. (Az ERIC indexelői legalább használhatják az identifier (azonosító) mezőt olyan kifejezésre, amely nem került még be a teauruszba.)

Hasonló probléma az, amikor egy, a teaurusz számára kiválasztott kifejezést nem használják a szakirodalomban (azaz nincs szakirodalmi alapja). Például a Compendexben a CD-ROM helyett egy elég mesterséges kifejezést - DATA STORAGE DIGITAL -ROM - használtak az 1990-es évek közepéig. Sok adatbázis a Kongresszusi Könyvtár tárgyszórendszerének (Library of Congress Subject Headings (LCSH)) módosított változatát használja, ez elég elavult, amikor új terminológiáról van szó, ugyanakkor prekoordinált katalógusok számára tervezték, nem pedig online rendszerek számára, amelyek kiválóan alkalmasak a posztkoordinációra, és bátorítják is azt. Az H. W. Wilson és az IAC adatbázis-családok a prekoordinált indexkifejezésekre szolgáltatnak példát. Dicséretükre el lehet mondani, hogy mindkettő sokkal több újabb kifejezést vett fel, mint az LCSH.

A kontrollált szótárak speciális problémája látható a PASCAL adatbázisban, amely többnyelvű kontrollált szótárat ajánl, de többre vállalkozik, mint amit képes megvalósítani. Nyilvánvaló, hogy az angol, spanyol és német szótárak összeállítói nem puristák és fanatikusok, amikor a francián kívül az egyéb nyelvek megbecsüléséről van szó. Úgy tűnik, hogy a spanyol kifejezéseket inkább egy számítógép fordította, nem pedig egy kétnyelvű szakember. Sok olyan van köztük, amelyek nem szerepelnek a szakirodalomban, még akkor sem, ha a prepozíciókat meghagyták volna. Prepozíció nélkül (pl. DETECCION ERROR DETECCION DE ERROR vagy inkább többes számban DETECCION DE ERRORES helyett) ezek a kifejezések úgy hangzanak, mint a turisták egy tijuani bevásárló központban, s az érzékelés hibáját jelenti hibaészlelés helyett.

Valójában a PASCAL sok félremagyarázott deszkriptora nyer más jelentést. Vegyük például a 10.2. ábra rekordját, amelyben a MOTOR INVESTIGACION jelentése motor-vizsgálat - ezt a Szövetségi Repülési Testület végezné egy repülőgép-baleset után. Amit szerettek volna mondani, az: „keresőmotor”. A „search engine” a spanyol számítástechnikai magazinokban BUSCADOR. A szoftver spanyol kifejezésként PROGRAMA, nem pedig LOGICIAL, amely a francia „logiciel” „spanyolosított” változata.

A német deszkriptorok mellett még a spanyolok is jónak látszanak a PASCAL-ban. (Érdemes megjegyezni, hogy mindig kevesebb spanyol deszkriptor van, mint francia, németből pedig sosincs háromnál több.) A német kifejezések olyan általánosak, hogy alkalmatlanok a keresésre. BESCHREIBUNG („description, leírás”) a leggyakrabban használt, s én inkább megspórolom véleményem leírását olyan választásról, amikor ez az egyetlen deszkriptor.



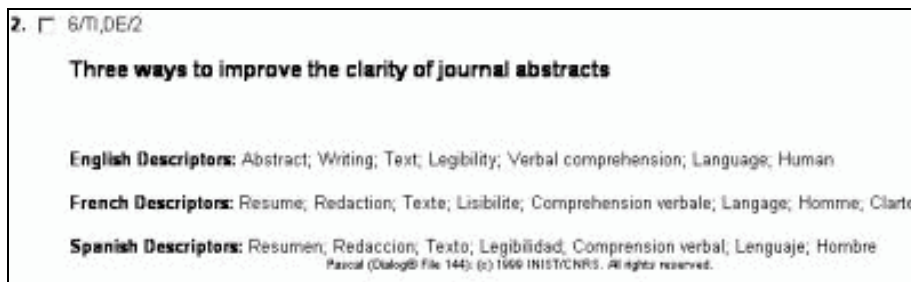
10.2. ábra: A szakirodalomban nem létező spanyol kifejezések

A legjobb tezausz sem segíthet, amikor az indexelő olyan kifejezéseket vesz fel, amelyek nem mentek keresztül minőségellenőrzésen s nem egyeztetik őket az érvényes kifejezések listájával. Az MHA mutatja a legszélsőségesebb példákat a tárgyi deszkriptorok, a legszen-
tebb tehenek teljes figyelmen kívül hagyására . még az elírásokat sem javítják (10.3. ábra).

Ref	Items	Index-term
E1	57	BEHAVIORAL SCIENCE
E2	1	BEHAVIORAL SCIENCE>
E3	35542	*BEHAVIORAL SCIENCES
E4	1	BEHAVIORAL SCIENCES CHILDREN AND YOUTH
E5	1	BEHAVIORAL SCIENCES INTELLIGENCE STUDIES
E6	1	BEHAVIORAL SCIENCES MENTAL OR EMOTIONAL DISOR
E7	1	BEHAVIORAL SCIENCES PSYCHOLOGY
E8	1	BEHAVIORAL SCIENCES PSYCHOSOMATIC MEDICINE
E9	1	BEHAVIORAL SCIENCES SUBCOMMITTEE RECOMMENDATI
E10	3	BEHAVIORAL SCIENCES.
E11	1	BEHAVIORAL SCIENCES. DRUG DEPENDENCE AND ABUS
E12	2	BEHAVIORAL SCIENCES
E13	1	BEHAVIORAL SCIENCES
E14	2	BEHAVIORAL SCIENCES
E15	1	BEHAVIORAL SCIENCES
E16	2	BEHAVIORAL SCINCES
E17	1	BEHAVIORAL SEINCNES

10.3. ábra: Rengeteg kirívó elírás az MHA deszkriptor mezőjében

Helytelen kifejezés kiválasztása a tezausból szintén általános probléma. Ez gyakran vitatható persze, de olyan esetekben azonban nem, amikor a kiválasztott deszkriptor egyértelműen rossz kifejezés. Az angol kifejezések még viszonylag jobban járnak, de ott is léteznek a nyelvészeti szempontból korrekt, de szemantikailag nem megfelelő kifejezések. Meglepő, amikor erre olyan cikkek esetében kerül sor, ahol az eredeti cím egyértelműen sugallja a megfelelő kifejezést, amely történetesen egy érvényes deszkriptor. Például a PASCAL a LEGIBILITY (olvashatóság) kifejezést használja, amikor a READABILITY-t (olvasmányosság) kellene használni olvashatóság értelemben. A 10.4. ábra világosan illusztrálja, hogy a cikk azt tárgyalja, milyen könnyen lehet egy dokumentumot megérteni (olvasmányosság), nem pedig az, hogy milyen nehéz olvasni (olvashatóság) a nyomtatás rossz minősége vagy a patikust is elpirulásra készítő kézírás miatt. Az olvasmányosságról szóló rekordok többségében rossz deszkriptort adtak meg a PASCAL-ban.



A cikk címének fordítása: Három mód, amellyel növelni lehet a folyóiratok referátumainak érthetőségét

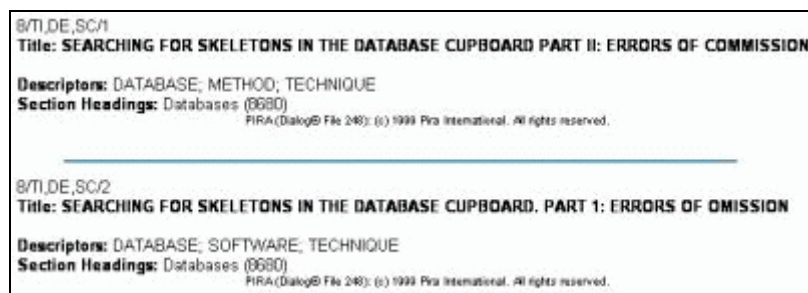
10.4. ábra: Rosszul megválasztott angol deszkriptor a PASCAL-ban

Esettanulmány a tárgyi indexelés minőségének értékeléséről

Ugyanazon cikkek különböző adatbázisokban készült rekordjainak halmazán keresztül fogjuk illusztrálni a tárgyi indexelés minőségértékelése négy legfontosabb mérőszámát. Olyan cikkeket választottunk ki, amelyek kapcsolódnak az online vagy a CD-ROM technikához. Egyeseket azért választottuk ki, mert jó alapot jelentenek a teszteléshez, a tárgyi indexelés következetességének illusztrálására a következő okok valamelyike alapján: Vagy kétrészes cikkek ugyanazt a témát (bár alkalmanként különböző szemszögből) feldolgozó rekordjai, vagy olyan cikkek rekordjai, amelyek először angolul jelentek meg, majd azután magyarra is lefordították őket.

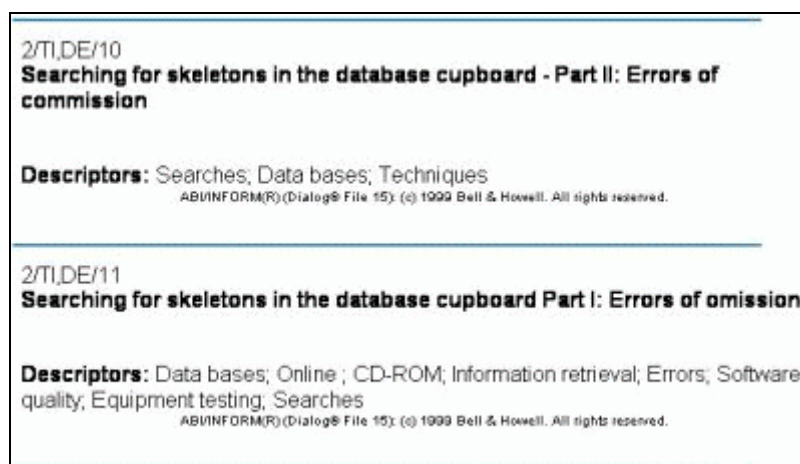
Sokkal több tesztet kellene elvégezni ahhoz, hogy következtetést vonhassunk le az indexelés minőségéről. A hely korlátozza az itt szereplő példák számát. A tárgyi indexelés minőségének tesztelése azt kívánja meg, hogy ismerjük a szakterületet és a primer dokumentumokat, így az értékelőknek olyan dokumentumok rekordjait kell kiválasztaniuk, amelyeket jól ismernek. Ez megbocsáthatóvá teszi olyan cikkek rekordjainak szerepeltetését, amelyeket a könyv szerzője írt. Egyes cikkeket a 11., a referátumok minőségét vizsgáló fejezetben is felhasználunk.

A specifikusság alapvető indexelési alapelv, vagyis az, hogy a kontrollált szótárból a fogalomra vonatkozó szakkifejezések közül a legspecifikusabbat adjuk. Ezt nehéz megtenni olyankor, amikor a teauruszban szereplő kifejezések nagyon általánosak, s nincsenek útmutatások azzal kapcsolatban, hogyan adjunk altárgyszavakat vagy alosztályokat, ha használatuk egyáltalán megengedett. Ez nyilvánvaló probléma a PIRA adatbázis esetében, amely mindenképpen elkerüli az összetett kifejezések használatát, s nem alkalmaz alosztályokat. A jellemző eset a „Searching for Skeletons” cikk mindkét részének indexelése (10.5. ábra). Az egyik a DATABASES, SOFTWARE és TECHNIQUE kifejezéseket használja, a másikban a DATABASE, METHOD és TECHNIQUE szerepel. A kifejezések a többes számot kívánnak meg, továbbá jelzőket, mint „search techniques” (keresőkifejezések) és „evaluation methods” (értékelési módszerek), ha már alosztályokat nem használnak.



10.5. ábra: A „Searching for Skeletons” cikk rekordjai a PIRA adatbázisban

Azt gondolhatnánk, hogy az ABI/INFORM ugyanazon a cikkpár esetében ugyanazokért a hiányosságokért hibáztatható, de a helyzet nem ez. Az ABI/INFORM olyan generikus fogalmakat használ, mint a TECHNIQUES (helyesen többes számban), de ezt egy fontos kifejezéssel (SEARCHES, keresések) együtt használják. Meg kell jegyezni azonban, hogy a DATA BASES kifejezés figyelmen kívül hagyja azt, hogy a szakirodalom sokkal gyakrabban használja a DATABASES formát, ahogy azt egy, az ABI/INFORM-ban végzett keresés is egyértelműen kimutatja. 101918 olyan teljes szövegű rekord van, amelyben ezt az egybeírt kifejezést használják, s csak 30895 olyan, amelyben a az ABI/INFORM-ban deszkriptorként használt két szóból álló forma (azaz data bases) jelenik meg a teljes szövegben.



10.6. ábra: A „Searching for Skeletons” cikk rekordjai az ABI/INFORM adatbázisban

A „Data Transfer” cikk 2. részének rekordjában a PROBLEMS és FUNCTIONS kifejezések nagyon általánosak, de ezeket számos specifikus kifejezéssel együtt sorolják fel, így a szövegösszefüggés világosabbá teheti az értelmüket (10.7. ábra). A furcsa deszkriptorok, a MANYPRODUCTS és a MANYCOMPANIES nem elírások. Olyan cikkekre vonatkozóan használják őket, amelyek sok termékkel és sok céggel foglalkoznak, s hasznosak lehetnek azok számára, akik mondjuk olyan cikkeket keresnek, amely sok nyomtatót értékel, vagy sok olyan cégről szól, amely például a 2000-es év problémájával (Y2K) foglalkozik.

10/TI,DE,CC/1
Data Transfer Capabilities of CD-ROM Software, Part II
Descriptors: CD-ROM; Data bases; Software; ASCII; Data transmission; Standards; Manyproducts; Comparative analysis
Classification Codes: 5240 (CN=Software & systems)
ABI/INFORM(R) (Dialog® File 15); (c) 1999 Bell & Howell. All rights reserved.
10/TI,DE,CC/2
Data Transfer Capabilities of CD-ROM Software: Part I
Descriptors: Software; CD-ROM; Problems; Functions; Output; Information storage; Manycompanies; Data bases
Classification Codes: 5240 (CN=Software & systems); 9190 (CN=United States)
ABI/INFORM(R) (Dialog® File 15); (c) 1999 Bell & Howell. All rights reserved.

10.7. ábra: A „Data Transfer” cikk rekordjai az ABI/INFORM adatbázisban

Még akkor is hiányozhat a szükséges specifikusság, ha összetett kifejezéseket adnak meg. Például az egyik legjobb minőségű adatbázis, az INSPEC rosszul szerepelt ebben a tesztben. A „Searching for Skeletons” mindkét rekordjához (10.8. ábra) általános és redundáns kifejezéseket adtak meg, hiányzik a legfontosabb, amely a minőség-ellenőrzés fogalmához kapcsolódik, bár erre tökéletes deskriptor található az INSPEC tezaurszában: quality control.

11/TI,DE,CC/2
Title: Searching for skeletons in the database cupboard. II. Errors of commission
Descriptors: information retrieval; information retrieval systems; information services
Class Codes: C7210 (Information services and centres); C7250 (Information storage and retrieval)
INSPEC (Dialog® File 2); (c) 1999 Institution of Electrical Engineers. All rights reserved.
11/TI,DE,CC/3
Title: Searching for skeletons in the database cupboard. I. Errors of omission
Descriptors: information retrieval; information retrieval systems; information services
Class Codes: C7250 (Information storage and retrieval); C7210 (Information services and centres)
INSPEC (Dialog® File 2); (c) 1999 Institution of Electrical Engineers. All rights reserved.

10.8. ábra: A „Searching for Skeletons” cikkek rekordjai az INSPEC-ben

A LISA által az átdolgozott kontrollált szótárból megadott két egyszavas kifejezés sokkal jobb ugyanezekre a rekordokra (ERRORS és DATABASES; hibák és adatbázisok és QUALIY és DATABASES; minőség és adatbázisok) (10.9. ábra).

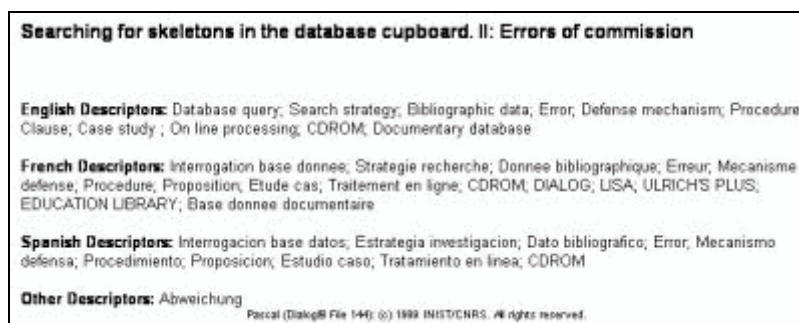
12/TI,BL,DE/2
Searching for skeletons in the database cupboard part II: errors of commission.
BLDSC Shelf Mark: 3535.802200
Descriptors: Databases; Quality
USA(LIBRARY&INFOSCI) (Dialog® File 61); (c) 1999 Reed Reference Publishing. All rights reserved.
12/TI,BL,DE/3
Searching for skeletons in the database cupboard Part 1: errors of omission.
BLDSC Shelf Mark: 3535.802200
Descriptors: Databases; Errors
USA(LIBRARY&INFOSCI) (Dialog® File 61); (c) 1999 Reed Reference Publishing. All rights reserved.

10.9. ábra: A „Searching for Skeletons” cikkek rekordjai az LISA-ban

A PASCAL-ban a francia, angol és - a mi példánkban - még a spanyol deskriptorok is megfelelőek. Néhányuk nagyon jó, mint a SEARCH STRATEGY (keresési stratégia) vagy a DATABASE QUERY (adatbázis lekérdezés). A német deskriptorok olyan általánosak szinte minden rekordban, hogy már haszontalanok. Még a generikus kifejezések közül is túl keveset

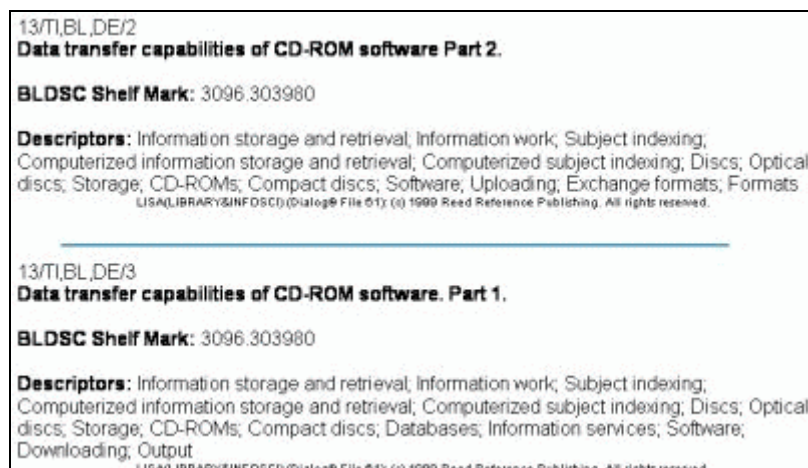
nagyon véletlenszerűen adnak meg, így nem adnak semmi kulcsot a cikkek tartalmáról. A 10.10. ábrán mutatott, a „Searching for Skeletons” 2. részére vonatkozó rekord német deszkriptorának, az ABWEICHUNG-nak, amelynek a jelentése DEVIATION, DISCREPANCY vagy ABNORMALITY (eltérés, az egyezés hiánya, rendellenesség) önmagában nincs értelme. Amikor egy gyors keresést végzünk arra vonatkozóan, hogy mennyire megfelelő a kifejezések kiválasztása, meg kell fontolni annak valószínűségét, hogy használnánk-e a kifejezés(ek)e)t a keresésben. Az ABWEICHUNG-nak önmagában nincs esélye erre.

Az indexelés kimerítő volta azt jelenti, hogy annyi releváns index kifejezést kell megadni, amennyit a cikk indokoltta tesz. Túl kevés index kifejezés megadása jelentős mértékben csökkenti a keresés teljességét. A túl sok hamis találatokhoz vezethet, bár az index kifejezések jó posztkoordinációja a keresőkérdésben csökkentheti ezek számát. Emiatt biztonságosabb az indexelő számára, ha inkább azzal hibázik, hogy több index kifejezést ad meg, mint kevesebbet, bár nem olyan mértékben, mint ahogy a LISA teszi.



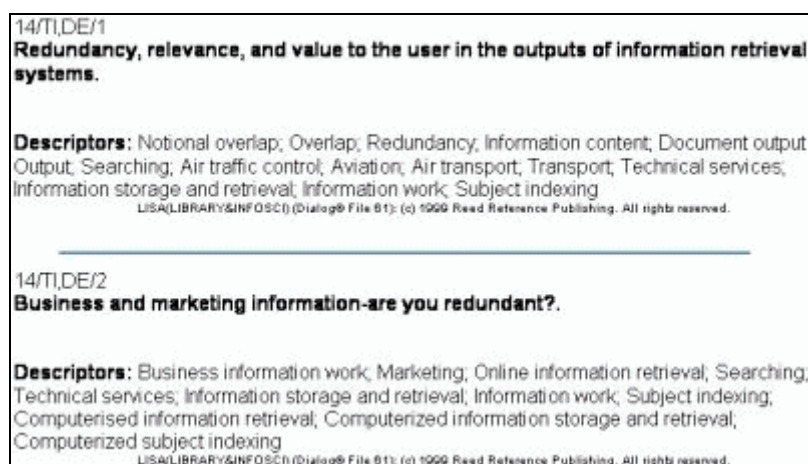
10.10. ábra: A „Searching for Skeletons” cikk 2. részének rekordja a PASCAL-ban

A LISA által a „Data Transfer” cikkhez megadott kifejezéseket a 10.11. ábra mutatja. Az INFORMATION STORAGE AND RETRIEVAL (információtárolás és -keresés), COMPUTERIZED INFORMATION STORAGE AND RETRIEVAL (számítógépesített információtárolás és -keresés), INFORMATION RETRIEVAL (információkeresés) nyilvánvalóan redundánsak és általánosak, csakúgy mint a másik fogalomra vonatkozó STORAGE MEDIA (tároló médium), OPTICAL DISCS (optikai lemezek), DISCS (lemezek) és COMPACT DISCS (kompakt lemezek), amikor a „CD-ROM” elegendő volna. A COMPUTERIZED SUBJECT INDEXING (számítógépesített tárgyi indexelés), SUBJECT INDEXING (tárgyi indexelés) és újra a SUBJECT INDEXING (tárgyi indexelés) nem csak abszurd módon redundánsak, hanem irrelevánsak is, mert a cikk sehol sem említi a tárgyi indexelést. A három legjobb, legspecifikusabb fogalom az UPLOADING (feltöltés), DOWNLOADING (letöltés) és az EXCHANGE FORMATS (csereformátumok), elegendőek is lettek volna a CD-ROM kifejezés mellett.



10.11. ábra: A „Data Transfer” cikkek rekordjai a LISA-ban

Ugyanannak a deskriptornak az ismétlődő használata olyan szakszerűtlen, így annak, hogy rekordok ezrei mutatták ezt a szimptómát, az egyedüli magyarázata az lehet, hogy egy program adta meg a deskriptorokat, amelyre senki nem figyelt. A 10.12. ábra rekordjai magukért beszélnek. Az illusztrációként kiválasztott rekordok eredeti cikkeinek témája nem véletlenül - mi más is lehetne, a redundancia. Egy indexelő ember aligha lett volna képes olyan sok redundáns kifejezést hozzárendelni a redundanciával foglalkozó cikkekhez.



10.12. ábra: A LISA rekordjai redundanciával foglalkozó cikkekhez

A specifikusságot sokszor úgy lehet elérni, hogy egy általánosabb kifejezést alosztályokra osztunk, hogy jelezzük a főtárgyszó bizonyos aspektusait. A Trade Industry Database (T&I) jó példát mutat erre a „Negotiating” cikk rekordjában (10.13. ábra) azzal, hogy a LIBRARIES-AUTOMATION (könyvtárak - gépesítés) és a CD-ROM--INSTALLATION (CD-ROM - telepítés) kifejezéseket adják meg.



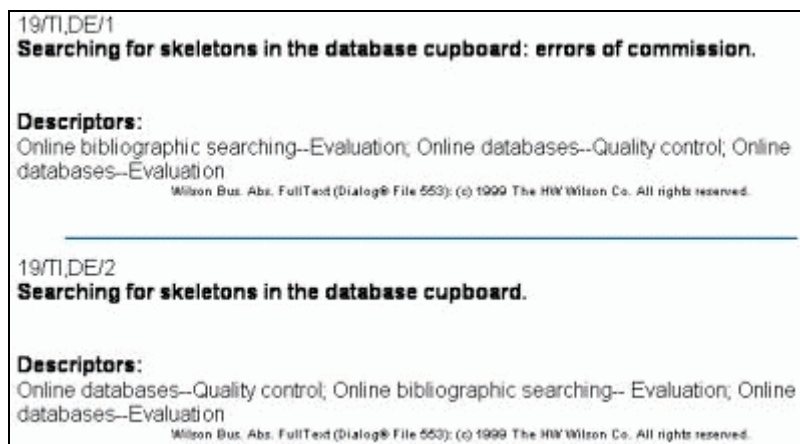
10.13. ábra: A T&I rekordja a „Negotiating” cikkhez

Ugyanez történik a „Quality of Abstracts” rekordjában (nem szerepel itt a példája), ahol mindkét főtárgyszóhoz tökéletes altárgyszó kapcsolódik: a QUALITY CONTROL (minőség-ellenőrzés) a fogalmakat szükséges mértékben specifikussá teszi. Furcsa, hogy a „Searching for Skeletons” cikk 1. részéhez kapcsolódóan az indexelő jó altárgyszót adott meg, a QUALITY CONTROL-t, a 2. résznél azonban kevésbé megfelelő a főtárgyszó-altárgyszó kombináció: DATA BASES-USAGE (adatbázisok - használat) (10.14. ábra). Ugyanaz vonatkozik a DATA BASES formátum használatára DATABASES helyett, amit korábban már az ABI/INFORM-nál megtárgyaltunk. Meg kell említeni a rekordokban a különbségeket a cím és az alcím sorrendjében, s ehhez hasonlóan a részek meghatározásának variációit is.



10.14. ábra: A „Searching for Skeletons” cikkek rekordjai a T&I adatbázisban

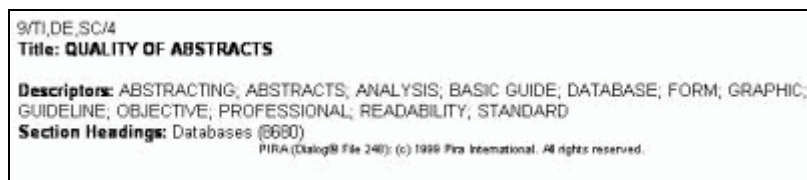
Ez a fajta alosztály az H. W. Wilson adatbázisok jellemzője, amelyek jók a specifikusságban, de ugyanakkor nem eléggé részletezőek. Az ilyen kérdéseket mindig az egész rekord tartalmával összefüggésben kell vizsgálni. Az H. W. Wilson cég jól ismert arról, hogy nagyon jó címkiegészítést használnak, amely épp annyira jó az online kereséshez, mint a további deskriptorok. Az H. W. Wilson adatbázisaiból vett mintarekordok között a Business Abstracts tökéletes egyensúlyt tart a túl mértékletes és a túl kimerítő indexelés között (10.15. ábra). Ez illusztrálja a hagyományos prekoordinált indexeléshez való erős ragaszkodást.



10.15. ábra: A „Searching for Skeletons” cikkek rekordjai a Business Abstractsben.

Az egy cikkhez hozzárendelt index kifejezések nagy száma nem feltétlenül biztosítja azt, hogy az megfelelően kimerítő legyen, ahogy azt világosan mutatta a LISA szokatlanul redundáns indexelési gyakorlatának példája. A PIRA ad „jó” példát arra, hogy ne csak a rekordokhoz hozzárendelt deskriptorok számát nézzük, amikor a tárgyi indexelés minőségét megítéljük. A „Quality of Abstracts” cikk meghintését 12 indexfogalommal (egyetlen szó kivétellel) nem lehet kritika nélkül hagyni (10.16. ábra).

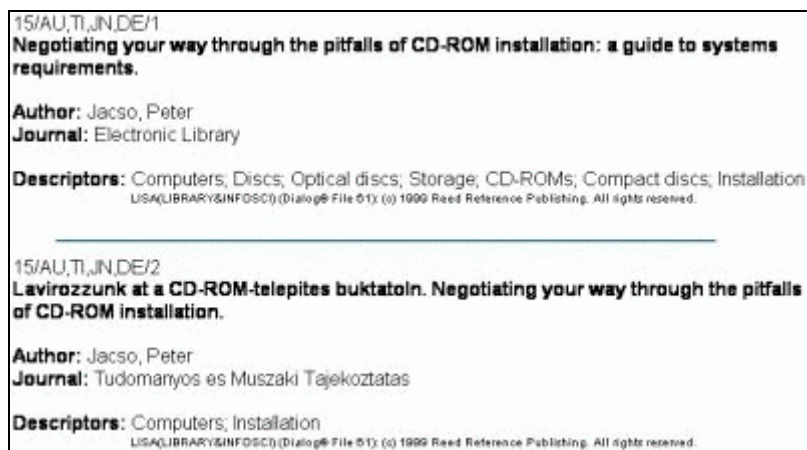
Úgy tűnik, hogy a PIRA rendkívüli mértékben kedveli egyszavas kifejezések megadását tárgyi deszkriptorként. A FORM (forma), GRAPHIC (grafikus) és a PROFESSIONAL (szakmai) kifejezések magukban sokkal inkább látszanak KWOC (keyword out of context, kulcsszó a szöveggörnyezeten kívül) indexfogalmaknak, mint deszkriptoroknak, és az egyes számú formák használata a legtöbb főnév esetében (kivéve az ABSTRACTS (referátumok) formát) nem általános gyakorlat. Az OBJECTIVE (cél) deszkriptor esetében ez nagyon zavaróvá válik (mert semmi nem különbözteti meg az objektív szótól, mint jelzőtől).



10.16. ábra: A „Quality of Abstracts” cikk rekordja a PIRA adatbázisban

Az ideális jelöltek a következetesség ellenőrzésére többrészes cikkek, konferencia-előadások, amelyeket később folyóiratcikké is átdolgoztak, valamint eredeti cikkek lefordított változatai.

Nyilvánvaló indexelési következetlenség figyelhető meg a „Negotiating” cikk eredeti angol nyelvű változatának és magyar fordításának LISA rekordjai között (10.17. ábra). Az eredeti cikk rekordjához hét, nagyrészt feleslegesen széles kifejezést adtak a CD-ROM-hoz, de legalább tartalmazza a legspecifikusabb kifejezéseket is. A magyar fordítás rekordja kihagyja a CD-ROM-ra vonatkozó deszkriptorokat, s csak két általános és felesleges kifejezést ad meg.



10.17. ábra: Egy cikk eredetijének és magyar fordításának rekordjai a LISA-ban

Az INSPEC esetében az angol nyelvű cikkhez megadott két deszkriptor tökéletes (CD_ROMs és INSTALLATION, CD-ROM-ok és telepítés). A „Negotiating cikk” magyar fordítása deszkriptorainak relevanciája megkérdőjelezhető (DP MANAGEMENT és INFORMATION SERVICES (az adatfeldolgozás menedzsmentje és információs szolgáltatások), mert túl általánosak. A szerző keresztnévének hibás kezdőbetűjét is szóvá kell tenni, bár ennek nincs hatása a tárgyi indexelésre (10.18. ábra).

16/AU, TI, JN, DE/1
Title: Negotiating your way through the pitfalls of CD-ROM installation
Author Jacso, J.
Journal: Tudományos és Műszaki Tájékoztatás
Descriptors: CD-ROMs, DP management, information services
<small>INSPEC (Dialog® File 4): (c) 1999 Institution of Electrical Engineers. All rights reserved.</small>
16/AU, TI, JN, DE/2
Title: Negotiating your way through the pitfalls of CD-ROM installation
Author Jacso, P.
Journal: Electronic Library
Descriptors: CD-ROMs, installation
<small>INSPEC (Dialog® File 4): (c) 1999 Institution of Electrical Engineers. All rights reserved.</small>

10.18. ábra: A „Negotiating” cikk rekordjai az INSPEC-ben.

Az ISA termékeny talajt biztosít a tárgyi indexelés következetességének ellenőrzéséhez, mert több ezer valódi duplikált rekordpár található benne, ahogy az részletesen tárgyalta Jacsó (1998a). Legtöbbjük érdekes ciklikus mintát mutat (10.19. ábra). A duplikátumokat általában három-öt évvel az első után vették fel. Ez a gyakorlat az 1990-es évek elején és közepén volt a legelterjedtebb, amikor az ISA-hoz adott rekordok száma minden aktualizáláskor irigylésre méltóan kerek volt (10.20. ábra), bár ez nem mindig érte el a havi 800-900 rekordnyi mennyiséget, amit az adatbázis korábbi készítője, az IFI/Plenum ígért.

23/BN, AU, TI, DE/1
ISA Document Number in Printed Publication: 9706482
Annual review of database developments: 1992.
Author (Affiliation): Basch, R. (Aubergine Information Services, Berkeley, CA)
Descriptors: ANNUAL REPORTS; BUSINESS INFORMATION; DATABASES; INFORMATION INDUSTRY; MARKETING; ONLINE SYSTEMS; POLITICAL ASPECTS; SOCIOECONOMIC ASPECTS
<small>Information Science Abs. (Dialog® File 202): (c) Information Today, Inc. All rights reserved.</small>
23/BN, AU, TI, DE/2
ISA Document Number in Printed Publication: 9207336
Annual review of database development, 1992.
Author (Affiliation): Basch, R. (Aubergine Information Services, Berkeley, CA)
Descriptors: ANNUAL REPORTS; DATABASES; DIALOG; EUROPE; INTERNATIONAL; NEXIS DATABASE; TELEVISION
<small>Information Science Abs. (Dialog® File 202): (c) Information Today, Inc. All rights reserved.</small>

10.19. ábra: A duplikátumok ciklikus mintája az ISA adatbázisban

A duplikált rekord párok deskriptorainak elemzése kisebb mértékű következetességet mutat, mint amit korábban Sievert és Andrews (1991) tapasztalt. Ez érvényes a széles tárgyi kategóriákra is, ahogy azt a mintarekordok illusztrálják. Figyelembe véve, hogy tárgyi kategóriáknak és alkategóriáknak csak 170 kódja van, s nem változtak abban az időszakban, amikor a legnagyobb mértékű volt a duplikátumok hozzáadása, szokatlanul magas szintűnek látszik a következetlenség. Ez egyrészt megmagyarázhatja azt, hogy az, aki hosszú időn keresztül az ISA szerkesztője volt, miért nem tudott a duplikátum tételek mennyiségéről és szívésságáról, s ez arra ösztönözte, hogy kijelentse, 1986-tól nem adtak duplikátum tételeket az adatbázishoz (Allcock 1997).

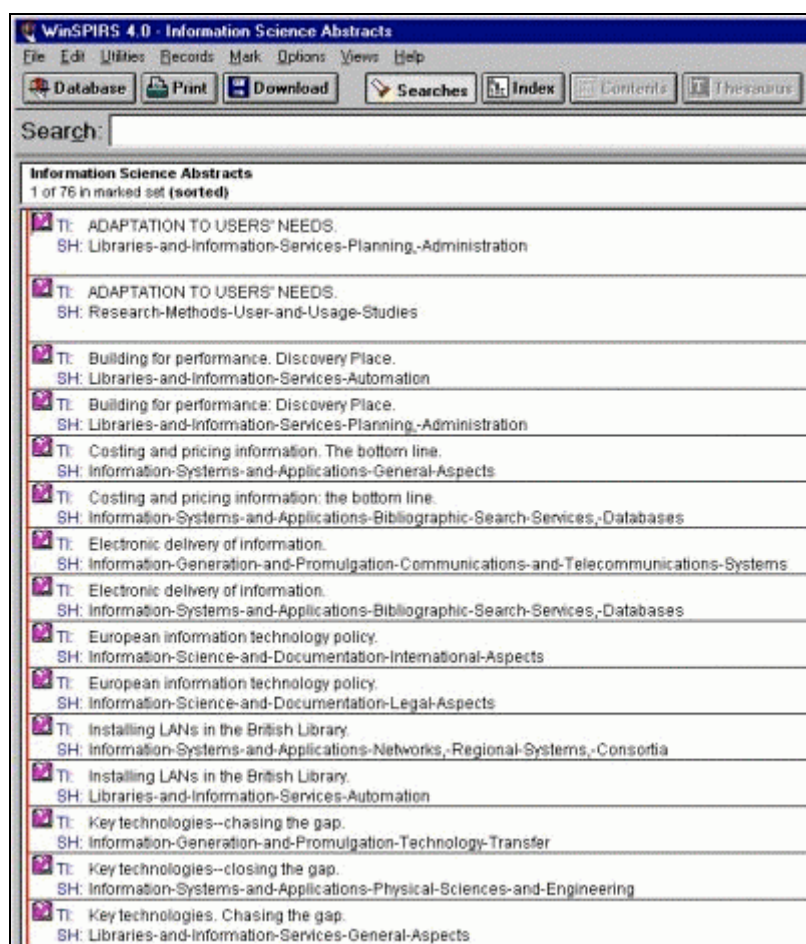
ID	Year	Count	ID
E13	202	700	UD=9501
E14	202	700	UD=9502
E15	202	700	UD=9503
E16	202	700	UD=9504
E17	202	700	UD=9505
E18	202	800	UD=9506
E19	202	850	UD=9508
E20	202	850	UD=9509
E21	202	800	UD=9510
E22	202	850	UD=9511
E23	202	850	UD=9512
E24	202	700	UD=9601
E25	202	700	UD=9602

10.20. ábra: Az aktualizálás rekordjainak tökéletesen kerek száma az ISA adatbázisban

Valóban, a duplikátumok különböző részekben jelentek meg (azaz nem libasorban jöttek ugyanabban a kategóriában), s ez nem könnyítette meg az olvasók és a szerkesztő számára, hogy észrevegyék ezt a problémát a nyomtatott változatban, amely a tételeket széles tárgyi kategóriák alá rendezi. A probléma másik része az, hogy az előfizetők sem panaszkodtak a duplikátumok miatt, mivel leginkább a referáló és indexelő folyóirat meghatározott részeit nézik meg, s nem olvassák végig A-tól Z-ig.

Az adatbázis új vezetése megígérte, hogy az ISA „erőteljes tisztítási folyamaton” fog keresztül menni, s ebbe beletartozik a duplikátum rekordok eltávolítása is. Ezt az erőteljességet növelte az a tény, hogy az, aki az IFI/Plenum régebbi vezető technikai szerkesztője ISA ügyekben, csatlakozott az új tulajdonos szervezetéhez, s kétségtelenül hidat jelent a régi és új ISA között. Duplikátumok ezreinek törlése időigényesebb folyamat, mint az adatbázishoz való hozzáadásuk volt.

Az erőfeszítések ellenére még mindig vannak duplikátumok, s újak is felbukkannak. Bár ez nem jó az ügyfelek számára, jó a kutatóknak, akik tárgyi indexelésen túl vizsgálhatják az indexelés következetességének egyéb szempontjait is. Ugyanakkor azonban a kutatóknak óvatosnak kell lenniük az ISA-ból visszakeresett adatok bibliometriai magyarázatával, amikor a kiadási és tematikus trendeket vizsgálják, egészen addig, amíg az összes duplikátumot el nem távolítják.



10.21. ábra: A széles tárgyi kategóriák következtelen meghatározása az ISA duplikátumaiban.

11. A referátumok minősége

A referátum az eredeti cikk rövid (150-250 szavas) összefoglalását jelenti. Bár a cikkek referátumainak közzlése időben megelőzte a folyóiratokat indexelő szolgáltatásokat, az információs szolgáltatások automatizálása azzal kezdődött, hogy számítógépeket használtak indexelő adatbázisok előállításához. Igazi áttörés volt, amikor a folyóiratokat és konferenciakiadványokat indexelő szolgáltatások elkezdtek referátumot is tartalmazni. Az 1990-es évek végére a csak indexelő adatbázisok száma (ellentétben a nyomtatott indexekkel) zuhanni kezdett.

A British Education Index, az Anthropological Index, a National Newspaper Index és a Library Literature a néhány ritka kivétel közé tartozik az adatbázispiacra, és 1999 nyarán a Library Literature-t a cikkek teljes szövegével egészítették ki (de referátumokkal nem) csaknem 80 jelentős könyvtár- és információtudományi folyóiratról. Ezt a továbbfejlesztett változatot Library Literature & Information Science Full Text Database-nek nevezik. A Population Indexet épp 2000-től nem egészítik ki új rekordokkal, annak ellenére, hogy kiváló volt a minősége és a hírneve. A referátumok hiánya a csak indexelő adatbázisokat egyre veszélyeztetettebb fájrá teszi.

Az H. W. Wilson cég további klasszikus indexeit, mint például az Art Index, Humanities Index és az Applied Science & Technology Index, korábban kiegészítették referátumokkal (bár csak az elmúlt néhány évre, s nem egészükre vonatkozóan), valamint sok száz folyóirat esetében az eredeti cikkek teljes szövegével. Egyre inkább az várható, hogy a dokumentumok teljes szövegű vagy képi változata (vagy mindkettő) elérhető lesz az adatbázisokban.

Mindezen fejlemények ellenére a referátumoknak lesz helyük a nap alatt. A jó referátumok segíthetnek a használóknak annak kiválasztásában, hogy melyik dokumentumot olvassák el a maga teljességében. Egy referátum utalhat a forrásdokumentumok (folyóiratcikkek, szabadalmi leírások, kongresszusi meghallgatások, könyvek, filmek) tartalmára, vagy összefoglalhatja a tartalmat. Az első az indikatív referátum, az utóbbi az informatív, amely közvetlenül bemutatja a forrás specifikus információinak egy részét. Mindkét típusra található példa a fejezetben. A referátum harmadik típusa a kritikai referátum, amely az összefoglalást a referáló, a szakterület specialistája értékelő és kritikai kommentárjaival egészíti ki. A kritikai referátumokat elsősorban házon belüli felhasználásra készítették, újabban rendszerezett Webkalauzok számára. Értékelésüket nem tárgyalja a fejezet.

A jó referátumok kitűnő szűrők a keresés folyamatában. A keresés korlátozása olyan rekordokra, amelyekben szerepel a „Microsoft”, a „Department of Justice” és az „antitrust” kifejezés a referátumban, valószínűleg növeli a visszakeresett referátumok relevanciáját, kizárva azokat a tételeket, amelyekben csak a teljes szövegben szerepelnek ezek a szavak. Azok esetleg csak mellékesen említik a folyamatban levő pert, vagy tartalmazzák a három fogalmat úgy, hogy azok nincsenek egymással kapcsolatban.

A referáló és indexelő szolgáltatások online és CD-ROM-os változatai sokkal könnyebbé teszik a referátumok elemzését és összehasonlítását. Az adatbázisból meghatározott feltételeknek (mint a téma, típus, a forrásdokumentum nyelve) megfelelő referátumok kiválasztása, s az azoknak megfelelő rekordok megtalálása másik adatbázisból összehasonlító elemzés céljából többé nem időrabló művelet. Nagyszámú referátum nyelvtani és stilisztikai elemzése olvashatósági szintjük meghatározása érdekében sokkal kevésbé fárasztó, mint amikor manuális elemzést alkalmaztunk. A referátumok összehasonlítása a forrásdokumentumokkal azonban olyan terület marad a referátumok elemzésében, amely emberi képességeket, hozzáértést és időt igényel. Annak megállapítása, hogy egy referátum milyen jól foglalja össze a forrásdokumentum tartalmát, időigényes művelet marad.

Fogalmak és irodalom

Számos kiváló könyv tárgyalja hosszan azt, hogy mitől lesz jó egy referátum és milyen típusú referátumokat használnak különféle célokra (Borko és Bernier 1975; Cleveland és Cleveland 2001; Tenopir és Lundeen 1988; Lancaster 1998). Lancaster könyvének második kiadása a friss szakirodalomról is áttekintést nyújt.

A referátumok az American National Standards Institute (ANSI) és a National Information Standards Organisation (NISO) közös szabványának tárgyát képezik. Az eredeti, 1971-es szabványt két alkalommal átdolgozták. A legutóbbi kiadást, a Guidelines for Abstracts ANSI/NISO Z39. 14-1997-et 2000 novemberének elején ingyenesen elérhetővé tette a weben a NISO a http://www.techstreet.com/cgi-bin/detail?product_id=52600 címen.

Ez a szokatlanul tömör szabvány összefoglalja a korábban említett monográfiákban tárgyalt alapelveket. Egy jól használható függelék egészíti ki, amely a referátumok különböző típusaival illusztrálja az irányelveket. A függelék nem része magának a szabványnak, sem a bevezető, amely a referátumok szabványosításának hátterét mutatja be. Bár a próbálkozások dicséretre méltóak, elsősorban azért, mert különös figyelmet szentelnek annak, hogy a számítógépes keresés szempontjából kívánatos jellemzőkkel kellene rendelkezniük a referátumoknak, meg kell jegyezni, hogy nem sok referáló és indexelő szolgáltatás követi ezeket minden szempontból.

A referátumok minőségi szempontjairól könyvek és cikkek kiváló választéka található, a működő rendszerekhez kapcsolódó tanulmányból azonban kevés van, annak ellenére, hogy a kutatók egyetértenek abban, hogy a referátumok jelentős mértékben hozzájárulhatnak a visszakeresés eredményességéhez.

A referátumok fontosságával kapcsolatos konszenzus olyan mértékű, hogy a friss szakirodalomban összesen egy olyan tanulmányt lehetett találni, amely egy mellékes megjegyzésében utalt arra a tényre, hogy a használók talán nem értékelik a referátumok jelenlétét olyan sokra, mint ahogy azt az információs szakemberek feltételezik. Watson és Perrin (1994) összehasonlította a MEDLINE és a CINAHL által feltárt anyagot a közös egészségügyi területeken, és azt állapították meg felmérésükben, hogy az értékelésbe bevont 10 oktató közül 6 nem érezte úgy, hogy a referátumok hiánya akadály lenne.

A referátumok fontosságát abból a szempontból, hogy növelhetik a teljességet, Sievert és Sievert (1991) kísérlete illusztrálta, akik azt állapították meg, hogy a rekordoknak a 12%-a, amely visszakereshető volt a FRANCIS adatbázisban végzett tárgyi keresésekkel, a Philosopher's Indexből azonban nem, nem tartalmazott referátumot vagy a szerzői referátumból vett olyan szakkifejezést, amely megfelelt volna a keresőkérdésben szereplő szakkifejezésnek. Weston és Lauderdale (1988) a PsycINFO adatbázisban szereplő disszertációk referátumai hiányának tulajdonította az alacsony teljességi értéket e dokumentumok kulcsszó szerinti keresésekor.

Bár létezik konszenzus a referátumok potenciális hozzáadott értékével kapcsolatban, nincs egyetértés abban, hogy elkülönítsék-e az indexelés és a referálás folyamatát, és hogy hogyan ítélik meg a szerzők által írt referátumokat. Everett Brenner (1989) azt mondja, hogy a referálási és indexelési készségek jelentős mértékben különböznek egymástól, egy jó indexelő nem szükségszerűen jó referáló is, s lehet, hogy nem hatékony, ha valakinek minden egyes dokumentum feldolgozásakor át kell váltania gondolkodását az indexelés és a referálás között. Másik oldalról Lancaster (1991) amellett érvel, hogy a referálás csak egy kis lépésnyire van az indexeléstől, és ugyanazt a fogalmi elemzést követeli meg a dokumentum elolvasása és lényegének kiemelése után, így a folyamatot hatékonyabbá teszi, ha a két tevékenységet kombinálják.

A szerzők által írt referátumok kérdésében Froehlich (1994) azt bizonyítja, hogy míg a kemény tudományok területén lehetséges, hogy egy szerző által írt referátum fokozza értékét és hitelét, „más szakterületeken nem lehet ez a helyzet, mert a szerző csak artikulálja a témával kapcsolatos megértést, s ez a megértés olyan lehet, amely vagy megfelel a szakterület ortodox felfogásának vagy megértésének, vagy nem.” (A cikk iróniája, hogy a szerző által készített nagyon jó minőségű, informatív referátum ellentmond a szerzők által írt referátumokat ellenző véleményének.)

Másik oldalról Milas-Bracovic és Zajec (1989) - összehangban a referátumok írására vonatkozó ANSI szabvány 1979-es változatának kikötésével - amellett érvel, hogy a szerzők által írt referátumok időt takarítanak meg és növelik a referátumok minőségét „mivel a szerzőktől mint a szakterület specialistáitól - el lehet várni, hogy a lehető legjobb referálók legyenek, feltéve, ha megismertetik őket a referátumok írásának szabályaival.”

A referátumokkal kapcsolatban az elmúlt néhány évben a legelsőrendűbb kérdés a klinikai orvostudományi folyóiratokban használt strukturált referátumok használata volt. Míg a könyvtár- és információtudományi folyóiratok alig vettek tudomást ezekről a fejleményekről (Harbourt, Knecht és Humprey [1995] később tárgyalandó mindenre kiterjedő, statisztikailag alapos és információgazdag tanulmányának figyelmet érdemlő kivételével) az orvosi folyóiratok szerkesztői és közreműködői intenzíven tárgyalták a strukturált referátumok érdemeit vagy érdemeinek hiányát, ahogy arra Haynes (1996), valamint Johnson és Bevan (1996) hivatkozik. Az ő mellette és ellene szóló érvek bármely szakterület információs szakemberei számára nagyon relevánsak lehetnek, és némi adaptálással a strukturált referátumok más természettudományi és - kisebb mértékben - társadalomtudományi területek számára is alkalmazhatók lehetnek.

A strukturált referátumok növekvő számban jelennek meg orvostudományi és más élettudományi adatbázisokban. Ahogy Froom és Froom (1993) idézi, az American College of Physicians Ad Hoc Working Group for the Critical Appraisal of the Medical Literature (Az amerikai orvosok szövetségének ad hoc munkabizottsága az orvosi szakirodalom kritikus értékelésére) kifejlesztett (majd később módosított) egy terjedelmes útmutatót azoknak a referátumoknak a tartalmáról és formájáról, amelyeknek a klinikai vizsgálatokról szóló és szemlecikkekben kell megjeleníteniük. Az útmutató azt javasolja, hogy a maximum 250 szóból álló referátum összetevőit fel kell címkézni, meghatározva az eredeti kutatás tanulmányának célját, tervét, helyszínét, a betegeket, a kezelést, a kimenet mérőszámait, az eredményeket és következtetéseket. A szemlékkel kapcsolatban ettől némileg eltérő összetevőket javasolnak.

Dixon (1988) azok közé tartozott, akik dicsérték a strukturált referátumokat, míg Heller (1990) úgy érezte, hogy „ez csak egy újabb roham a kreatív orvosi kommunikáció ellen.” Az emocionálisan túlfűtött véleménycserén túl némi kvantitatív és kvalitatív értékelést is megjelentettek, és a folyóiratok állást foglaltak. Míg a *Journal of Clinical Epidemiology* visszaütötte a strukturált referátumok használatát, a legmagasabb impakt faktorú folyóiratok, mint a *New England Journal of Medicine*, a *Journal of the American Medical Association*, a *Chest* és a *Gut* némileg különböző változatokban alkalmazta a strukturált referátumokat.

Narine et al. (1991) lényeges hiányosságokat talált a *Canadian Medical Association Journal*, az útmutatások egyik korai alkalmazója által megjelentetett 33 referátum minőségében. Taddio et al. (1994) kiterjesztette Narine et al. (1991) vizsgálatait 150 nem strukturált és 150 strukturált referátumra, az előbbieket saját pontozásos modelljét alkalmazva, s azt találták, hogy a két rétegezett minta összesített minőségi középértéke 0,57 és 0,74 volt.

Froom és Froom (1993) 130, az *Annals of Internal Medicine* című folyóiratban, a strukturált referátumok fogalmának első alkalmazójában megjelent cikk referátumát és teljes szövegét vizsgálta. Megvizsgálták, hogy a strukturált referátumok mennyire feleltek meg az

útmutatónak, s mennyire segítették a klinikai szakembereket annak eldöntésében, vajon az információ mennyire érvényes az ő betegek populációjára. Azt találták, hogy a strukturált referátumok nagyon hiányosak a betegek kiválasztásának összetevőihöz és szerzők következtetéseihöz kapcsolódó információk bemutatásában.

Egy eltérő véleményt megfogalmazó, de kiegyensúlyozott cikkben Haynes (1993) kimutatta a felmérés korlátozott hatókörét és megjegyezte, hogy részletesebb vizsgálatokra lenne szükség. Bár bevallja, hogy „elszomorító, hogy olyan sok információ hiányzik” a strukturált referátumokból, arra figyelmeztet, hogy nem az a kulcskérdés, hogy a referátum mennyire felel meg az útmutatónak, hanem inkább az, hogy a jól strukturált referátumok mennyire felelnek meg céljaiknak. Ezek a célok: jobb kommunikáció az olvasókkal, tökéletesített lektorálás, hatékonyabb visszakeresés a bibliográfiai adatbázisokból.

Salager-Meyer (1991) 77 referátumot vizsgált, amelyek 33 biomedikai folyóiratban jelentek meg 1987 és 1989 között, s úgy találta, hogy 48%-ukban volt egy vagy több strukturális hiba, mint például egy fontos összetevő kihagyása, az összetevők illogikus sorrendje, például az, hogy a következtetések megelőzték az eredményeket.

Valószínűleg idáig Harbourt, Knecht és Humphreys (1995) tette a legfontosabb lépést abba az irányba, hogyan lehet értékelni a strukturált referátumok jellemzőit és hatását az online és CD-ROM-os visszakeresés szempontjából. A szerzők megvizsgálták a MEDLINE adatbázisba 1989 elejétől 1991 végéig bekerült csaknem egymillió rekord jellemzőit, köztük 3873 rekordot strukturált referátumokkal. Azt találták, hogy a strukturált referátumot tartalmazó rekordokban 3,5-szer annyi MeSH kifejezés volt, mint a teljes MEDLINE mintában, bár ez a növekedés kisebb volt akkor, amikor klinikai cikkekkel (2,1) és a szemlékkel (1,9) hasonlították össze őket. Ezen kívül a strukturált referátumok (a címkék nélkül) csaknem 70%-kal hosszabbak voltak, mint a teljes MEDLINE minta. A statisztikai adatok egyértelműen azt jelzik, hogy mind a klinikai folyóiratok száma, mind a strukturált referátumot tartalmazó cikkek száma folyamatosan növekszik, bár a strukturált referátumokat tartalmazó mintarekordok 79%-a a folyóiratok 17%-ából jött össze.

Kulkarni, Gupta és Viswanathan (1996) a strukturált referátumot tartalmazó MEDLINE rekordok növekedési ütemét vizsgálták meg, összehasonlítva az angol nyelvű klinikai folyóiratcikkek teljes számával, s azt állapították meg, hogy csak 28,5%-ukban volt strukturált referátum. Bár hosszú időbe telhet, amíg a strukturált referátumok kerülnek a klinikai orvostudományi cikkek esetében többségbe, hozzáférési pontok számának jelentős mértékben megnőtt száma valószínűleg segít a keresőknek mind a visszakeresés, mind a megjelenítés során.

Ezt a feltételezést Hartley, Sydes és Blurton (1996) tesztelte kis mintán, a *British Journal of Educational Psychology* 30 cikkéről készült egy-egy strukturált és nem strukturált referátumon. Ötvenkét egyetemistát kértek meg arra, hogy találjanak választ két kérdésre nyolc egyik fajta formában bemutatott referátum alapján. Ezt a másik formában megjelenő nyolc referátumhoz kapcsolódó két-két kérdés követte. Egy kiegészítő tesztben 56 diákot kértek meg arra, hogy találjanak öt referátumot, amely bizonyos fajta vizsgálatról adott jelentést, majd pedig öt további, amely egy másik fajta vizsgálatot mutatott be. Az időre és a hibákra vonatkozó adatok (amelyeket automatikusan rögzítettek) alátámasztották azt a hipotézist, hogy a használoknak könnyebb strukturált referátumokat használni, mint nem strukturáltakat.

Hartley, Sydes és Blurton (1996) tanulmányozták a strukturált referátumok tipográfiai aspektusait is, és hogy ezek milyen hatással vannak a használok preferenciáira. Vizsgálatok sorát végezték el három szakaszban összesen több mint 400 hallgatóval, hogy meghatározzák, az olvasóknak milyen preferenciái vannak a strukturált referátumok alcímének különböző tipográfiai elrendezéseivel kapcsolatban, valamint összességében a referátumok elhelyezé-

sével (európai) szabványoldalon (amely majdnem teljesen azonos az USA-ban használt 8,5x11 hüvelykes levélméret formátummal). A legkedveltebb változat az volt, ahol a kövér nagybetűkkel írt alcímeket üres sorral választották el minden egyes bekezdés fölött (bevezetés, cél, módszer stb.). A formátumra vonatkozóan a legkedveltebb formátum az volt, amikor az a kéthasábos cikk fölött középre volt helyezve.

Annak az egyik lehetséges oka, hogy az RNdex Top 100 sikeresen lépett be a zsúfolt piacra, az volt, hogy minden rekordjukhoz tartozik referátum, és ezeket a referátumokat a fájl előállítója, az American College of Physicians Ad Hoc Working Group for the Critical Appraisal of the Medical Literature ajánlásait követve készíti. Azt javasolják, hogy a cikkek közvetlen klinikai gyakorlattal rendelkező szerzői készítsék a referátumokat, s így a cél, módszerek és eredmények alapvető szempontjait következetesen, szabványosított formában írják le (Haynes 1987).

Annak érdekében, hogy a *Referativnii Zhurnal Khimii* (RZK), egy orosz kémiai referáló és indexelő folyóirat referátumainak minőségét megpróbálják megállapítani, Katritzky, Cato és Deyrup (1993) az RZK kísérleti angol változatából 923 referátumot vizsgált meg. Azután - egy orosz születésű vegyész segítségével - összehasonlítottak 102 véletlenszerűen kiválasztott angol nyelvű folyóiratcikkre készült referátumot, amelyek mind a Chemical Abstractsben (CA), mind az RZK-ban megtalálhatók voltak, hogy meghatározzák, vajon

egy vegyész könnyen tud-e információhoz jutni a referátumból,
a kutatás célját egyértelművé tették-e,
a cikk legfőbb fogalmai szerepelnek-e a referátumban,
a referátumban megtalálható-e a legfontosabb következtetés,
a kémiai szerkezetek szerepelnek-e benne,
voltak-e jelentős különbségek a referátumban az adatok mennyiségében,
voltak-e nyilvánvaló hibák, és
a referátum minősége kapcsolatban volt-e a cikk földrajzi származási helyével.

A kísérleti angol változatban volt néhány elírás és hiba, bár nem a szerző neve, a folyóirat címe, az oldalszám és a dátum mezőben, amelyet nagyon jó minőségűnek találtak. A 102 referátum mély vizsgálata azt találta, hogy az RZK referátumai jól fogalmazottak és hosszabbak voltak, több adatot adtak az összetételről és a fizikai jellemzőkről, mint a CA referátumai. Legnagyobb gyengeségük az elméleti és általános következtetések leírásában, a vegyületek szerkezetének grafikai reprezentációjának hiányában volt. Nem volt különbség az angol és az orosz dokumentumok kezelésében.

Milas-Bracovic és Zajec (1989) horvát tudományos folyóiratokban megjelent, kutatásokról beszámoló cikkek referátumainak minőségét vizsgálta, újszerű megközelítést alkalmazva. A szerzők kiválasztottak 1372 tanulmányt és azok referátumát (53%-ot a természettudományok, orvostudomány és műszaki tudományok területéről, 47%-ot a társadalomtudományok, humán tudományok és művészetek területéről), 160 Horvátországban megjelenő folyóirat közül 98-ból.

A referátumokat információs szakemberek elemezték, hogy meghatározzák a makrostruktúrák mintáit, benne azokat a részeket (bevezetés, módszerek, eredmények és tárgyalás), amelyek megléte elvárható egy jó referátumban, és az informatív (62%), indikatív (9%) és vegyes típusú (29%) referátumok százalékos arányát. Jelentős különbségeket találtak a két részhalmaz között mind a makrostruktúrában, mind a referátum típusainak százalékos megoszlásában.

Ezt követően elemezték a referátumokat, hogy meghatározzák, a referátumok milyen mértékben térnek el a referálási szabványtól a nyelvtan (22%), a harmadik személyű forma (6%), terjedelem (15%), önállóság (self-containment) (5%) és tömörség (17%) szempontjából. Itt is lényeges különbségek voltak a két mintaként használt részhalmaz között.

A minőséget ellenőrző próbaként ezt követően a referátumokat összehasonlították azzal a változattal, amelyet a *Bulletin Scientifique* francia referáló és indexelő kiadvány számára készítettek, hogy meghatározzák, milyen százalékban használták fel őket változtatás nélkül (58%), átírva (41%), és mennyit írtak teljesen újra (1%). Ebben a kategóriában a két részhalmaz közötti különbségek minimálisak voltak. Csak társadalomtudományi, művészeti és humántudományi cikkeket kellett újraírni.

Tenopir és Jacsó (1993) a referátumok leginkább mérhető szempontjait tesztelte, három, a nagyközönségnek szóló periodikumokat feldolgozó adatbázis mintáit használva: *Magazin Article Summaries* (MAS) 967 rekorddal; a *Resource One*-t 1070 rekorddal és *Readers' Guide Abstracts* (RGA) 839 rekorddal. A *Journal of Social Psychology* cikkeinek referátumait két társadalomtudományi adatbázisban: a PsycLITben (a PsycINFO CD-ROM-változata) 315 rekordot és a Sociofile-ban 379 rekordot is tesztelve.

A Grammatik-IV stílus ellenőrző program segítségével határozták meg az olvasási szint kvantitatív értékeit; a bekezdések, mondatok és szavak hosszúságát; a szenvedő szerkezet; a prepozíciók; a hosszú, közepes és rövid mondatok százalékarányát. A nyers számokat azután az olvashatóság és az ANSI szabvánnyal való kompatibilitás szempontjából rendezték és értelmezték. Az informativitást úgy értékelték, hogy minden egyes adatbázisból 300 véletlenszerűen kiválasztott rekordot megvizsgáltak, hogy meghatározzák az indikatív, informatív és kombinált típusú referátumok százalékarányát.

Az RGA-t találták a leginformatívabbnak annak döntő többségében terjedelmes és alapos referátumai miatt, amelyek majdnem ugyanolyan pontszámot kaptak az olvashatóságra és a fokozatok szintjére, mint a MAS. A Resource One-ban vannak a legkönnyebben olvasható referátumok, elsősorban a rövid mondatok miatt. A társadalomtudományi adatbázisok között az olvasás könnyősége szempontjából a PsycLIT pontszámai sokkal jobbak voltak, mint a Sociofile-éi, de ez utóbbiban sokkal hosszabb referátumok voltak.

A referátumok értékelésének folyamata

A forrásdokumentumok és a belőlük előállított különböző fájlok készítői által létrehozott referátumok olvasása marad a legjobb módszer a referátumok értékelésére a korábban idézett tankönyvek szempontjai alapján. A referáló és indexelő folyóiratok digitális verziójának és a keresőszoftvernek a kombinációja azonban olyan lehetőségeket kínál, amelyekkel a folyamat számos fázisa könnyebbé válik. Olyan rekordok kiválasztása, amelyekben van referátum, könnyű dolog olyan szoftverrel, amely lehetővé teszi a szűrést ilyen feltétel szerint. Ezt a lépést egy meglevő halmaz alapján lehet elvégezni. Az ilyen halmaz hasonló módon hozható létre, mint bármely másik (azaz tárgy, a folyóirat címe, a szerző neve, dátum stb. alapján). A korlátozás mind parancsmódban, mind menü módban elvégezhető, ez utóbbiban egy filter kocka használatával. A 11.1. és 11.2 ábrák illusztrálják a két megközelítést.

Set	Term Searched	Items
S1	PAIN?/ENG,MAJ,HUMAN,ABS	21160

11.1. ábra: Korlátozás a DIALOG-ban parancsmódban

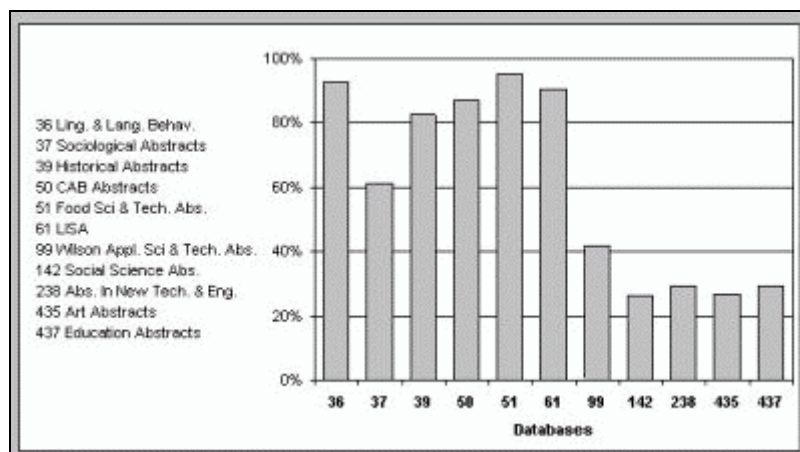
The screenshot shows the Medline Ovid web interface. At the top, it says "Medline 1996 to September 1999 Week 2". Below this is a toolbar with icons for Author, Title, Journal, Search Fields, Tools, Combine, Limit, Basic, Change Database, and Logoff. A "Search History" table is visible, showing a single entry with a dash in the first column and a dash in the third column. Below the table, there is a "Run Saved Search" button. The main search area has a text input field with "toxoplasmosis" and a "Perform Search" button. There are also checkboxes for "Map Term to Subject Heading" (checked), "Full Text Available", "EBM Reviews", "Human", "English", "Reviews", "Abstracts" (checked), and "Latest Update". At the bottom, there are dropdown menus for "From: 1969" and "To: 1999".

11.2. ábra: Korlátozás az Ovid haladó menümódjában

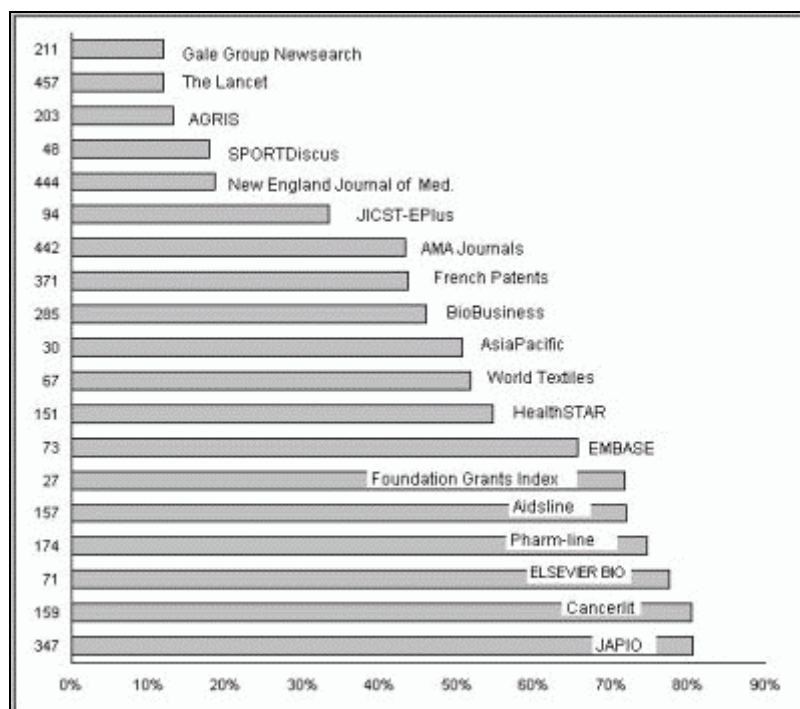
Csak azért, mert az adatbázis címe tartalmazza az *abstracts* szót, még nem biztos, hogy minden rekordban van referátum. Az ilyen adatbázisokból vett minták azt mutatják, hogy számos adatbázisban a rekordok változó százalékában nincs referátum (11.3. ábra). Egyes adatbázisok dokumentációja megjegyzi, hogy csak egy bizonyos évtől kezdve vagy egy adott dokumentumtípushoz kapcsolódóan található referátumok, de sok másik semmilyen figyelmeztetést nem ad a referátumok hiányáról.

Sajnos sok adatbázis, amelynek a címében szerepel az *Abstracts* szó, nem teszi lehetővé a keresési halmaznak a referátumokkal rendelkező rekordokra való szűkítését, így nem lehetett tesztelni. Az, hogy ilyen korlátozási lehetőséget nem adnak, első látásra logikusnak látszik, feltételezve azt, hogy az olyan adatbázisokban, mint a LISA és az ISA, minden rekordban kell lennie referátumnak, de nem valószínű, hogy ez érvényes feltételezés lenne. Ha referáló adatbázisokból származó halmazokban sok referátum nélküli rekordot találunk, ez nem csak frusztráló, hanem pénz pazarlás is lehet, mivel azt jelenti, hogy drágább, referátumot tartalmazó megjelenítési formát választunk, ennek megfelelően többet fizetünk, csak azért, hogy rájövünk, számos rekordban nincs referátum.

A referátumok hiánya olyan adatbázisokban, amelyek címében nem szerepel az *Abstracts* szó, érthetőbb és elég elterjedt, ahogy azt a 11.4. ábra mutatja. Egyesek figyelmeztetik a használókat a dokumentációban arra, hogy ez a fontos adatelem nem található meg minden rekordban. (Az más kérdés, hogy hány használó olvassa el a dokumentációt.)



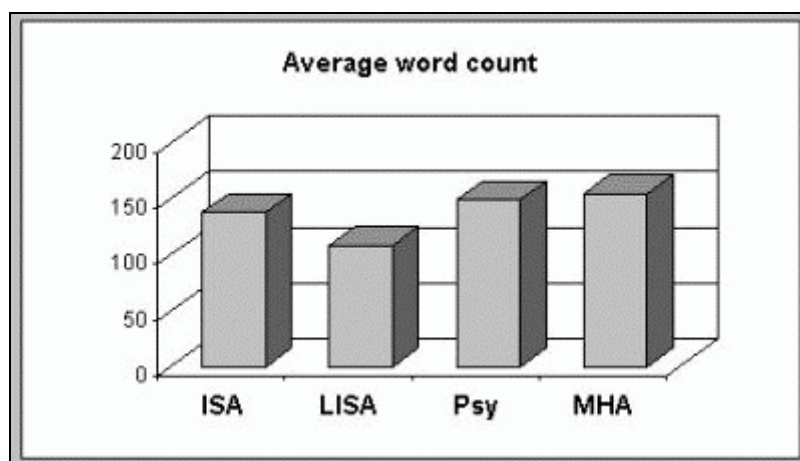
11.3. ábra: A referátumokat tartalmazó rekordok aránya olyan adatbázisokban, amelyeknek a címében szerepel az *Abstracts* szó



11.4. ábra: A referátumokat tartalmazó rekordok aránya olyan adatbázisokban, amelyeknek a címében nem szerepel az *Abstracts* szó

Sok adatbázisban a referátumok sokkal rövidebbek, mint az ANSI által javasolt 150-250 szó. Elég könnyen lehet tesztelni egy adatbázisban a referátumok átlagos hosszát, ha letöltünk néhány száz mintarekordot, és kihasználjuk egy szövegszerkesztő szószámlálási lehetőségét. Ez drága lehet egy online adatbázisban, amelyben rekordonként kell fizetni, de ingyenes a CD-ROM-os adatbázisok esetében. A 11.5. ábra illusztrálja a referátumok átlagos hosszát a LISA, ISA, PsycINFO és Mental Health Abstracts adatbázisban. Ugyanazt a széles tárgykörre vonatkozó keresést végeztük el mindkét adatbázis pár referátum mezőjében, azután az első 500 rekordot töltöttük le mindegyikből, és értékeltük a Microsoft Word szószámlálási lehetőségének segítségével. Ez talán nem statisztikailag reprezentatív minta, de nagyon jól működik azok számára, akik meg akarják érezni, milyen terjedelmű referátumok szerepelnek

az egyes adatbázisokban. A PsycINFO és az MHA elég közel vannak egymáshoz, míg az ISA kétségtelenül előnyben van a LISA-val szemben.



11.5. ábra: A referátumokban szereplő szavak átlagos száma négy adatbázisban

Közismerten a legrövidebb és legkevésbé informatív referátumok a receptekről, könyv- és filmkritikákról készült rekordokban szerepelnek. Például egy bestsellerről, a *Bridget Jones' Diary* (Bridget Jones naplója) című könyvről két tucat kritika rekordja szerepel az EBSCO Academic Search Elite adatbázisában. A referátum valamennyiben ugyanaz az egy sor (11.6. ábra), kivéve azt az egyet, ahol a *Bridget* szót *Briget* formában szerepel (azt igazolva, hogy ezek a „referátumok” nem a kivágás és beillesztés (cut and paste) módszerrel kerülnek egyik rekordból a másikba).

Title:	Adult books: Fiction.
Subject(s):	BRIDGET Jones's Diary (Book)
Source:	Booklist, Jul98, Vol. 94 Issue 21/22, p1859, 1/4p, 1bw
Author(s):	Wilkinson, Joanne
Abstract:	Reviews the book 'Bridget Jones's Diary,' by Helen Fielding .
AN:	904470
ISSN:	0006-7385
Database:	Academic Search Elite
Title:	The marriage mystique.
Subject(s):	BRIDGET Jones's Diary (Book)
Source:	New Yorker, 08/03/98, Vol. 74 Issue 22, p70, 6p, 1c
Author(s):	Merkin, Daphne
Abstract:	Reviews the book 'Bridget Jones's Diary,' by Helen Fielding .
AN:	916645
ISSN:	0028-792X
Database:	Academic Search Elite

Helen Fielding „Bridget Jones” című könyvének kritikája.
Helen Fielding „Briget Jones” című könyvének kritikája.

11.6. ábra: Könyvkritikák rekordjainak egyforma „referátumai” az Academic Search Elite-ben

A (később tárgyalandó) tesztek azt mutatták, hogy a Bell & Howell általában nagyon jó, informatív referátumokat készít az információs technikával foglalkozó folyóiratcikkekről. Ugyanez nem mondható el a könyvkritikák rekordjairól. Az Academic Search Elite-ben látott, nem informatív „referátumok” jellemzőek a Periodical Abstracts PlusText adatbázisra is (11.7. ábra). Érdekes, hogy testvéradatbázisában, a Newspaper Abstracts Dailyben a legtöbb könyvkritikáról terjedelmes referátum található, hasonló ahhoz, mint ami a 11.8. ábrán láthatóak a *Bridget Jones' Diary*-ről.



Gilbert kritikája a Helen Fielding által írt „Bridget Jones naplójá”-ról.
Durnell kritikája a Helen Fielding által írt „Bridget Jones naplójá”-ról.

11.7. ábra: Könyvkritikák rekordjainak „referátumai” a Periodical Abstract PlusText adatbázisban

Ha nem hallott a „Bridget Jones naplójá”-ról ..., nos, akkor önnek jó. A regény, Angliában bestseller egy 30-as éveiben járó szingli nő történetét mondja el, aki egy kiadónál dolgozik, olyan méretű szenvedéllyel, ami egy halott tekintetéből sugárzik. Ms. (Bridget) Jones meséjét napló formájában mondja el, minden bejegyzés végén feljegyezve a súlyát, a régebbi és potenciális szeretőihez intézett sikertelen telefonhívások, elszívott cigaretták, elfogyasztott kalóriák, bekapott falatok és alkoholos italok mennyiségét. Gyötrődik amiatt, hogy a férfiak miért bánnak rosszul vele, mennyire önmagukkal elfoglaltak és érzelmileg visszamaradottak, azután órákat tölt azzal, hogy vágyakozóan bámulja a telefont. A könyv Angliában ünnevelt hírességgé tette a szerzőt, Helen Fieldinget, s csaknem minden amerikai női magazin, a Vogue-tól a Mirabelláig és a Jane-ig dicsérte a Viking kiadónál való megjelenése után. Sok nő látszólag nevet azon, hogy felismerik önmaguk vonásait Bridget (Jones) zűrzavaros bizonytalanságában (épp úgy, ahogy beavatottan özszemosolyognak Ally McBeallel). Bridget univerzuma párokkal van benépesítve, akiket önelégült házasoknak nevez, és szinglinek nevezett egyedülálló barátokkal.

Bridget ambivalens a házassággal kapcsolatban - azok, akik elérik, azonnal önelégült bugrisokká válnak - nem hagy kétséget a szingli állapot felől: a független nő hiszékeny, vénlány, olyasvalaki, aki megérdemli a lekicsinylő címkét. A cselekmény egyszerű: olyan férfit keres, aki válaszol a telefonhívásaira, s szeretne négy kilót leadni.



**11.8. ábra: Egy könyvkritika terjedelmes referátuma
a Newspaper Abstracts Daily adatbázisban**

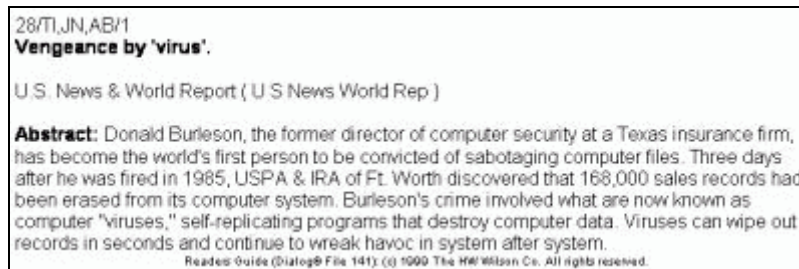
A referátumok átlagos hossza csak az első, nyers megközelítés a referátumok értékelésekor. A referátumok hosszúsága nem garantálja informatívságukat. A hosszabb nem feltétlenül jobb, bár a hosszabb referátumoknak természetesen nagyobb az esélyük arra, hogy a primer dokumentum alapvető információi szerepeljenek bennük. A referálásról és indexelésről szóló számos klasszikus munka szempontokat határoz meg arra vonatkozóan, mitől lesz jó egy referátum. Ez az ítélet azonban továbbra is szubjektív marad, s megköveteli a referátumok és a primer dokumentumok elolvasását.

Esettanulmány a referátumok minőségének értékeléséről

A legkönnyebben akkor kezdhetünk hozzá az értékeléshez, ha ugyanakkor a cikknek a referátumait tanulmányozzuk. Ez viszonylag könnyű a nagyközönségnek szóló periodikumokat feldolgozó adatbázisok esetén, s nehezebb a tudományos és műszaki dokumentumok referátumaival. A 11.9., 11.10. és 11.11. ábra ugyanannak a cikknek három adatbázisban megtalálható referátumát mutatja.

A Readers' Guide Abstractsből származó messze a legjobb. Ez választ ad a négy w-re: who, what, when és why (ki, mit, mikor, miért). Az Academic Search Elite referátumából hiányzik néhány fontos részlet. A Periodical Abstracts Plus Text referátuma nem említi meg azt az alapvető tényt, hogy a vádlottat elbocsátották néhány nappal a bűntény elkövetése előtt. Erre az információra szükség lett volna a cím megértéséhez.

Donald Burleson, egy texasi biztosítási cég volt számítógépes biztonsági igazgatója lett a világ első embere, akit számítógépes fájlok szándékos megrongálásaért elítéltek. Három nappal 1985-ös elbocsátása után az USPA & IRA of Ft Worth felfedezte, hogy 168000 eladási rekordja törlődött számítógéprendszeréből. Burleson bűntényéhez kapcsolódott az, amit most számítógépes „vírusként” ismerünk, ezek olyan önmagukat sokszorozító programokat, amelyek elpusztítják a számítógépes adatokat. A vírusok másodpercek alatt kitörölhetnek rekordokat, s folytatják bosszúálló pusztításukat rendszerről rendszerre.



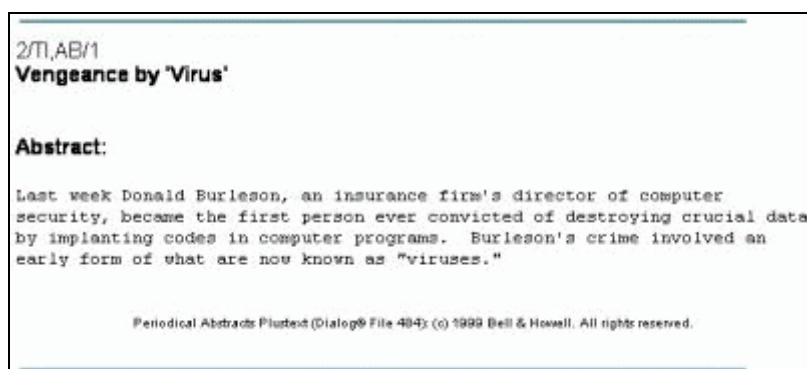
11.9. ábra: A „Bosszú” cikk referátuma az RGA-ban

Olyan számítógépes programokba telepített kódokkal foglalkozik, amelyek fontos adatokat elpusztítanak (vírusok). Az elmúlt héten Donald Burleson lett az első olyan személy, akit ilyen szándékos rongálásért elítéltek. Büntény; Költségek.



11.10. ábra: A „Bosszú” cikk referátuma az Academic Search Elite-ben

Az elmúlt héten Donald Burleson, egy biztosítási cég számítógépes biztonsági igazgatója lett az első olyan személy, akit elítéltek azért, hogy fontos adatokat tönkretett azzal, hogy kódokat telepített számítógépes programokba. Burleson bünténye kapcsolódott annak korai formájához, amit ma „vírusok”-ként ismerünk.



11.11. ábra: A „Bosszú” cikk referátuma a Periodicals Abstracts PlusTextben

A „Crimson Copycat” („Harvardi Hamiska”) szerkesztőségi cikk rekordjai azt illusztrálják, hogy ugyanabból a cikkből milyen különböző megközelítésű referátumok készültek. Az Academic Search Elite referátuma (11.12. ábra) egy meghatározott eseményre koncentrál, s a legtöbb részletet adja, míg a Periodical Abstracts PlusText (11.13. ábra) meg sem említi az aktuális eseményt, de jól összegzi a szerző gondolatait. Az RGA kombinálja ezt a két megközelítést, tökéletes egyensúlyt tartva.

Kommentár a plágiumról. Beszámol arról, hogy Shervert Frazier, a National Institute of Mental Health (a mentális egészség országos intézete) volt igazgatója, az American College of Psychiatrists (a pszichiáterek amerikai kollégiuma) volt elnöke, a Harvard Medical School pszichiátriai professzora, és a bostoni McLean Kórház fő pszichiátere elismerte a plágiumot és lemondott. Frazier plágiumát a University of Rochester diplomás hallgatója, Paul Scatena fedezte fel.

Title: The *crimson copycat*.
Subject(s): FRAZIER, Shervert
Source: U.S. News & World Report, 12/12/88, p90, 1p
Author(s): Leo, J.
Abstract: Commentary on plagiarism. Report that Shervert Frazier, former director of the National Institute of Mental Health, former president of the American College of Psychiatrists, professor of psychiatry at Harvard Medical School and psychiatrist in chief at Boston's McLean Hospital admitted to plagiarism and resigned. Frazier's plagiarism was uncovered by graduate student Paul Scatena of the University of Rochester.
AN: 8800029690
ISSN: 0041-5537
Database: Academic Search Elite

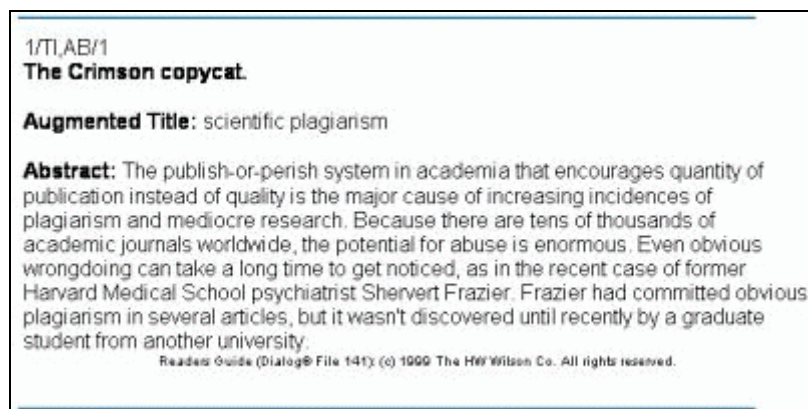
11.12. ábra: A „Harvardi Hamiska” cikk referátuma az Academic Search Elite-ben

A publikálj vagy pusztulj rendszer, amely most áthatja az egyetemi előremenetelt, nemcsak a közepszerűséget bátorítja, hanem növeli a plagizálás csábítását. Ennek a rendszernek a melléktermékei nemcsak olvashatatlan cikkek, de valójában nem olvasásra szánt egyetemi folyóiratok ezreinek kialakulása.

31/TI,JN,AB/1
The Crimson Copycat
 US News & World Report
Abstract:
 The publish-or-perish system which now permeates the ranks of academia, not only encourages mediocrity, but heightens the temptation to plagiarize. The byproducts of this system are not only unreadable articles but also the rise of thousands of academic journals that are not actually meant to be read.

11.13. ábra: A „Harvardi Hamiska” cikk referátuma a Periodicals Abstracts PlusTextben

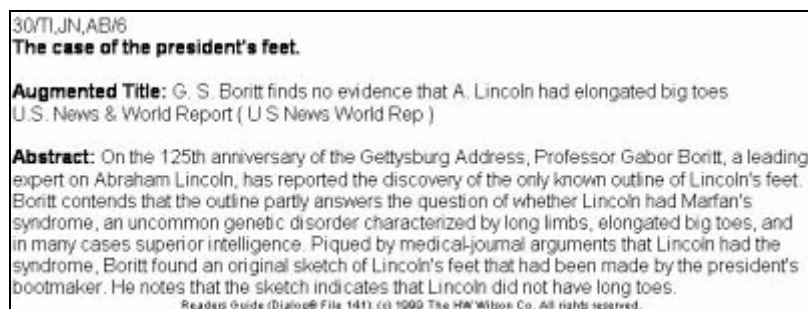
A publikálj vagy pusztulj rendszer, amely az egyetemi körökben a publikációk mennyiségét bátorítja a minőség helyett, a fő oka plágiumok és a közepszerű kutatások növekvő számának. Mivel a világon egyetemi folyóiratok ezrei léteznek, a visszaélések lehetősége hatalmas. Még a nyilvánvaló törvénysértések észrevételéhez is hosszú időre van szükség, mint ahogy Shervert Frazier, a Harvard Medical School volt pszichiátere mostani esetében is történt. Frazier nyilvánvaló plágiumot követett el számos cikkben, de ezt csak most fedezte fel egy másik egyetem diplomás hallgatója.



11.14. ábra: A „Crimson Copycat” cikk referátuma az RGA-ban

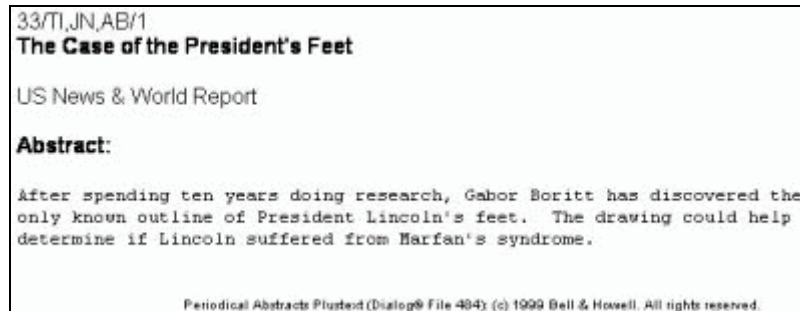
A harmadik cikk (amelyre egy volt hallgató, Marylin Rappun hívta fel a figyelmemet sok évvel ezelőtt) szintén magába sűríti az RGA néhány kivételesen értékes jellemzőjét. Ez az egyetlen referátum, amely közli a cikk következtetését, hogy nincs bizonyíték arra, hogy Lincoln Marfan szindrómában szenvedett volna (11.15. ábra). Az RGA bónusza, hogy tartalmazza a csattanót, mint a cím kiegészítését, így ez még a rövid megjelenítési formában is látható, és a *Marfan* szó kereshetővé válik a címindexben is. A másik két referátum kihagyja ezt a fontos következtetést. Ugyanaz az informativitás érvényes a Marfan szindrómát meghatározó mondatra is. Az RGA röviden megmagyarázza ezt, a Periodical Abstracts PPlusText (11.16. ábra) csupán jelzi, hogy ez egy genetikai rendellenesség, az Academic Search Elite (11.17. ábra) célzást sem tesz a betegséggel kapcsolatos következtetésekre.

A Gettysburgi beszéd 125. évfordulóján Professor Gabor Boritt, Abraham Lincoln egyik legjobb szakértője bejelentette, hogy felfedezte az egyetlen ismert ortopéd lábnyomatot Lincoln lábáról. Boritt azt állítja, hogy a lenyomat részben megválaszolja azt a kérdést, hogy vajon Lincoln szenvedett-e Marfan-szindrómában, egy ritka genetikai rendellenességben, amelyet hosszú végtagok, megnyúlt nagylábujjak és sok esetben kiváló intelligencia jellemez. Az orvosi folyóiratok azon állításaitól felajzva, hogy Lincoln szenvedett ebben a szindrómában, Boritt megtalálta Lincoln lábának eredeti lenyomatát, amelyet az elnök csizmakészítője használt. Megjegyzi, hogy a vázlat azt jelzi, hogy Lincolnnak nem voltak hosszú lábujjai.



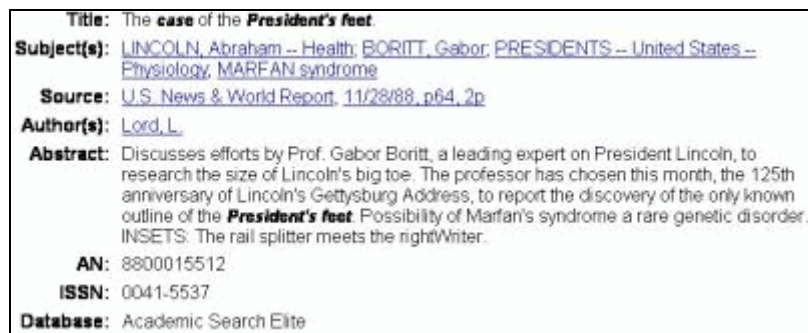
11.15. ábra: A „Lincoln” cikk referátuma az RGA-ban

Miután tíz évet töltött kutatással, Gabor Boritt felfedezte Lincoln elnök lábfejeének egyetlen ismert körvonalát. A rajz segíthet meghatározni, hogy Lincoln szenvedett-e a Marfan-szindrómában.



11.16. ábra: A „Lincoln” cikk referátuma a Periodical Abstracts PlusTextben

Megtárgyalja Prof. Gabor Boritt, Lincoln elnök vezető szakértője próbálkozásait Lincoln nagylábujja méretének megállapítására. A professzor ezt a hónapot, Lincoln Gettysburgi beszédének 125. évfordulóját választotta ki arra, hogy bejelentse az elnök láblenyeomatának egyetlen ismert körvonala felfedezését. A Marfan-szindróma, egy ritka genetikai rendellenesség lehetősége.



11.17. ábra: A „Lincoln” cikk referátuma az Academic Search Elite-ben

A leggyakorlatiasabb megközelítés az összehasonlító elemzés céljaira az, ha a folyóiratok címe alapján összeállítunk egy alapvető halmazt, olyan folyóiratokat választva, amelyekről tudjuk, hogy nem tartalmaznak referátumokat. A könyvtár- és információtudomány területén az ilyen folyóiratok közé tartozik az *Online*, a *Database* (most *EContent*), a *Searcher* és a *Computers in Libraries*. Különbözik nagyon sok egyforma vagy legalább nagyon hasonló referátumot kapnánk, amelyeket a szerzői referátumokat tartalmazó folyóiratok eredeti cikkeiből vettek, ilyen a legtöbb egyetemi tudományos folyóirat. A fájlok készítői örömmel veszik át ezeket a referátumokat, sok időt és pénzt megtakarítva. Minden szakterületen vannak azonban jó minőségű folyóiratok, amelyek nem tartalmaznak a szerző vagy a szerkesztő által készített referátumokat. Ezek a folyóiratok tökéletes jelöltek a referáló és indexelő szolgáltatások által készített referátumok összehasonlítására és értékelésére.

Ezt követően ugyanannak a cikknek a különböző adatbázisokban megtalált duplikátumai egy összehasonlítható halmazt képeznek. Ez a halmaz értékelhető elolvasva és egy olcsó, alapos nyelvi ellenőrző programon végigfuttatva.

A referátumok tipikus erősségeinek és gyöngeségeinek illusztrálására a szerző néhány cikkét (egy esetben társszerzőségben a könyv szerkesztőjével) választottuk ki. Ez gondoskodott arról a követelményről, hogy az értékelőnek ismernie kell az eredeti cikket. A kiválasztott cikkek a

CD-ROM Professionalben, az Online-ban és a Database-ben jelentek meg, ezek egyike sem tartalmaz referátumokat, ebből következően jelentős különbségek voltak várhatók.

Néhány, a könyvtár- és információtudományi anyagot kiválóan feldolgozó adatbázis nem jelenik meg a példák között. Az H. W. Wilson cég Library Literature adatbázisa egyáltalán nem tartalmaz referátumokat. Az ISI Social SciSearch adatbázisa csak eredeti referátumokat vesz fel, minden módosítás nélkül. A következőkben tömör, illusztratív elemzéseket mutatunk be, cikkről cikke, a következő munkákhoz kapcsolódóan:

- „Quality of Abstracts” (A referátumok minősége)
- „What Is in a(n) (Up)Date? Currency Test Searching of Databases” Mi van az aktualizálásban (dátumban)? (Adatbázisok frissességének vizsgálata)
- „Data Transfer Capabilities of CD-ROM Software,” (CD-ROM szoftver adatátviteli képessége, 1. és 2. rész)

A „Quality of Abstracts” cikk

A referátumok minőségével foglalkozó cikkekre vonatkozóan az ABI/INFORM-ban szerepel a leginformatívabb és legjobban szerkesztett referátum (11.18. ábra). Ez tartalmazza a cikk legtöbb témáját, kivéve a PsycLIT és a Sociofile referátumai olvashatóságának alapvető összehasonlító elemzését (amelyet az összes többi adatbázis is kihagyott). Meghatározza a három, a nagyközönségnek szóló periodikumokat feltáró adatbázist, s ugyanígy a szempontokat, valamint a teszteléshez használt szoftvereszközöket.

Érdekes módon a Periodical Abstracts PlusText adatbázis (az ABI/INFORM testvérpublikációja) nem használja ugyanazt a referátumot (11.19. ábra), ami a több fájlt előállítók általános gyakorlata. Ez a megközelítés biztosan többbe kerül a fájl előállítójának, de igazolható az eltérő célközönséggel, amely valószínűleg különböző referátumokat igényel. A Periodicals Abstracts PlusTextben található referátum sokkal rövidebb és nem tartalmazza a számszerű adatokat, de megemlíti a legfontosabb elemzési szempontokat és meghatározza a nagyközönségnek szóló folyóiratokat feldolgozó adatbázisokat. Ez nagyon jól összehasonlítható a H. W. Wilson cég Business Abstracts adatbázisával (11.20. ábra) és az ISA-val (11.21. ábra). Az utóbbi dicséretet érdemel azért, mert több részletet megemlít a cikkben tárgyalt témáról, s elmarasztalást a „the authors focuses” nyelvtani hibájáért.

A nyomtatott, online és CD-ROM-os indexelő és referáló kiadványok ismert tájékoztató eszközök a könyvtárakban. Az American National Standards Institute (ANSI, az amerikai országos szabványosítási intézet) a referátumok három típusát ismeri el: 1. indikatív vagy leíró, 2. informatív és 3. vegyes típusú (indikatív és informatív). A referátumok minőségét vizsgálták a három legfőbb, a nagyközönségnek szóló periodikumokat feldolgozó, referátumokat tartalmazó indexben - Reader's Guide Abstracts, Resource-One és Magazine Article Summaries -, kvantitatív vagy objektív teszteket alkalmazva annak érdekében, hogy általános képet kapjanak a minőségről. A referátumokat a következő szempontokból vizsgálták: 1. a stílus következetessége és az olvashatóság, 2. az, hogy milyen mértékben vették figyelembe az ANSI szabványát 3. informativitás. Az olvashatóságot a Grammatik-IV szoftverrel határozták meg. Az eredmények azt jelzik, hogy a Resource-One referátumai a legkönnyebben olvashatóak, de a Reader's Guide átlagosan 110 szavas referátumai felelnek meg legjobban az ANSI szabványai által ajánlott terjedelemnek (100-250 szó). A Reader's Guide-ot találták a leginformatívabbnak, referátumainak 81%-a tartozott az informatívok közé.

DIALOG(R)File 15:(c) 2000 Bell & Howell. All rts. reserv.

Quality of abstracts

ABSTRACT: Printed, online, and CD-ROM indexing and abstracting publications are familiar reference tools in libraries. The American National Standards Institute (ANSI) recognizes 3 types of abstracts: 1. indicative or descriptive, 2. informative, and 3. combination (indicative and informative). The quality of the abstracts in 3 major CD-ROM general periodical indexes that contain abstracts - Reader's Guide Abstracts, Resource-One, and Magazine Article Summaries - was tested using quantitative or objective tests to provide a generalized view of quality. The abstracts were examined for: 1. consistency of style and readability, 2. the extent to which the ANSI standard was observed, and 3. informativeness. Readability was determined using Grammatik-IV software. The results indicate that the abstracts of Resource-One are the easiest to read, but the Reader's Guide abstracts, with an average of 110 words, most closely match the ANSI standards for recommended length (100 to 250 words). The Reader's Guide was also judged the most informative, with 81% of its abstracts of the informative type.

11.18. ábra: Referátum az ABI/INFORM-ból

A három legfőbb, a nagyközönségnek szóló lapokat feldolgozó indexben található referátumok minőségét hasonlították össze a stílus következetessége, az olvashatóság, az American National Standards Institute (ANSI) szabványának megfelelés mértéke és az informativitás szempontjából. A három tanulmányozott index az H. W. Wilson cég által kiadott Reader's Guide Abstracts, a UMI Periodical Abstracts Ondiscje és az EBSCO Magazin Articles Summaries című adatbázisa volt.

DIALOG(R)File 484:(c) 2000 Bell & Howell. All rts. reserv.

Quality of abstracts

ABSTRACT: The quality of abstracts found in the three major CD-Rom general periodical indexes is compared for consistency of style and readability, the extent to which the American National Standards Institute (ANSI) standard is observed and informativeness. The three indexes studied were H. W. Wilson Co's Readers' Guide Abstracts, UMI's Periodical Abstracts Ondisc and EBSCO's Magazine Article Summaries.

11.19. ábra: Referátum a Periodical Abstracts PlusTextből

Leírja egy referátum minősége megítélésének szempontjait, benne az alkalmazott referálási stílussal és a referátum tartalmával és terjedelmével, s ezt használták fel a referátumok minőségének értékelésére három adatbázisban: az H. W. Wilson cég Reader's Guide Abstracts-jében, a UMI Periodical Abstracts Ondiscje és az EBSCO Magazin Articles Summaries című adatbázisában.

DIALOG(R)File 553:(c) 2000 The HW Wilson Co. All rts. reserv.

Quality of abstracts.

AUGMENTED TITLE: H.W. Wilson's Readers' Guide Abstracts, UMI's Periodical Abstracts Ondisc and EBSCO's Magazine Article Summaries

ABSTRACT: The criteria for judging the quality of an abstract, including the abstracting style used and the abstract's content and length, are described and used to evaluate the quality of abstracts in three databases: H.W. Wilson Company's Readers' Guide Abstracts, UMI's Periodical Abstracts Ondisc, and EBSCO's Magazine Article Summaries..

11.20. ábra: Referátum az H. W. Wilson cég Business Abstracts adatbázisából

A cikk azokat a tényezőket vizsgálja, amelyek hozzájárulnak az információs adatbázisok jó minőségű referátumaihoz, az American National Standards Institute 1979-ben alkotott szabványaira koncentrálva. A szerzők a három, a nagyközönségnek szóló lapokat feldolgozó index, a Reader's Guide Abstracts, a Periodicals Abstracts On Disc és a Magazine Article Summaries referátumainak előnyeire és hátrányaira koncentrálnak. A szerzők megmagyarázzák, mi a referátum, kik írják a referátumokat, mi kerül be a referátumba, a referátum terjedelmét, a referátum olvashatóságának mérőszámait és egyéb változókat.

DIALOG(R)File 202:(c) Information Today, Inc. All rts. reserv.

Quality of abstracts.

The article examines those factors that contribute to high-quality abstracts contained in information databases, focusing on standards established in 1979 by the American National Standards Institute. The authors focuses on the benefits and drawbacks of abstracts contained in three general periodical indexes, including Readers' Guide Abstracts, Periodical Abstracts On Disc and Magazine Article Summaries. The authors explain what an abstract is, who writes abstracts, what goes into an abstract, length of abstract, measures of abstract readability and other variables.

11.21. ábra: Referátum az ISA-ból

A Trade & Industry Database-ben található referátum (11.22. ábra), amelyet az Information Access Company (ma a Gale Group része) készített, szokatlanul aránytalan, a cikk első negyedére koncentrál. Meg sem említi a tesztelés folyamatát és szempontjait, sőt még a tesztelt adatbázisokat sem. A referátum második mondata egy részének („to help focus a review of an on-line, printed or CD-ROM database” az eredeti cikk szerzője számára sincs értelme. A referátumok készítésének funkcióját korlátozza, azt állítva, hogy „a referátumok lehetővé teszik a keresők számára, hogy az eredeti dokumentum rövidebb változatát megnézhessék.” Ennek semmi köze a médiumhoz; a nyomtatott referáló/indexelő publikációban szereplő referátumnak megvan ugyanez az előnye.

A referátumok lehetővé teszik az online keresőknek, hogy megnézzék egy eredeti dokumentum rövid változatát. A referátumok további keresési kifejezéseket is szolgáltatnak, hogy segítsék egy online, nyomtatott vagy CD-ROM-os adatbázis áttekintését. A referátumot írhatja az eredeti cikk szerzője is; a professzionális szolgáltatások azonban általában saját referátumokat készítenek. A referálókat alkalmazhatják íráskészségük miatt, de egyes esetekben a referálás és indexelés feladatait kombinálhatják. A referátumok minősége megítélhető olvashatóságuk és informativitásuk alapján.

DIALOG(R)File 148:(c)2000 The Gale Group. All rts. reserv.

Quality of abstracts.

ABSTRACT: Abstracts allow on-line searchers to view a short version of an original document. Abstracts also provide additional search terms to help focus a review of an on-line, printed or CD-ROM database. An abstract may be written by the author of an original article; however, professional services typically write their own abstracts. Abstracters may be hired for writing skills, but in some cases abstracting and indexing tasks may be combined. The quality of abstracts can be judged by readability and informativeness.

11.22. ábra: Referátum a Trade & Industry Database-ből

A LISA-ban található referátum (11.23. ábra) megfelelő módon ír a témákról és az elemzési szempontokról, jelzi, hogy statisztikai adatok szerepelnek a cikkben, de nem határozza meg a tesztelt adatbázisokat. Nem teszi meg ezt az ERIC referátuma sem (11.24. ábra), amely a statisztikai adatok jelenlétéről is hallgat, de tartalmazza a hivatkozások számát. Az általában jó referátumokat készítő INSPEC biztosan nem bizonyítja ezt ezzel a cikkel (11.25. ábra). A referátum a cikk előzményekről szóló részéből két bekezdés másolata. Teljes mértékben hiányzik belőle a cikk lényege, s félrevezeti az olvasót azzal kapcsolatban, miről is szól a cikk.

A kivételek, amikre az előbb utaltunk, azt jelzik, milyen fontos referátumok tucatjainak elemzése annak érdekében, hogy megfelelő képet kapjunk a referátumok minőségéről, s ne ítéljük meg az adatbázist néhány nem jellemző módon gyenge rekord miatt.

A referátumok lehetővé teszik olvasóik számára, hogy megítéljék cikkek relevanciáját, összefoglalást nyújtanak az eredeti dokumentumról, és esetleg helyettesíthetik is azt. Meghatározza a referátumokat és foglalkozik azzal, hogy kik írják az American National Standard Institute (ANSI) és más források szerint. Megkülönbözteti az indikatív és informatív referátumokat. Az ANSI és az ERIC az informatív referátumokat támogatja. Megtárgyalja a referátum tartalmát és a referálás műveleteit, a megírás stílusát, a minőség, olvashatóság és informativitás tesztjeit. Statisztikai adatokat mutat be, három, a nagyközönségnek szóló adatbázis referátumait elemezve, továbbá a referátum terjedelméről és típusáról. GLC.

```
DIALOG(R)File 61:(c) 2000 Reed Reference Publishing. All rts. reserv.

Quality of abstracts.

ABSTRACT: Abstracts enable users to judge the relevance of articles,
provide a summary and may be a substitute for the original document.
Defines abstracts and considers who they are written by according to the
American National Standards Institute (ANSI) and other sources.
Distinguishes between indicative and informative abstracts. Informative
abstracts are preferred by ANSI and ERIC. Discusses the content and
procedures for abstracting, writing style, tests of quality and
readability and informativeness. Presents statistics analyzing abstracts
from 3 general interest databases and on abstract length and type. GLC.
```

11.23. ábra: Referátum a LISA-ból

Áttekinti a referátumok értékelésekor figyelembe vett tényezőket. A referátumok típusai, tartalma, terjedelme, megírásának stílusa, a referátumok átfogó szempontjai, olvashatóság, az olvashatóság tesztelése, az ANSI (American National Standards Institute) szabványainak megfelelés és az informativitás a tárgyalt témák közé tartozik. (12 hivatkozás) (KRN)

```
DIALOG(R)File 1:(c) format only 2000 The Dialog Corporation. All rts.
reserv.

Quality of Abstracts.

Reviews the factors to be considered in evaluating abstracts. Types of
abstracts, content, length, writing style, comprehensive criteria for
abstracts, readability, readability testing, compatibility with ANSI
(American National Standards Institute) standards, and informativeness are
among the topics covered. (12 references) (KRN)
```

11.24. ábra: Referátum az ISA-ból

Egy tényező van, amely megkülönbözteti egymástól az indexeket, s fontos szempont lehet referátumaik minőségének értékelésekor. Egy referátum minőségének megítéléséhez meg kell érteni, mitől lesz jó egy referátum. A „jó” megítélése mindig kockázatos, de a referátumok esetében van némi egyetértés a szakértők között. Az American National Standards Institute (ANSI) 1979-ben kidolgozott egy szabványt a referátumokról (jelenleg vizsgálják átdolgozás érdekében), amely sokféleképp meghatározza, milyennek kell lennie egy jó referátumnak. Az ANSI meghatározása szerint a referátum „egy dokumentum tartalmának rövidített, pontos reprezentációja”.

```
DIALOG(R)File 2:(c) 2000 Institution of Electrical Engineers. All rts.
reserv.

Title: Quality of abstracts

Abstract: One factor that differentiates between indexes and could be an
important consideration in evaluation is the quality of their abstracts. To
judge the quality of abstracts, it is necessary to understand what makes a
good abstract. Judgments of 'good' are always dangerous, but, in the case
of abstracts, there is some consensus among the experts. The American
National Standards Institute (ANSI) developed a standard for abstracts in
1979 (it is currently being reviewed for reaffirmation) that in many ways
dictates what a good abstract should be. ANSI defines an abstract as 'an
abbreviated, accurate representation of the contents of a document'.
```

11.25. ábra: Referátum az INSPEC-ből

A „What Is in an (Up)date?” cikk

A „What Is in an (up)date” cikkről újra az ABI/INFORM-ban található a legalaposabb (bár nem tökéletes) referátum (11.26. ábra). Jól foglalja össze a primer dokumentum megjelenési ideje és rekordjának egy adatbázisba való bekerülése közti időbeli elmaradás legfontosabb okait. Világosan meghatározza a frissesség tesztelésének három technikáját, de még csak nem is utal az illusztrációként használt példák nagy számára, ugyanakkor szükségtelenül pazarolja a helyet az eredeti cikk egy mellékes megjegyzésére („End users are less interested...”, A végfelhasználókat kevésbé érdekli...)

Számos akadály lehet az oka annak, hogy egy új közlemény miért kerül be megkésve egy adatbázisba. A késések okai közé tartozik az, amikor a forráspublikációkat késve jelentetik meg, egy referáló-indexelő szolgáltatás hátraléka, a CD-ROM-gyártó üzem túlvállalja magát, és a postai szolgáltatások problémái. A végfelhasználókat kevésbé érdeklik az okok, mint a tény: a tipikus vagy átlagos idő, ami a publikáció megjelenésének dátuma és az az alapján készült rekordnak az adatbázisban való kereshetősége között eltelik. Minél kisebb az időkülönbség, annál frissebb az adatbázis. A frissesség hagyományos mutatói a publikáció frissességét hónapokban mérik, összehasonlítva a forrásdokumentum megjelenésének dátumát és a referáló-indexelő kiadvány megjelenésének dátumát. Ezek a dátumok azonban nem biztosan egyértelműek, és torzíthatják az eredményeket, ha fenntartások nélkül használják őket. A frissesség tesztelésének technikái közé tartoznak a következők: 1. a megjelenés dátuma kontra az aktualizálás dátuma technikája, 2. az aktualizálás dátuma kontra a megjelenés dátuma technikája, és 3. az azonosító szám kontra a megjelenés dátuma technikája. Az online adatbázisok frissebbek, mint CD-ROM-os megfelelőik.

```
DIALOG(R)File 15:(c) 2000 Bell & Howell. All rts. reserv.

What Is in a(n) (Up)Date? Currency Test Searching of Databases

ABSTRACT: There can be a number of obstacles to delay the entry of current citations in a database. Delays include source publications published late, a backlog in the abstracting-indexing service, overcommitment by the CD-ROM manufacturing plant, and postal service problems. End users are less interested in the reasons than in the fact: the typical or average travel time between the date of publication and the date of availability of the related record for searching. The shorter the time gap, the more current the database. Traditional currency indicators measure the currency of publication in terms of months, comparing the date of issue of the source document and the date of issue of the abstracting-indexing publication. However, these dates may be ambiguous and may distort the results if used without reservation. Techniques for currency testing include: 1. the publication date versus update date technique, 2. the update date versus publication date technique, and 3. the accession number versus publication date technique. Online databases are more current than their CD-ROM siblings.
```

11.26. ábra: Referátum az ABI/INFORM-ból

A LISA referátuma (11.27. ábra) dicséretet érdemel azért, mert megmutatja a cikk alapvető szempontjait - hogy a technikák lehetővé teszik rekordok ezreinek vagy akár egy egész adatbázisnak a tesztelését. A negatív oldalhoz tartozik az, hogy a két mondat a referátum közepén bosszantóan redundáns („Describes techniques that can be used for currency measurement. 3 basic currency measurement techniques are described.”; Leír technikákat, amelyek használhatók a frissesség tesztelésére. 3 alapvető frissességet mérő technika leírása.)

Az online és CD-ROM keresőszoftverek kiváló lehetőségeket biztosítanak az adatbázisok frissességének, a források megjelenése és az adatbázisba való bekerülés időkülönbségének mérésére, rekordok ezreire vagy akár az adatbázis teljes tartalmára vonatkozóan. Bemutat technikákat, amelyek használhatók a frissesség mérésére. Három alapvető frissességet mérő technikát mutat be: A megjelenés dátuma kontra az aktualizálás dátuma; az aktualizálás dátuma kontra a megjelenés dátuma; és a raktári jelzet kontra a megjelenés dátuma. Megmutatja, hogy néhány év frissességi tesztjeit hogyan lehet történeti perspektívába állítani, hogy egy adatbázis frissességének jellemzői hogyan változtak meg. Megjegyzí, hogy egy adatbázis mért frissességére milyen hatással lehet a fájl aktualizálásának rendszeressége.

```

DIALOG(R)File 61:(c) 2000 Reed Reference Publishing. All rts. reserv.

What is in a (n) (up) date? Currency test searching of databases.

ABSTRACT: The on-line and CD-ROM search software offers excellent
capabilities for performing measurements of a data base's currency and
the time-gaps between publication of the sources and appearance in the
data base, involving thousands of records or even the entire content of
the data base. Describes techniques that can be used for currency
measurement. 3 basic currency measurement techniques are described.
Publication Date versus Update Date; Update Date versus Publication Date;
and Accession Number versus Publication Date. Shows how currency test
searches accross a few years time span can yield a historic perspective
of how the currency profile of a data base has changed. Notes that
measured currency of a data base can be affected by the pattern and
regularity of the updating schedule for that file. N.L.M.

```

11.27. ábra: Referátum a LISA-ból

Az ISA referátumának (11.28. ábra) pozitív összetevője az, hogy egyértelművé teszi, hogy a cikk a frissesség értékelését mutatja be. A Trade & Industry Database (11.29. ábra) jó összefoglalást ad a frissesség tesztelésének technikáiról, de nem írja le az összefüggéseket. A Periodicals Abstracts PlusText referátuma (11.30. ábra) csak két mondatból áll, és a második csupán a cikk egy megjegyzésének a parafrázisa. Az H. W. Wilson cég Business Abstracts adatbázisának referátuma (11.31. ábra) nagyon elszomorító ettől a fájlkészítőtől, amely általában a legjobb referátumokat állítja elő. Nem csak arról van szó, hogy a referátum nagyon rövid, hanem arról is, hogy az egyetlen mondat nem ad semmi olyan újat, amit nem tudhatnánk a cím és a címkiegészítés mezőkből.

A cikk az adatbázisok frissességét vizsgáló tesztkeresések kérdését vizsgálja. Azt állítja, hogy az online és CD-ROM-os keresőszoftverek kiváló lehetőségeket biztosíthatnak nagyméretű frissességi tesztkeresések elvégzéséhez. A specifikus összehasonlítási technikák közé tartoznak: a megjelenés dátuma kontra az aktualizálás dátuma technikája; az aktualizálás dátuma kontra a megjelenés dátuma technikája; azonosító szám kontra a megjelenés dátuma technikája. A frissesség jellemzőinek értékelését is bemutatják.

```

DIALOG(R)File 202:(c) Information Today, Inc. All rts. reserv.

What is a(n) (up)date? Currency test searching of databases.

This article examines the issue of currency test searching of
databases. It argues that online and CD-ROM search software may offer
excellent capabilities for doing wide-scale currency test searches.
Specific comparisons of techniques include: publication date versus
update date technique; update versus publication date technique;
accession number versus publication date technique. Currency profile
evaluations are also presented.

```

11.28. ábra: Referátum az ISA-ból

A bibliográfiai adatbázisok frissességét számos módszerrel lehet tesztelni. A megjelenés dátumának összehasonlítása rekordok adott halmazában az aktualizálás dátumával ugyanarra a halmazra vonatkozóan némi jelzést ad a feldolgozás időbeli lemaradásáról. Az ezzel ellentétes módszer, amely az aktualizálás dátumát vizsgálja, némi jelzést ad arra, hogy egy adott aktualizáláson belül hogyan oszlanak meg a különböző korú rekordok. A rekord azonosító számának összehasonlítása a megjelenés dátumával lehetővé teszi a frissesség ellenőrzése ott, ahol az aktualizálási dátumok nem elérhetőek.


```
DIALOG(R)File 148:(c)2000 The Gale Group. All rts. reserv.

What is in a(n) (up)date? Currency test searching of databases.

ABSTRACT: The currency of bibliographic databases can be tested using a
number of methods. Comparison of the publication date for a given set of
records with the update date for those records gives some indication of the
processing timelag. The converse method of comparing update date with
publication date offers some indication of the distribution of records of
different ages within a given update. Comparison of record accession
numbers with publication date allows for currency checking where update
dates are not available.
```

**11.29. ábra: Referátum a Trade & Industry
Database-ből**

Bemutatja és megtárgyalja az online adatbázisok frissességet tesztelő, nagyméretű keresésekre vonatkozó vizsgálat eredményeit. A végfelhasználókat jobban érdekli a kiadvány megjelenésének dátuma és a keresésre való elérhetőség közötti tipikus és átlagos idő.

```
DIALOG(R)File 484:(c) 2000 Bell & Howell. All rts. reserv.

What Is in a(n) (Up)date?--Currency Test Searching of Databases

ABSTRACT: The findings of a study to determine widescale currency test
searches of online databases are presented and discussed. End-users are
more interested in the typical or average travel time between the date of
publication and the date of searching availability.
```

**11.30. ábra: Referátum a Periodical Abstracts
PlusTextből**

Online és CD-ROM adatbázisokban szereplő publikációk frissességét értékelő technikákat mutat be.

```
DIALOG(R)File 553:(c) 2000 The HW Wilson Co. All rts. reserv.

What is in a(n) (up)date? Currency test searching of databases.
AUGMENTED TITLE: online and CD-ROM

ABSTRACT: Techniques for evaluating the currency of publications in online
and CD-ROM databases are presented.
```

**11.31. ábra: Referátum az H. W. Wilson cég
Business Abstractséből**

A „Data Transfer Capabilities” cikk

A „Data Transfer Capabilities” cikk legjobb referátuma az ABI/INFORM-ban található (11.32. ábra). Ez a leginformatívabb valamennyi közül, és az egyetlen, amely világosan meghatározza a hét output és letöltési követelményt. Nagyon jól összefoglalja a cikk lényegét is.

A legtöbb CD-ROM-szoftver kifinomult keresési technikákat ajánl, de gondjai vannak, amikor átfogó output lehetőségeket kell ajánlani. Ez tipikus lehet az online adatbankok esetében, amelyek szintén a keresési jellemzőket hangsúlyozzák az adatátvitel helyett. Az adatátviteli képességek vizsgálata áttekinti az output és a letöltés követelményeit. Az eredmények a merevlemezen vagy a hajlékony lemezen való tárolására letöltésként utalnak. Ennek a fájlnak a host programra való átvitelét feltöltésnek nevezik. Az output és a letöltés követelményeit négy kategóriában lehet összefoglalni: 1. az output rendeltetési helyének meghatározása, 2. az átvivendő rekordok megjelölése, 3. előredefiniált és megjelenítési nyomtatási formátumok, 4. a használó által meghatározott megjelenítési és nyomtatási formátumok, 5. az output rekordjainak sorba rendezése, 6. az oldal formázása, 7. az átvitel formáinak választéka. A CD-ROM-termékek jelentős különbségeket mutatnak ezekben a kategóriákban. Ennek eredményeként a specifikus letöltési jellemzők fontos tényezők lehetnek a CD-ROM-vásárlók döntésekor. Az átviteli lehetőségek a CD-ROM adatbázisból letöltött output szerkezetének formájától függ.

DIALOG(R)File 15:(c) 2000 Bell & Howell. All rts. reserv.

Data Transfer Capabilities of CD-ROM Software: Part I

ABSTRACT: Most CD-ROM software offers sophisticated search techniques but falls short when it comes to offering extensive output capabilities. This may be typical of the online databanks, which also emphasize search capabilities instead of data transfer. A study of data transfer capabilities reviews the scope of output and downloading requirements. Storing results on a hard disk or a floppy diskette is referred to as downloading. The transferring of this file into the host program is called uploading. Output and downloading requirements can be grouped into 7 categories: 1. specification of output destination, 2. marking of records to be transferred, 3. predefined and display print formats, 4. user-specified display and print formats, 5. sorting of output records, 6. page formatting, and 7. choice of transfer formats. CD-ROM products show significant differences within these categories. As a result, the specific downloading features may be an important factor in the purchasing decisions of CD-ROM buyers. Transfer options depend on the structural format of the output downloaded from a CD-ROM database.

11.32. ábra: Referátum az ABI/INFORM-ból

Az ISA-ban található a második legjobb referátum (11.33. ábra), amely részleteket emel ki az eredeti szövegből, ami nagyon általános gyakorlat, s azokat összefűzi. Az eredeti szöveghez képest az egyetlen apró módosítás az első mondat: „but suffers from limited ability to offer extensive output capabilities”, ezt a következővel helyettesíti: „but falls short when it comes to offering extensive output capabilities”, bármilyen látható indok nélkül.

A legtöbb CD-ROM-szoftver kifinomult keresési technikákat ajánl, de szenved attól, hogy korlátozottak a képességei arra, hogy átfogó output lehetőségeket ajánljon. A CD-ROM adatbázisokban és az egyéb helyi rendszerekben nincsenek közvetlenül a rekordok sorba rendezéséhez, formázásához, megjelenítéséhez vagy kinyomtatásához kapcsolódó költségek. Az ilyen termékeknél elvárható lenne, hogy hatékonyabb output lehetőségeket ajánljanak, azonban nagyon kevés CD-ROM-termék ajánl megfelelő lehetőségeket a keresés eredményének formázásához. Legalább azt lehetne elvárni egy CD-ROM-terméktől, hogy egy keresés eredménye hozzáférhető legyen olyan adatsere-formátumokban, amelyek alkalmasak arra, hogy szövegszerkesztőbe, szövegkezelőbe, táblázatkezelőbe és grafikai szoftverbe importálják. Ez lehetővé tenné a keresés eredményének a megfelelő host programhoz való eljuttatását további feldolgozásra. Az output tartalmát és a letöltés követelményeit vizsgálják, a CD-ROM adatbázisokból való letöltés technikai szempontjaira koncentrálna, egy olyan folyamatra, amely elmentés, átvitel, megtartás, exportálás vagy másolás néven ismert.

DIALOG(R)File 202:(c) Information Today, Inc. All rts. reserv.

Data transfer capabilities of CD-ROM software. Part I.

Most CD-ROM software offers sophisticated search techniques, but suffers from limited ability to offer extensive output capabilities. In CD-ROM databases and other local systems, there are no direct associated costs to sorting, formatting, displaying or printing records. For such products, it would be reasonable to expect more powerful output options, yet very few CD-ROM products offer appropriate facilities for formatting the output of a search. The least one should expect from any CD-ROM product is that the output of a search be available in data exchange formats that are appropriate for importing into word processors, text management, spreadsheet, database management, and graphics software. This would allow the transfer of the search output to an appropriate host program for further processing. The scope of output and downloading requirements is examined, focusing on the technical aspects of downloading from CD-ROM databases, a process also known as saving, transferring, keeping, exporting, or copying.

11.33. ábra: Az ISA referátuma

Az INSPEC referátuma (11.34. ábra) közel áll az ISA-éhoz és hasonló összefűzési technikát alkalmaz, de az eredeti szöveg más bekezdéseiből. A LISA két mondatos referátuma elég szegényes, alig ad valamit a címben már szereplő információhoz (11.35. ábra). Furcsa, hogy több szót szentel a cikk második részét tárgyaló résznek, mint az elsőnek.

A legkevesebb, amit el lehet várni egy CD-ROM adatbázistól az, hogy a keresés eredménye elérhető legyen olyan adatcsere-formátumokban, amelyek megfelelőek szövegszerkesztőbe, szövegkezelőbe, táblázatkezelőbe, adatbázis-kezelőbe és grafikai szoftverbe való átvitelre. Ez lehetővé tenné, hogy a keresés eredményét eljuttassák a megfelelő host programhoz további feldolgozásra. A tanulmány áttekinti az output tartalmát és a letöltés követelményeit. Az output összes korlátját le lehet győzni, ha képesek a keresés rekordjait feltölteni a megfelelő host programba. Az átvitel lehetőségei a CD-ROM adatbázisból letöltött output szerkezeti formájától függ.

```
DIALOG(R)File 2:(c) 2000 Institution of Electrical Engineers. All rts.
reserv.

Title: Data transfer capabilities of CD-ROM software. I
Abstract: The least one should expect from any CD-ROM product is that the
output of a search be available in data exchange formats which are
appropriate for importing into word processors, text management,
spreadsheet, database management and graphics software. This would allow
the transfer of the search output to an appropriate host program for
further processing. This study reviews the scope of output and downloading
requirements. All of the output restrictions discussed can be overcome if
you are able to upload the records of the search into an appropriate host
program. The transfer options depend on the structural format of the output
downloaded from a CD-ROM database.
```

11.34. ábra: Referátum az INSPEC-ből

Áttekinti a CD-ROM adatbázis szoftverek output és letöltési követelményeinek hatókörét. A 2. rész (az 1991 márciusi számban) a potenciális host programok által az adatok importálására legáltalánosabban használt csereformátumokat fogja megtárgyalni és illusztrálni. N. L. M.

```
DIALOG(R)File 61:(c) 2000 Reed Reference Publishing. All rts. reserv.

Data transfer capabilities of CD-ROM software. Part 1.

ABSTRACT: Reviews the scope of output and downloading requirements of
CD-ROM data base software. Part 2 (March 91 issue) will discuss and
illustrate the exchange formats most commonly used by potential host
programs for importing data. N.L.M.
```

11.35. ábra: Referátum a LISA-ból

A cikk második részét újra csak az ABI/INFORM referálta legjobban, amelyet elegendő részlet és jó szerkezet jellemez. Az INSPEC referátuma (11.37. ábra) nem említi meg a speciális formátumokat és nem utal arra, hogy mind a szöveges, mind a grafikus csereformátumokat tárgyalja a cikk. A négy csereformátum - amelyeket nem nevez meg a referátum - csak a szöveges fájlokra illik, a cikk sajátos grafikus formátumokat is tárgyal.

Sok CD-ROM program output lehetőségeinek korlátai nehezé teszik azt, hogy a keresés eredményeit közvetlenül a szoftverbe vagy host programba vigyék át. Bár az lenne a legjobb, ha a CD-ROM keresőszoftverek jobb veleszületett letöltési és adatátviteli képességekkel rendelkeznenek, ki lehet használni az egyéb szoftverek átviteli lehetőségeit a CD-ROM-os keresés eredményeinek szerkesztésére és további feldolgozására. Van néhány olyan szabványosított formátum, amelyet megért a legtöbb fontosabb program, s így hidat képeznek a programok között. A négy legáltalánosabb formátum a következő: 1. szerkesztetlen vagy sima ASCII, 2. kötött mező formátum, 3. körülhatárolt ASCII formátum, és 4. MARC (géppel olvasható katalogizálás) kommunikációs forma. A legjobb grafikus fájlkonverziós program, a Hijaak 30 formátumot sorol fel. Szerencsére vannak széles körben használt formátumok, s a legtöbb CD-ROM-terméket, amely ajánl grafikus formát, képes átvinni képeket, rajzokat, clip-artokat és line-artokat ezeknek a ténylegesen szabványos formátumoknak valamelyikébe. A legismertebb és legáltalánosabban használt formátum a PCX, amelyet a PC Paintbrush csomag tett népszerűvé.

DIALOG(R)File 15:(c) 2000 Bell & Howell. All rts. reserv.

Data Transfer Capabilities of CD-ROM Software, Part II

ABSTRACT: The limitations in output capabilities of many CD-ROM programs make it difficult to transfer the results of the search from a CD-ROM directly into the software or host program. Although it would be ideal to have better downloading and data transfer capabilities inherent in the CD-ROM retrieval software, one can use the transfer capabilities of other software programs to edit and further process the results of the CD-ROM search. There are a few standard formats that are understood by all the mainstream programs and thus provide a bridge among these programs. The 4 most common formats are: 1. unstructured or plain ASCII, 2. fixed field format, 3. delimited ASCII format, and 4. MARC (machine readable cataloging) communications format. The best graphic file conversion program, Mijack, lists 30 formats. Fortunately, there are some widely used formats, and most of the CD-ROM products that offer graphics are capable of transferring the images, drawings, clip-art, and line-art in one of these de facto standard formats. The simplest and most commonly accepted format is the PCX format made popular by the PC Paintbrush package.

11.36. ábra: Referátum az ABI/INFORM-ból

Az I. részt lásd ugyanitt, 4. köt.1. sz. 63-69.p. (1991). Az első rész a CD-ROM szoftverek output és letöltési lehetőségeit tárgyalta. Itt a szerző azt illusztrálja, hogy csere/átviteli formátumoknak milyen szerkezeti lehetnek használhatóak a rekordoknak a potenciális szoftver host programokba való feltöltésére. A szerző megvizsgálja a CD-ROM-okról származó mozgó számokat, szövegeket és képeket. Van néhány szabványosított forma, amelyet megértenek a legfőbb programok, s ezzel hidat képeznek a programok között. A szerző a négy legáltalánosabb formátumra koncentrál, valamint arra, hogy a legfontosabb CD-ROM-termékek hogyan képesek outputot szolgáltatni ezekben a formátumokban.

DIALOG(R)File 2:(c) 2000 Institution of Electrical Engineers. All rts. reserv.

Title: Data transfer capabilities of CD-ROM software. II

Abstract: For pt.I see ibid., vol.4, no.1, p63-9 (1991). In the first part, the output and downloading options of CD-ROM software were discussed. Here the author illustrates what exchange/transfer format structures may be used to upload records into potential software host programs. The author looks at moving numbers, text and images from CD-ROMs. There are a few standard formats which are understood by all the mainstream programs and thus provide a bridge among these programs. The author focuses on the four most common formats and the capability of the major CD-ROM products to produce outputs in these formats.

11.37. ábra: Referátum az INSPEC-ből

Az ISA-ban két referátum található, az első elfogadható (11.38. ábra), de sokkal gyengébb, mint a második (11.39. ábra), amelyet az ERIC adatbázisból vettek át. Bár ez rövidebb, mint az ABI/INFORM referátuma, a részleteket különösen jó elrendezésben szolgáltatja.

A CD-ROM programok output képességeinek korlátait tárgyalja. CD-ROM-okról származó mozgó számokat, szövegeket és képeket, például szövegszerkesztőt, adatbázis-kezelőt és táblázatkezelőt mutat be. Az alapvető ASCII használatának előnyeit tárgyalja. Grafikák átvitelét lehetővé tevő programokat kutat.

DIALOG(R)File 202:(c) Information Today, Inc. All rts. reserv.

Data transfer capabilities of CD-ROM software, Part II.

Limitations in output capabilities of CD-ROM programs are discussed. Moving numbers, text and images from CD-ROM, e.g., word processing, database management, and spreadsheets is described. The advantages of using basic ASCII are discussed. Programs that enable the transfer of graphics are explored.

11.38. ábra: Az első referátum az ISA-ból

Leír és összehasonlít számos csere/átviteli formát, amelyeket rekordok szoftver host programokba való feltöltésére használnak, köztük (1) szerkesztetlen vagy sima ASCII; (2) rögzített mező; (3) körülhatárolt ASCII; (4)tagolt; (5) MARC (Machine Readable Cataloging) és (6) grafikus. Arra a következtetésre jut, hogy a CD-ROM-készítőknek erősíteniük kell a programok beépített output kezelőit, egynél több csereformátumban kell letöltési lehetőségeket ajánlaniuk. (A referátum forrása: ERIC)

```

DIALOG(R)File 202:(c) Information Today, Inc. All rts. reserv.

Data transfer capabilities of CD-ROM software, part II.
Describes and compares several exchange/transfer formats used to
upload records into software host programs, including: (1) unstructured
or plain ASCII; (2) fixed field; (3) delimited ASCII; (4) tagged; (5)
MARC (Machine Readable Cataloging); and (6) graphic. It is concluded
that CD-ROM producers should strengthen the built-in output management
of programs and offer download options in more than one exchange
format.. (Abstract Source: ERIC)

```

11.39. ábra: A második, az ERIC-ből importált referátum az ISA-ból.

A LISA kétmondatos referátuma (11.40. ábra) megint alig informatív és van benne egy sajtóhiba is („ooptions”). Furcsa módon a Trade & Industry Database-ben a cikk egyik részéhez kapcsolódóan sem volt referátum. Különös, hogy az ERIC-ben csak a cikk 2. részéről volt rekord (a 11.39. ábra mutatja) - ez azonban a cikkeknek a forrásdokumentumokból való kiválasztásának problémájához kapcsolódik, s nem a referálás minőségének kérdéséhez.

Az 1. rész a CD-ROM szoftverek output és letöltési lehetőségeivel foglalkozott. A 2. rész azt illusztrálja, hogy milyen csere- és átviteli formátumszerkezetek használhatók a rekordok feltöltésére potenciális szoftver host programokba. N. L. M.

```

DIALOG(R)File 61:(c) 2000 Read Reference Publishing. All rts. reserv.

Data transfer capabilities of CD-ROM software Part 2.

ABSTRACT: Part 1 considered the output and downloading ooptions of CD-ROM
software. Part 2 illustrates what exchange and transfer format structures
may be used to upload records into potential software host programs.
N.L.M.

```

11.40. ábra: Referátum a LISA-ból

A referálás gyakorlata időről időre változik. Van Camp (1993) jelentette, hogy a CINAHL adatbázis csak az ápolás magfolyóiratainak cikkeiről tartalmazott referátumokat, s azt is csak akkor, ha a referátum megtalálható volt az eredeti cikkhez kapcsolódóan; ebből következően csak a rekordok 17%-ában volt referátum. Ez az arány azonban jelentős mértékben javult, amikor a CINAHL az összes lehetséges cikkből kezdte felvenni a szerzői referátumokat.

Ez jó emlékeztető arra, hogy a referátumok minősége jelentős mértékben javulhat, de a referátumokat sohasem javítják ki visszamenőlegesen, ahogyan ez más adatelemekkel megtörténhet. A referátumok minőségének értékelése nagyon időigényes művelet, így a mini minták sokkal valószínűbbek, mint a széles körű értékelések. A kis minták használata esetleg nem igazságos egy adatbázis esetében, amely legtöbbször alapos és informatív referátumokat ad.

12. Az árral kapcsolatos megfontolások

A régi mondás: „Azt kapod, amiért fizetsz” gyakran nem érvényes az adatbázisokra. Sokszor meg lehet kapni ugyanazt a tartalmat olcsóbban vagy akár ingyen is. Az, hogy mit fizet a használó, a cégek vagy főiskolák, attól a megállapodástól függ, amit az információs szolgáltatással kötöttek. A díjak és a díjszabási rendszerek természetesen állandóan változnak, s még adott pillanatban sem azonosak minden ügyfél számára. A cégek, oktatási intézmények és egyének számára meghatározott díjak nagyban különbözhetnek. Ugyanazon a kategórián belül a díjak különbözőek lehetnek az alkalmazottak vagy a diákok teljes számától, vagy az adatbázishoz egyszerre hozzáférő, párhuzamos használók számától függően.

A díjak alapulhatnak a tranzakciók mennyiségén (keresőkérdeések, a keresőkérdeések által használt erőforrások, a megjelenített, elmentett vagy kinyomtatott rekordok száma és formátuma). Ez a használat mértéke szerinti fizetési módként ismert. Ennek alternatívájaként korlátlan használatot biztosító, egységes díjat meghatározó megállapodás is köthető. Ugyanannak az információs szolgáltatónak számos díjszabási módja lehet az ügyfelek sokféle típusához kapcsolódóan. A költségek összehasonlítása összetett feladat.

A különböző szolgáltatók által kínált változatos díjszabási rendszerek vizsgálatán és összehasonlításán túl értékelni kell a tartalmat és a szoftver szempontjából a hozzáférés hatékonyságát, könnyűségét, kényelmét is. Még akkor is, ha egy adatbázis tartalma azonos két szolgáltatónál, ha az egyik szolgáltatónak gyatra szoftvere van, nagyobb költséget jelenthet a végfelhasználónak az, hogy több időt tölt a szoftverrel való küszködéssel, vagy egyszerűen képtelen bizonyos típusú kereséseket elvégezni, amelyek a valamivel drágább másik szolgáltató szoftverével azonnal rendelkezésre állnak.

Senki sem kételkedik abban, hogy a DIALOG szoftvere összességében sokkal jobb, mint sok kormányzati adatbázis ingyenes webes változatának legtöbb keresőprogramja. Senki sem kérdőjelezi meg, hogy az információs szakember számára, aki az adatbázisok kereséséből él, hasznosak ezek a hatékony opciók. De vajon az átlagos felhasználónak igazán szükségük van azokra a legfejlettebb módszerekre, amelyeket a DIALOG az Explode, Rank vagy Map parancsokkal felajánl? Egy cég használói kihasználnák azt a lehetőséget, hogy a DIALOG-nál több adatbázisban kereshetnek egyszerre? Megéri ez a plusz költséget? Az egyszerűség, a természetes nyelven való keresés lehetőségei, a kérdések helyesírásának ellenőrzése, a használók kalauzolása a kontrollált szótár fogalmai között fontosabb jellemzők lehetnek, mint azok, amelyeket a szuperkeresők behunyt szemmel is használnak és megkövetelnek.

Hasonló tényezők lehetnek érvényesek az adatbázisok tartalmára, amikor összehasonlítjuk az alternatívákat és azok költségkihatásait. Gyakran olcsóbban vagy akár ingyen is hozzáférhetünk egy adatbázishoz, bár talán csak egy korlátozott részhalmazához. A korlátozás vonatkozhat az időszakra, dokumentumtípus(ok)ra, nyelvre vagy ezek kombinációjára.

Különbségek a tartalomban

A könyvtárosok alig tudnák elképzelni, hogy nyomtatott formában, CD-ROM-on vagy online módon ne férjenek hozzá valamelyik folyóirat címjegyzékhez. Most, hogy a PubList ingyen érhető el a Weben, ideje újra átgondolni a lehetőségeket. A PubList az Ulrich's International Periodicals Directory részhalmaz. Nincsenek benne rekordok a megszűnt címekről, s nincs meg benne minden adatelem, amelyeket az Ulrich's rekordja tartalmaz, de mégiscsak nyújtja a leginkább szükséges információkat (12.1. ábra), az eredeti fájl hibáival együtt.

Mennyire fontos tudni a megszűnt folyóiratokról egy adott könyvtár gyakorlatában? A CODEN-ről? A Dewey osztályozási jelzetről? A dokumentumszolgáltatókról? A British Library raktári jelzetéről? Ezek nem érhetők el a PubList verzióban, csak a pénzes szolgáltatásnál (12.2. ábra). Ugyanígy nem érhető el Kongresszusi Könyvtár jelzete, sem azon indexelő és referáló szolgáltatások neve, amelyek feldolgozzák a folyóiratot. De a Kongresszusi Könyvtár jelzete az Ulrich's pénzes változata esetében is, csak a rekordok kevesebb mint 20%-ában érhető el, amely nem teljes (és ijesztően elavult) információkat ad az indexelő és referáló szolgáltatásokban való feldolgozottságról (12.3. ábra)

PUB LIST

Home Search About Us Help

The premier online global resource for information about print and electronic publications.

EContent

Quick Title Search

Search

Services Available for this Title

Order Articles

Main Publisher Advertising Full Record

Title: Database (Wilton)

SubTitle: the magazine of database reference and review

Publication Type: Trade Publication

Status: Active **Frequency:** Bi-Monthly

ISSN: 0162-4105

Editor: Ed. Marydee Ojala

Registered with CCC: Registered

Online: Available Online

Subject(s) (click on subject to see a listing of other titles on PubList.com in that area):
 • [COMPUTERS \(Database Management / General\)](#)

Circulation (Paid): 4,500

Publisher: [Online, Inc.](#)

Address: 462 Danbury Rd., Wilton, CT 06897-4007

Publisher Homepage: <http://www.onlineinc.com/database> **E-mail:** dbmag@onlineinc.com

Phone: 203-761-1466 **Fax:** 203-761-1444

Advertising: Available

12.1. ábra: Mintarekord a PubListből

EContent

Formerly (until June 1999): Database: ISSN 0162-4105

Status: Active

Publisher: Online, Inc.
 213 Danbury Rd
 Wilton
 CT 06897-4006
Telephone: 203-761-1466, 800-248-8466
E-mail: dbmag@onlineinc.com
URL: <http://www.onlineinc.com/database/>,
<http://www.ecmag.net/>
FAX: 203-761-1444

Editor: Pub. Jeffrey K. Pemberton, Ed. Marydee Ojala

Country of Publication: United States (US)

First Published: 1978; Online - full text edition: 1978

Frequency: bi-monthly

Circulation: 4,500 (paid)

Special Features: Index

Document Type: Trade

Price: \$55 domestic ; \$90 elsewhere; \$65 CANADA (effective 2000)

ISSN: 1525-2531

CODEN: ECONF4

Dewey Decimal: 025.04

Call Number:

LC Call Number: Z699.A1

12.2. ábra: Mintarekord az Ulrich'sből (1. rész)

LC Call Number:	Z699.A1
British Library Shelf Mark:	BLDSC Shelfmark 3659.530425
Document Suppliers:	IDS, Ei, CISTI, CINDOC, CASDDS, AskIEEE
Availability:	Also available online
Online:	
Online Vendors:	Bell & Howell Information & Learning, Gale Group
Abstracting and Indexing Services:	Applied Science & Technology Index, Business Periodicals Index, Library Literature, Social Science Citation Index, INSPEC: Computers & Control Abstracts, A B I - INFORM, Current Index to Journals in Education, Computer Contents, Compumath Citation Index, Cumulative Index to Nursing and Allied Health Literature, Current Contents, Consumers Index, Key to Economic Science, Legal Information Management Index, Library Hi Tech News, Microcomputer Index, Trade & Industry Index, Library & Information Science Abstracts, Legal Resource Index, SoftBase, WPM
Listed in:	Ulrich's International Periodicals Directory Scientific & Technical Books & Serials in Print
Subject Headings:	COMPUTERS (0000621X); DATA BASE MANAGEMENT (0000621X)
Notes:	Features articles on a variety of topics of interest to online database users, and electronic content providers and users. Includes database search aids. back issues avail. Text in English Copyright Clearance Center
Rights & Permissions:	Jenny Pemberton

12.3. ábra: Mintarekord az Ulrich'sból (2. rész)

Nem említi meg például azokat az adatbázisokat, amelyek jól feldolgozzák az *EContent* című folyóiratot: Business & Industry, Business & Management Practices, az H. W. Wilson cég Applied Science & Technology adatbázisa, PASCAL, Information Science Abstracts, PIRA, Periodical Abstracts PlusText vagy az H. W. Wilson cég Business Abstracts adatbázisa (12.4. ábra). Megemlíti a Microcomputer Indexet, amely sok évvel korábban Microcomputer Abstractsre változtatta címét (s valóban nagyon informatív referátumokat tartalmaz), majd 2000 elején Internet & Personal Computing Abstractsre keresztelték át.

Set	Term Searched	Items	File
S1	JN=ECONTENT?	889	
S1	JN=ECONTENT?	89	553
S1	JN=ECONTENT?	107	484
S1	JN=ECONTENT?	203	440
S1	JN=ECONTENT?	14	248
S1	JN=ECONTENT?	132	202
S1	JN=ECONTENT?	28	144
S1	JN=ECONTENT?	43	211
S1	JN=ECONTENT?	109	99
S1	JN=ECONTENT?	77	13
S1	JN=ECONTENT?	87	9
Show Database Details for:			
[9: Business & Industry(TM)			

12.4. ábra: Próbakeresés az *EContent* feldolgozottságáról

Vajon ezek az extra és gyakran nagyon hiányos és pontatlan adatelemek megérik a CD-ROM-változat csaknem 1000 dolláros előfizetési díját, amely csak akkor érhető el, amikor a könyvtár nyitva van, vagy amikor a tájékoztató pult működik, ahol valószínűleg még korlátozottabb időben lehet megkapni a CD-ROM-ot az olvasójegy ellenében? Ami rossz a PubList-ben, az rossz az Ulrich's pénzért szolgáltatott változataiban is, ilyen például a szerkesztő neve, amely Ojala, nem pedig Ofala.

A PubList szoftvere nem rendelkezik mindazokkal a jellemzőkkel, mint az Ovid, DIALOG vagy Lexis-Nexis, de jobb a tartalom megjelenítésének formája, hatékonyabb az eredmények listájának formája, s olyan ugrópontjai vannak a tárgyi, e-mail és URL mezőkből, amelyekre például a DIALOG még mindig nem képes.

A szoftverek különbségei

Egyre több kormányzati adatbázis érhető el ingyen a weben. Ezek megkérdőjelezzik az értelmét a kereskedelmi szolgáltatótól való változat előfizetésének, különösen akkor, ha a tartalom ugyanaz, a szoftver pedig sokkal jobb, mint egyik vagy másik kereskedelmi szolgáltatóé.

Vegyük például az ERIC adatbázist. Egy tucatnyi kereskedelmi rendszeren elérhető, s számos ingyenes változata van a weben. Közülük kettőben megvan mind a folyóiratok anyagát tartalmazó részhalmaz (amely az EJ részhalmazként is ismert), mind a kutatási jelentéseket tartalmazó részhalmaz (ED részhalmaz). Miután 1999 közepén mindkét ingyenes változatot olyan módon fejlesztették, hogy a rekordok 1966-ig visszamenőleg megtalálhatók, tartalmuk megegyezik a kereskedelmi információszolgáltatóknál elérhetővel. Valójában a legjobb ingyenes változat az volt, amely az ERIC Assessment and Evaluation Clearinghouse-nál volt elérhető (<http://ericae.net/>), sok fontos, az oktatók számára jól használható dokumentum teljes szövegét is tartalmazva. Az ERIC/AE ezen extráin kívül az egyedüli különbség a szoftverben - és az árban van. (Ahogy már az 1. fejezetben utaltunk rá, az ERIC/AE-t a többi ügynökséggel együtt megszüntették, az ERIC ingyenes változata, a SearchERIC a <http://SearchERIC.org> címen érhető el.) Az ERIC feltehetőleg legrosszabb kereskedelmi változatát a CARL rendszer ajánlja fel több ezer dolláros éves előfizetési díjért - a végfelhasználók számától függően. Valószínűleg az összes közül egyik legjobb az ERIC Assessment and Evaluation Clearinghouse (ERIC/AE) épp említett változata volt. Érdekes, hogy ezt az adófizetők pénzének felhasználása nélkül fejlesztették ki, s ingyenesen szolgáltatták.

Mindkét változat elérhetővé teszi az ERIC teauruszát (ERIC Thesaurus), de a CARL csak külön adatbázisban, míg az ERIC/AE gyönyörűen integrálva a bibliográfiai fájlal. S mi több, ez automatikusan aktivizálódik, amint a keresőkérdésben lévő szavakkal egyezést talál. Ha a használó egy olyan kifejezést ad meg, amely az ERIC-ben nem preferált kifejezés (mint például *managers*), a keresési varázsló felkínálja a preferált kifejezést (*administrators*), s vizuálisan kellemes módon a szélesebb, szűkebb és rokon kifejezéseket, a tárgykörre vonatkozó megjegyzéseket, s a nem preferált kifejezéseket (12.5. ábra).

12.5. ábra: A kifejezések leképezése és a tezaurusz megjelenítése az ERIC/AE-ben.

A használók három lehetőség közül választhatnak a szavak egyeztetésére. A legegyszerűbb a pontos egyezés (*manager*). A második megtalálja a szó szabályos többes számú és birtokos formáját (*managers*, *manager's*). A harmadik a szóból képzett további változatokat is megtalálja (*management*, *managerial*, *managed*, *managing*).

A keresés leszűkíthető a szerzői, cím- vagy deskriptor mezőkre, vagy a számos dokumentumtípus (könyv, konferencia előadás, irodalmi kritika stb.) valamelyikére legördülő menük segítségével. Három előre meghatározott megjelenítési formátum van, és a „Find Similar” (találjon hasonlót) opció automatikusan megkeresi azokat a dokumentumokat, amelyekben a deskriptoroknak ugyanaz a csoportja található meg, mint az eredeti dokumentumban. Szép tulajdonság, hogy a használók további deskriptorokat adhatnak meg és törölhetnek másokat, vagy akár kikerülhetik a teljes folyamatot, a keresőkérdésben szereplő szavak tezaurusz-kifejezésként való keresését.

A hab a tortán az, hogy a rekord azonnal megjeleníthető franciául, olaszul, spanyolul, portugálul és németül. Ez a jellemző nem tökéletes, a számítógépes fordítás érdekes dolgokat produkál, például az „Errors of Commission” helyett „The Errors of the Commission”, de érthető, és a spanyol deskriptorok sokkal jobbak, mint a PASCAL-ban találhatók. A referátumok elegendőek bárki számára, aki nem tud elég jól angolul ahhoz, hogy megértse, miről szól a cikk vagy a jelentés, s a fordítás villámgyors. Az eredmények rövid listája redundáns, s nem lehet egyszerre az egész adatbázisban keresni. Két részre bomlik, 1990 előttire és 1990 utánira, de a legtöbb használó számára az utolsó évtized elegendőnek látszik. Messze lekörözi a CARL által felajánlott változatot, s jobb a többi kereskedelmi rendszernél, kivéve az Ovidot és a National Information Services Corporation (NISC).

A CARL változatban a használóknak először ki kell választaniuk az ERIC Thesaurus adatbázist, abban navigálniuk, a kiválasztott kifejezés(ek)e)t leírni egy papírra, azután átváltani az ERIC bibliográfiai fájlba és beírni a kereső kifejezéseket. A keresési és megjelenítési lehetőségek nem különbek, s csak az ERIC/AE-ből lehetséges közvetlen hozzáférés a többi oktatással kapcsolatos anyag tömegéhez (soknak a teljes szövegéhez).

Az árak anomáliái

Időigényesebb feladat két különböző, de ugyanazt a diszciplínát feldolgozó adatbázis értékét összehasonlítani, de gyakran megéri. Ez különösen igaz a DIALOG-ra, amely megváltoztatta díjszabási módszerét, a kapcsolati idő szerinti fizetés helyett bevezették a DialUnit-ot, amely homályos, és amellyel nehezen megbecsülhető a keresés költsége. A DIALOG megígérte, hogy 2000 végéig visszatér régi díjszabási rendszeréhez, de 2001 elejéig ez nem történt meg.

A DialUnit árak mindenestre használhatók két vagy több DIALOG-beli alternatíva összehasonlítására. Például az ár semmiképpen sem igazodik az értékhez a PsycINFO és a Mental Health Abstracts esetében. A jó minőségű, nagyon átfogó, 1,6 millió rekordos PsycINFO esetében a DialUnit ára 3,25 dollár, míg az MHA 4,00 dollárt fizettet 513 ezer rekordhoz való hozzáférésért, amelyek közül 90% 15 évnél régebbi.

Az MHA adatbázis egy évben kevesebb új rekorddal egészül ki, mint a PsycINFO egy hónapban. Az utolsó, decemberi aktualizálás után 26563 1999-ben megjelent tétel volt a PsycINFOban, míg az MHA-ban csak 1473. Az MHA folyóiratbázisa nem egyszerűen összehasonlíthatatlan a PsycINFOéval, de figyelmen kívül hagyja állítólagos elsődleges szakterületének teljes mértékben alapvető folyóiratait is. 1997-ben az 50 legmagasabb impakt faktorral rendelkező pszichiátriai és pszichológiai folyóirat közül az MHA 4-et dolgozott fel. A PsycINFOban 49 volt megtalálható. Még a DialUnit árak fényében is még mindig lehetnek használók, akik hajlandóak megfizetni az abszurd árat egy ilyen szörnyű termékért, mint az MHA.

Az MHA aktualizálása 2000-ben érte el a mélypontot, 2000 júniusában frissítették utoljára. Kilenc hónap kellett ahhoz, hogy 2001 márciusában a DIALOG bejelentse, az adatbázist lezárják, többé nem aktualizálják. Valószínű, hogy úton van a megszűnés felé, ahogy azt a könyv szerzője korábban megjósolta (Jacsó, 2000b).

A könyvtár- és információtudományi adatbázisok esetében sem igazán értékarányosak a DialUnit árai, s ugyanígy nem azok a megjelenítés árai esetében sem. A LibLit 1 dolláros díja egy teljes rekordra drága egy indexelő rekord esetében, ha a LISA 1,10-es és az ISA 1,20-as áraival hasonlítjuk össze, amelyek referátumot is tartalmazó rekordokra vonatkoznak. A LISA és az ISA megjelenítési árainak különbsége reális, de az ISA 6,00 dolláros DialUnit ára a LISA 3,75 dollárosához képest irreális. Mindkettő rossz üzletnek számít, amikor a kereső olyan folyóiratokban keres cikkeket, mint a *Searcher*, *Online*, *Database*, *EMedia Professional*, *Computers in Libraries* és *Multimedia Schools*, mivel mindegyiküket feldolgozza a Microcomputer Abstracts (most Internet & Personal Computing Abstracts), amely csak 2,50 dolláros díjat számít fel és sokkal kurrensebb, mint a LISA vagy az ISA.

Jobbak-e a referátumai, mint a LISA-ban találhatók? Feltétlenül. Jobbak-e, mint az ISA-éi? Csaknem egyformák. Miért? Mert az ISA 1999 közepe óta a korábban felsorolt folyóiratokban megjelent cikkek rekordjainak többségét a Microcomputer Abstractsból kölcsönözte - néhány hónappal később és néhány adatelemet - mint az URL és a referáló kódja - elhagyva. A Microcomputer Abstracts (Internet & Personal Computing Abstracts) és az ISA tulajdonosa most ugyanaz a cég, az Information Today, de különböző emberek vezetik. Az ISA feltétlenül hasznot húz abból a minőségi, hozzáértő munkából, amely régóta a Microcomputer Abstracts jellemzője, de a két adatbázis árkülönbsége abszurd marad.

Vannak fizetős adatbázisok, amelyek reménytelen helyzetben vannak. A Magill's Survey of Cinema adatbázisban gyakran jó a cselekmény leírása, de még az alacsony, 1,50 dolláros DialUnit ára és a teljes rekord 1,30 dolláros ára is elrettentő lenne, ha az ingyenes, információkban gazdag, varázslatosan megvalósított Internet Movie Database-zel hasonlitanánk össze, még akkor is, ha nem hagyták volna abba az aktualizálását 1997 februárjában. Való-

színűleg ki fogják vonni a DIALOG adatbázisai közül, mivel nincs több esély arra, hogy bevételhez lehessen jutni ebből a termékből.

Azon is csodálkozni lehet, hogy főiskolai és közművelődési könyvtárak miért költenek 520 dollárt a Bowker's Complete Video Directoryra, amikor ez sokkal kevesebbet nyújt tartalom-ban, frissességben, a használat könnyebbségében, mint az IMDb vagy az All-Movie Guide, mindkettő ingyenes webes adatbázis. A Complete Video Directoryban szerepelhet néhány szakképzési és vallásos oktatási film, amely az utóbbi kettőben nincs, de ez nem lehet ok arra, hogy pénzt költsünk egy olyan CD-ROM-ra, amely csak a könyvtár nyitvatartási idejében érhető el (és milyen áron), csak egy munkaállomásról, csüggesztően pontatlan, következtelen, hiányos és elavult információkkal.

Online információs szolgáltatások versenye

Ugyanazon adatfájl különböző változatai árának összehasonlítása fontos tényező az értékelésben. A sok keresést végző használók számára biztosan kifizetődik, ha megtudják, melyik a legjobb üzlet. Az „Azt kapod, amiért fizetsz” mondás nem érvényes azokra a példákra, amelyek Jacsó 1996-os cikkéből származnak (Jacsó, 1996). A korábbi árképzési példát az indokolja, hogy a DIALOG-nak most olyan árképzési rendszere van, a hírhedt DialUnit, amelyet még a DIALOG belső emberei sem tudnak megmagyarázni, s amely összehasonlíthatatlanná teszi azt a többi szolgáltatással.

Viszonylag kis árkülönbségek miatt nem lehet érdemes váltani. A Dissertation Abstracts óránként 60 dollárba került a DIALOG-nál és 50-be az OCLC-nél. Egy olyan alkalmi használó számára, aki a DIALOG-ot ismeri, ez a különbség nem lehet elegendő. Ugyanakkor az Economic Literature Index (EconLit) esetében az árkülönbség már jelentős. A DIALOG 60 dollárt számol fel óránként a kapcsolati időért, míg az OCLC csak 30-at.

A megjelenítési költségeket is figyelembe kell venni az összehasonlításakor. A Dissertation Abstracts esetében az OCLC 0,75 dollárt számolt fel rekordonként, a DIALOG 1,00 dollárt. Az EconLit esetében az OCLC-é volt a magasabb rekordonkénti ár (0,75 dollár, szemben a DIALOG 0,50 dollárával), de ez nem ellensúlyozta azt, hogy az OCLC feleannyit számított fel az adatbázis kapcsolati idejéért. Nem érvényesül az a minta, hogy A szolgáltatás mindig olcsóbb, mint B szolgáltatás. Az OCLC óránként 50 dollárt fizettetett a Microcomputer Abstractsért, míg a DIALOG csak 30-at, a rekordonkénti díj teljesen azonos volt. A PAIS 60 dollár volt az EPIC-nél és 30 dollár a DIALOG-nál, a rekordonkénti díj pedig 0,75 és 0,60 dollár volt.

Az egyes adatbázisok szoftverének jellemzői is fontosak lehetnek, s igazolhatják a magasabb díjakat. Az EconLit esetében azonban nem ez volt az helyzet, mivel egyik szolgáltatás sem tette elérhetővé online módon a tezaurszt. A DIALOG a GeoRef esetében elérhetővé teszi a tezaurszt, s ezzel együtt is olcsóbb volt, mint a tezaursz nélküli OCLC változat (60 és 70 dollár a kapcsolati időért, 0,85 és 0,75 dollár a teljes rekordért).

Csak azért, mert A szolgáltatás jobb, mint B szolgáltatás, még nem kell lemondani arról, hogy még jobb üzletet keressünk, alacsonyabb árakat vagy jobb működést (vagy mindkettőt). A Sociofile olcsóbb volt a DIALOG-nál, mint az OCLC-nél, mind a kapcsolati időt, mind a nyomtatás díját tekintve, annak ellenére, hogy a DIALOG még az adatbázis tezaurszát is hozzáférhetővé tette. Az Ovid azonban még jobb lehetőségeket kínált: 24 dollárt óránként, 0,75 dollárt teljes rekordonként - s ennek volt a legjobb szoftvere is, közte a legjobb tezaursz kezelési lehetőségekkel.

Még jelentősebb a különbség az ABI/INFORM esetében. Ennek sem az OCLC-nél, sem a DIALOG-nál nem elérhető a teaurusza. Az Ovid könnyedén legyőzi őket, teaurusszal (bár van benne pár hiba) (Johnson, 2000) és a legalacsonyabb, 60 dolláros óradíjjal (az OCLC-é 80, a DIALOG-é 90 dollár volt). A teljes rekordok esetében az OCLC-é volt a legalacsonyabb díj (1,50 dollár), de az Ovid 2,00 dolláros díja még mindig jobb volt, mint a DIALOG 2,20 dollárja.

Nagyon gyakran az adatbázisoknak a fájl készítője által készített webes változatai jelentik a legjobb üzletet. Azt is figyelembe kell venni, hogy fájl készítőjének változata lehetővé teszi a tranzakciók szerinti árképzést, míg a más adatbázis kiadók változatai csak fix előfizetési díjakra korlátozódnak, csak azoknak a tagoknak ajánlva fel a tranzakciók szerinti díjat, akik ezen felül éves tagdíjat is fizetnek.

A web forradalmasította az árképzési alternatívákat. Akik idáig fizetős szolgáltatást használtak a nagyközönségnek szóló periodikumok teljes szövegű változatához, ugyanazon cikkek közül nagyon sokat fognak megtalálni a weben kedvezőbb áron. Például az egyéni használók 9,99 dollár havi díjért korlátlan keresési és nyomtatási lehetőséget kapnak az Electric Library gyűjteményéből, amely nagy mértékben azokon a fájlokon alapul, amelyet a nagyközönségnek szóló periodikumok legnagyobb tartalomszolgáltatóinak némelyike készít.

Egész mostanáig senki nem csökkentette az árakat olyan mértékben, mint a Northern Light, amelynek nagy gyűjteménye van magas impakt faktorú, nagy presztízsű folyóiratokból. 1999-ig 1,00 dollárt kértek egy cikkért, amelynek az ára 3-5-ször akkora lett volna hagyományos online szolgáltatóknál, a kapcsolati időn vagy a keresési díjakon felül, a Northern Lightnak ilyen díjai nincsenek (12.6. ábra). Bár a Northern Light 1999-ben emelte árait, azok összehasonlítva még mindig kedvezőbbek, és a szoftverük is kiváló.

2000-ben azonban új szereplők jelentek meg a színen olyan ajánlatokkal, amelyeknek nehéz ellenállni. A Contentville-nél ingyenes a keresés és a referátumok, s a teljes szövegek árai is versenyképesek a Northern Lightéival. Az, hogy a Bell & Howell gyűjteményéből 1,6 millió disszertáció referátumát ingyen szolgáltatja (12.7. ábra), a Contentville-t sok online információszolgáltató félelmetes versenytársává teszi, mivel azok a disszertációk referátumainak kereséséért és megjelenítéséért is fizettetnek. (Meg kell jegyezni, hogy a Contentville csak a referátum első 130 szavát jeleníti meg, de valószínű, hogy a jövőben a teljes referátumot megmutatja majd.) A Find Articles szolgáltatás a Gale Group adatbázisának egy részhalmozán alapul. Mind a keresés, mind a cikkek teljes szövege ingyenes, feltételek nélkül.

A XanEdu szolgáltatás, amelyet 2000-ben vezetett be a Bell & Howell, áttörést jelent az érték/ár csatában. Bár nem ingyenes, úgy tűnik, mintha az lenne, ha tartalmat, a szolgáltatás módját és az előfizetési díjat figyelembe vesszük. 1,3 millió rekordja van különböző Bell & Howell adatbázisokból. Havi 7,50 dolláros, rögzített előfizetési díjért (minimum 4 hónapra kell előfizetni) a használóknak korlátlan keresési és megjelenítési/nyomtatási lehetőségük van 1600 folyóirat, magazin és újság tartalmához. Bár ez a folyóiratbázis sokkal kisebb, mint a Northern Light speciális gyűjteményéé, annak legjavát képviseli.

Az Amerikai Egyesült Államok és az Egyesült Királyság legjobb tudományos folyóiratai, szakmai és a nagyközönségnek szóló magazinjai és legmagasabbra becsült napilapjai elérhetőek a XanEdunál teljes szövegű, szöveg plusz grafika és képi formátumban, épp úgy, mint a teljes ProQuest szolgáltatásnál. A szolgáltatást főiskolai hallgatóknak és oktatóknak szánják, de bárki előfizethet rá. A bevezető ár 2000 szeptemberében 9,99 dollár volt négy hónapra.

16,000 ways to love mom
free shipping & free gift wrap [collect.com](#) [click here](#)

U.S. News & World Report

The following is a free summary of a premium quality document from Northern Light's [Special Collection](#). If you wish to read the entire document, please press the "Read Document Now" button, or use our "Shopping Cart" to select multiple documents for batch purchasing.

[Read Document Now](#)
[Add to Shopping Cart](#)
[Money Back Guarantee](#)

Title: Patently absurd claims

Summary: Clueless courts and a broken patent system help Amazon

[Return to Results](#)
[New Search](#)
[Shopping Cart](#)

Source: U.S. News & World Report
Date: 03/2020, 2000
Price: \$1.00
Document Size: Very Short (less than 1 page)
Document ID: MA20000328010000289
Citation Information: (ISSN: 0041-5537), Vol. 128 No. 11, pg. 56
Author(s): Randall E. Stross

12.6. ábra: A Northern Light 1 dolláros ára egy hosszú cikkért

DIRECTION OF ATTENTION AND HETEROSOCIAL ANXIETY IN ADOLESCENT MALES
by [Rebecca Lynn Johnson](#)

Degree : PHD
School : THE CATHOLIC UNIVERSITY OF AMERICA (0043)
Date : 1983
Pages : 340
Subject: Psychology, Clinical (0622)

Formats Available	Price	Citizen Club Price
<input checked="" type="radio"/> UNBOUND	\$29.95	\$28.45 (You Save 5%)
<input type="radio"/> SOFTBOUND	\$57.50	\$54.62 (You Save 5%)
<input type="radio"/> HARDBOUND	\$69.50	\$66.02 (You Save 5%)

[ADD TO CART](#)

ABSTRACT
Heterosocial anxiety is a significant concern of young adults. The present study examined a cognitive-attentional model of heterosocial anxiety. It was hypothesized that anxiety would be related to a self-evaluative attentional set, while confidence would be associated with more task-focused attention. Additionally, determinants of attentional focus were explored, as well as the relationships among behavioral and cognitive factors. Subjects were 35 Heterosocially Shy and 37 Heterosocially Confident high school boys, selected from a group of 444 boys on the basis of extreme scores on screening measures of social anxiety and dating behavior. Subjects were audiotaped as they conversed with a female confederate under one of three instructional conditions: Self-Evaluative, Task-Focused, or Natu

12.7. ábra: Disszertációk ingyenes referátumai a Contentville-ben.

Nem mindig közvetlenül a pénz az, amely igazolja egy adatbázis értékelését. Lehet

- az átfogóbb szakterületi lefedettség
- több fontos folyóirat elérhetősége
- több értéknövelő információ
- az indexelés vagy a referálás (vagy mindkettő) jobb minősége
- az információ időszerűsége, vagy
- a szoftver kényelme vagy hatékonyabb működése (vagy mindkettő),

ami egy adatbázist jobbra tehet a többinél. Mary Ellen Bates (2000) kitűnő modellt és példákat nyújt üzleti információs források és szolgáltatások összehasonlítására.

Bates tanulmányának tényleges számai természetesen változhatnak az idők folyamán, és saját forgatókönyvet kell készítenünk el az összehasonlításra, de a példák informatívak és világossá teszik, hogy nem mind arany, ami fénylik, s néhány ingyenes szolgáltatás többbe kerülhet a keresőnek időben, mint a jól szervezett fizetős szolgáltatások. De ahogy ez a könyv is bemutatta, sok, a szakemberek által használt drága, hagyományos adatbázist túl magasra értékelnek, azoknak nagyon súlyos hiányosságai vannak.

Valószínű, hogy a darwini alapelv a legrátermettebb túléléséről érvényesülni fog (Jacsó 2000b), hogy a legrosszabb minőségű kereskedelmi adatbázisok el fognak tűnni, s a mérsékelten árazott alternatívák fogják felváltani őket, amelyek azt nyújtják, amit ígérnek. Egy másik jó megállapításokat tartalmazó cikkében Bates megmagyarázza, hogy „Nincs ingyenes ebéd”, de elismeri, hogy „elég nagy a választék mérsékelten árazott választási lehetőségekből, különösen akkor, ha hajlandók vagyunk a keresés funkcionalitásának egy részéről lemondani.” E könyv szerzőjének véleménye szerint *létezik* ingyenes ebéd, amely egészségesebb az étrendünk és a pénztárcánk szempontjából, mint a drága, állott, a tápérték szempontjából szegényes ebéd (amelyet egyes hagyományos rendszerek ajánlanak), amelyek a műanyag evőeszközöket és a papírpoharat sem érdemlik meg.

Fontos dolog az, hogy mindig jól nézzünk körül, amikor jobb adatbázisokat akarunk találni, s legyenek jó szempontjaink az összehasonlításhoz. Ez a célja ennek a könyvnek: figyelmeztetni a lehetőségekre, hogy gyorsan meghatározzuk, melyik adatbázisra célszerű előfizetni, vagy akár melyik alkalmas egy fontos alkalmi kereséshez. A kritériumok használoról használóra változnak; még egyetlen használó esetében is pillanatnyi információs igényeinek megfelelően.

Remélem, hogy ezen kritériumok ismerete segíteni fogja az olvasót formális, teljes keresztszetszerű értékelésben csakúgy, mint informális, pillanatnyi mini értékelésekben, hogy hozzáértő döntést hozhasson, ahelyett, hogy az adatfájlok készítőinek és az adatbázisok kiadóinak gyakran túl harsány és félrevezető állításait fogadja el.

Ábrák jegyzéke

2.1. ábra: A deskriptorok nagymértékű elírása az MHA adatbázisban.	16
2.2. ábra. Több indextétel kiválasztása egyszerre az Information Science Abstractsból.	17
2.3. ábra. Részlet az ISA szabadalomtulajdonosi indexéből.	17
2.4. ábra. A Dun & Bradstreet adatbázis felelősség elhárítási nyilatkozata.	18
2.5. ábra. Közvetlen ugrópont a hiba jelzésére az Internet Movie Database-ből.	19
2.6. ábra. Adatbázis statisztika a Computer Science Bibliography adatbázisban (részlet)	20
3.1. ábra. Az archINFORM adatbázis tárgyával kapcsolatos nyilatkozata	25
3.2. ábra. Az AIAA magyarázata adatbázisának tárgyról	25
3.3. ábra. Az Information Science Abstracts adatbázis tárgyának meghatározása	26
3.4. ábra. Egy Wilson adatbázis tárgyának meghatározása	27
3.5a. ábra. Nincs rekord a Library Literature adatbázisra.	28
3.5b. ábra. Nincs rekord a Microcomputer Abstracts adatbázisra.	28
3.6. ábra. Az EBSCO adatbázisában külön rekordok találhatók egy folyóirat jelenlegi és korábbi címeire	29
3.7. ábra. Az Ulrich'sban egyetlen rekord található egy folyóirat jelenlegi és korábbi címeire	29
3.8. ábra. A Mark Knopflerre mint művészre vonatkozó keresés eredményei az Amazon.com-nál	31
3.9. ábra. A Mark Knopflerre mint művészre vonatkozó keresés eredményei a Borders.com-nál	31
3.10. ábra. Az ISA divatos kifejezéseket tartalmazó oldala	33
3.11. ábra. A divatos kifejezések eredményének mátrixa	34
3.12. ábra. A tárgyi feltártsággal kapcsolatos állítások a Mental Health Abstracts-ben	35
3.13. ábra. A DIALINDEX-ben végzett keresések eredményei	36
3.14. ábra. A Find Database funkció a WebSPIRS-ben.	37
3.15. ábra. A rosszul meghatározott PSYCH adatbázis csoport a DIALOG-nál.	38
4.1. ábra. Három könyvtár- és információtudományi adatbázis mérete	40
4.2. ábra. Az információ- és könyvtártudományi részhalmaz méretének megállapítása a Social SciSearch és a Trade & Industry adatbázisokban	41
4.3. ábra. Az MHA adatbázis online és CD-ROM változataira egy-egy rekord található a Gale címjegyzékében	43
4.4. ábra. Az MHA online változatainak négy rekordja a World Databasesben.	44
4.5. ábra. Abszurd információt visznek át egyik rekordból a másikba a World Databasesben.	44
4.6. ábra. A szemlecikkek feldolgozási különbségeinek hatása.	45
4.7. ábra. Több rekord ugyanarról a filmről a Bowker cég Complete Video Directoryjában.	46
4.8. ábra. Az Annie Hall mesterrekordja az Internet Movie Database-ben.	46
4.9. ábra. Az Annie Hall DVD-specifikus információjának kapcsolódó rekordja.	47
4.10. ábra. A duplikátumok detektálásának eredményei	48
4.11. ábra. Duplum-eredmények ellenőrzése kontroll-csoportokkal.	49
4.12. ábra. Részlet az ISA duplikátumainak listájáról	50

4.13. ábra. Még mindig maradtak duplikátumok az ISA-ban.	50
4.14. ábra. Újabban hozzáadott duplikátumok.	51
4.15. ábra. Könyvtár- és információtudományi folyóiratok hivatkozási felezési ideje.	53
4.16. ábra. A könyvtár- és információtudományhoz kapcsolódó adatbázisok retrospektivitása	53
4.17. ábra. Pszichiátriai folyóiratok hivatkozási felezési ideje.	54
4.18. ábra. Magfolyóiratok későn kezdett feltárása	55
4.19. ábra. Egy folyóiratról vett tételek rangsorolása a kiadási év szerint, időrendben.	56
4.20. ábra. Új folyóiratok, amelyek feltárását ígerte az ISA és tényleges feltártságuk	57
4. 21a. ábra. Az adatbázis meghintése néhány tétellel fontos forrásokból a MAS adatbázisban.	57
4. 21b. ábra. Fontos források jól kiegyensúlyozott feltártsága a Readers' Guide Abstracts adatbázisban.	58
4.22. ábra. A Library Quarterly hullámvasútszerű feltártsága az ISA-ban.	60
4.23. ábra. Az Information Technology and Libraries hullámvasútszerű feltártsága az ISA-ban.	60
4.24. ábra. Hézagok a nagy presztízsű Program feltártságában az ISA-ban.	61
4.25. ábra. Hézagok a Scientometrics című magfolyóirat feltártságában az ISA-ban.	61
4.26. ábra. Az Online című folyóirat feldolgozásának felfüggesztése vagy megszüntetése a LISA adatbázisban.	62
4.27. ábra. Az AGRICOLA korábban kezdte a jelentős mértékű feltárást, mint ahogy azt állították.	63
4.28. ábra. A PsycINFO és az MHA adatbázisban feldolgozott anyag mennyisége évenkénti bontásban	64
4.29. ábra. Lehetetlen kiadási évek az AGRICOLA adatbázisban.	65
4.30. ábra. A Y2K bug az ISA egyik rekordjában	65
4.31. ábra. Az ISA adatbázis tárgyi feltártságára vonatkozó, revideált állítás.	66
4.32. ábra. Az ISA adatbázishoz adott rekordok számának gyors csökkenése a legutóbbi években.	67
4.33. ábra. Néhány ígéretes állítás egy korai ISA szerkesztőségi cikkből.	68
4.34. ábra. Csökkenő mélységtényező 1995 és 2000 szeptembere között	69
4.35. ábra. Az MIS Quarterly feltártsága a lefelé vezető lejtőn van.	69
4.36. ábra. A Library Administration and Management csökkenő feltártsága az ISA-ban.	70
4.37. ábra. Az NFAIS Newsletter teljes 1999-es anyaga feltáratlan az ISA-ban.	70
4.38. ábra. A hírügynökségi adatbázisok listája az aktualizálás bannerével	73
4.39. ábra. Az aktuális nap feldolgozottságának ellenőrzése a napilap-adatbázisokban.	73
4.40. ábra. Részlet az Ovid adatbázis listájáról a dátumok bannerével.	74
4.41. ábra. Részlet a SilverPlatter adatbázis listájáról a dátumok bannerével.	74
4.42. ábra. Részlet az H. W. Wilson cég adatbázis listájáról a dátumok bannerével.	74
4.43. ábra. Részlet a DIALOG adatbázisainak listájáról a dátumok bannerével.	75
4.44. ábra. Az ISA aktualizálásának mintája 1999-ben.	76
4.45. ábra. A LISA aktualizálásának mintája 1999-ben.	76
4.46. ábra. A Library Literature aktualizálásának mintája 1999-ben.	77
4.47. ábra. Az ISA UD és ZD indexei értékeinek összehasonlítása a tényleges aktualizálási minták megállapítása érdekében	78

4.48. ábra. A rekordok megjelenési év szerinti megoszlása a könyvtár- és információ-tudományi adatbázisok 1999-es aktualizálásaiban.	79
4.49. ábra. A rekordok megjelenési év szerinti megoszlása a könyvtár- és információ-tudományi anyagot tartalmazó adatbázisok 1999-es aktualizálásaiban.	79
4.50. ábra. Tesztkérdés a publikáció évéhez képesti aktualizálás késés meghatározására az ISA-ban	80
4.51. ábra. Az aktualizálás késése a publikáció évéhez képest a LIBLIT adatbázisban.	80
4.52. ábra. Az RQ aktualizálási mintája az ISA-ban	82
4.53. ábra. Az RQ kiadási év mintája az ISA-ban	82
4.54. ábra. Az RQ tényleges aktualizálási éve és kiadási éve mintája a LibLit adatbázisban.	83
4.55. ábra. A CD-ROM Professsional tényleges aktualizálási éve és kiadási éve mintája az ISA-ban.	83
4.56. ábra. Az 1999 nyarán megjelent Compton's 2000-ben benne voltak az 1999 márciusában bejelentett 1998-as Oscar-díjas filmek.	84
4.57. ábra. Az 1999 nyarán megjelent Compton's 2000-ben benne volt az NBA 1999-as NBA bajnok.	84
4.58. ábra. A Microsoft Encarta 2000-ben benne voltak az 1998-as Oscar-díjas filmek	85
4.59. ábra. A Grolier 2000-ben csak 1997-ig voltak benne az Oscar-díjas filmek.	85
4.60. ábra. Husszein király elavult életrajza 1999 végén.	86
5.1. ábra. Már nem kapható könyvek rekordjai a Books in Print DIALOG változatában	88
5.2. ábra. Rekordok könyvkritikával a DIALOG Books in Print adatbázisában	88
5.3. ábra. CD-ROM adatbázisok feltártsága négy címjegyzékben	89
5.4. ábra. A PsycINFO és a Dissertation Abstracts által feldolgozott anyag összehasonlítása a címekben végzett keresés alapján.	90
5.5. ábra. Disszertációk rekordjai az ISA-ban	91
5.6. ábra. Konferencia-előadások az ISA, LibLit és LISA adatbázisokban	91
5.7a. ábra. A LibLit H. W. Wilson változata	93
5.7b. ábra. A LibLit SilverPlatter változata	94
5.7c. ábra. A LibLit Ovid változata	94
5.8. ábra. A konferencia-előadásokat következtelenül azonosítják az ISA adatbázisban.	95
5.9. ábra. A szabadalmi rekordok száma az ISA adatbázisban	96
5.10. ábra. Túl sok kritikátípus található a Biographies Plus Illustrated adatbázisban	96
5.11. ábra. Az Ovid Mental Health Collection adatbázisában szereplő folyóiratok helyezése impakt faktoruk szerint	97
5.12. ábra. A legmagasabb impakt faktorral rendelkező pszichiátriai és pszichológiai folyóiratok 1997-ben.	99
5.13. ábra. A legtöbb hivatkozással rendelkező pszichiátriai és pszichológiai folyóiratok 1997-ben.	100
5.14. ábra. A legproduktívabb pszichiátriai és pszichológiai folyóiratok 1997-ben.	100
5.15. ábra. Az MHA által 1997-ben feldolgozott folyóiratok listája (részlet)	100
5.16. ábra. Nagy presztízsű könyvtár- és információ-tudományi folyóiratok feldolgozottsága hat adatbázisban	102
5.17. ábra. 94 kémiai és matematikai folyóirat feldolgozottsága négy témafigyelő szolgáltatásban.	103
5.18. ábra. Folyóiratok közötti átfedés az IAC Expanded Academic Indexben és az H. W. Wilson cég adatbázisaiban	103

5.19. ábra. Átfedések az Avery Index és az Architectural Periodicals Index által feldolgozott folyóiratok között	104
5.20. ábra. Részlet a PsycINFO folyóiratlistájának bevezetőjéből.	105
5.21. ábra. A PsycINFO folyóiratlistája	105
5.22. ábra. A UMI folyóiratlistájának egyik formátuma	106
5.23. ábra. Az H. W. Wilson cég egy adatbázishoz kapcsolódó folyóiratlistája	106
5.24. ábra. Folyóiratok keresése a WebSPIRS adatbázisokban	107
5.25. ábra. Keresés a Mental Health Abstracts JN és SO indexeiben	108
5.26. ábra. A DIALOG adatbázisainak nemzetközi feltártságára vonatkozó állításai	109
5.27. ábra. A PASCAL adatbázis folyóiratainak földrajzi megoszlása	109
5.28. ábra. Az USA-beli, brit, kanadai és ausztrál folyóiratok dominanciája az Ulrich's-ban	110
5.29. ábra. Az időszaki kiadványok földrajzi megoszlása az EBSCO The Serials Directory adatbázisában	110
5.30. ábra. Az ISDS adatbázis időszaki kiadványainak földrajzi megoszlása	111
5.31. ábra. Az időszaki kiadványok földrajzi megoszlása az ISI Journal Citation Reports adatbázisának társadalomtudományi sorozatában	111
5.32. ábra. A Latin American News adatbázis forrásainak földrajzi összetétele	113
5.33. ábra. A nem az USA-ban és Kanadában megjelent folyóiratokból származó rekordok csökkenő aránya az EMBASE adatbázisban	114
5.34. ábra. Aránytalanul sok cseh nyelvű rekord található a Business & Industry Database-ben	115
5.35. ábra. A nemzetközi folyóiratok megfelelő keveréke figyelhető meg a LibLit adatbázisban	116
5.36. ábra. Az All Movie Guide országok szerinti listája	117
5.37. ábra. Az Internet Movie Database országstatisztikája	117
5.38. ábra. Az EMBASE adatbázis által feldolgozott idegen nyelvű dokumentumok számának csökkenése ötvenkénti adatok alapján.	118
5.39. ábra. A nem angol nyelvű dokumentumok csökkenő aránya a Social SciSearch adatbázisban	119
5.40. ábra. A nem angol nyelvű dokumentumok csökkenő aránya a Sociological Abstracts adatbázisban	119
5.41. ábra. A nem angol nyelvű dokumentumok csökkenő aránya a LISA adatbázisban	120
5.42. ábra. A LibLit nyelv szerinti korlátozása a WebOvidban	120
5.43. ábra. Nyelv szerinti korlátozás az ISA WebSPIRS változatában	121
5.44. ábra. Japán és koreai nyelvű anyagok az MHA és a PsycINFO adatbázisokban	122
5.45. ábra. „Segítő” információ a nyelv szerinti keresésről a Complete Video Directoryban	123
5.46. ábra. Pezsgő ígéretek a Bowker's Complete Video Directory lehetőségeiről	123
5.47. ábra. A tényleges keresés keserű valósága	124
5.48. ábra. Segítő információ a nyelv szerinti keresésről a The Serials Directoryban.	124
5.49. ábra. Az ISA-ról íródott szerkesztőségi cikkek listája az adatbázis saját weboldalán	125
5.50. ábra. Állítások a kizárólag a LISA-ban vagy az ISA-ban feldolgozott témákról	127
5.51. ábra. A JASIS-ben megjelent cikkek százalékos megoszlása a szerzők munkahelye szerint	128
5.52. ábra. A LISA-ban és az ISA-ban végzett tesztkeresések eredményei	129

5.53. ábra. A gyors javítás körülbelül annyira volt hatásos, mint magának a sikerekről beszámoló bekezdésnek a helyesírás-ellenőrzése	130
5.54. ábra. Régi és új hiányok a Journal of Scholarly Publishing feldolgozásában	131
5.55. ábra. Néhány releváns cikk a Journal of Scholarly Publishing című folyóiratból, amelynek feldolgozását 1998 után az ISA abbahagyta	132
5.56. ábra. Folyóiratok feltártságának gyors csökkenést mutató trendje az ISA-ban az adatbázis által alapvetőnek mondott témákban	132
5.57. ábra. Az ISA által legfontosabbnak mondott, feltárt területek az adatbázisleírás 2001 februári változatában.	133
5.58. ábra. Az M.D. Computing az orvosi informatikával foglalkozó társadalomtudományi cikkek között a leghivatkozottabb források közé tartozik.	134
5. 59a. ábra. Folyóiratok és mérőszámaik a Journal Citations Reports 2000 mesterséges intelligencia kategóriájában	135
5. 59b ábra. Az Artificial Intelligence különböző címváltozatokkal a leghivatkozottabb folyóirat a mesterséges intelligenciával foglalkozó cikkekben a SciSearch adatbázisban	135
5. 59c ábra. Az Artificial Intelligence különböző címváltozatokkal a leghivatkozottabb folyóirat a mesterséges intelligenciával foglalkozó cikkekben a Social SciSearch adatbázisban	136
5. 59d ábra. Az Artificial Intelligence feldolgozottsága teljesen megszűnt az ISA adatbázisban, annak ellenére, hogy a szakemberek abszolút mértékben elismerik.	136
5. 60a. ábra. Az IEEE Transactions on Information Theory helye az impakt faktor szerinti rangsorban	137
5. 60b. ábra. Az ISA magfolyóiratainak utolsó listája	138
5.61. ábra. Az „új” folyóiratok feldolgozottsága 2001 júliusában	140
5. 62a. ábra. Az adatbázis-leírás helyesen rendszertelennek jelöli az ISA aktualizálási gyakoriságát	141
5. 62b. ábra. Az ISA adatbázis-leírásának következő változata új hírként közli, hogy évente 9 alkalommal fogják aktualizálni	141
5. 63a. ábra. Az ISA adatbázis tényleges aktualizálási gyakorisága	142
5. 63b. ábra. A megközelítőleg hathetes aktualizálási időköz ígérete	143
5. 63c. ábra. Az aktualizálási címkék jelzik a frissesség problémáját	144
5. 63d. ábra. 2000 első félévének tényleges aktualizálási időközei	144
5.64. ábra. Az ISA korábbi szakterületeinek csaknem egyharmadát eltüntették az ISA adatbázis-leírásának 2001 februári változatában.	145
5.65. ábra. Az ISA adatbázis növekedésének csökkenő üteme a tulajdonosváltás óta	146
5.66. ábra. Kollázs az ISA adatbázisának főiskolai weboldalakon található leírásairól	147
5.67. ábra. Utalás az ISA által feldolgozott anyagra az ISA honlapján	148
5.68. ábra. A JASIS-ben megjelent cikkben kitagadott könyvtári témák feldolgozottsága az ISA-ban	148
5.69. ábra. Az egyik, az ISA adatbázis-leírásában kiemelten említett téma (multimédia információs rendszerek) feldolgozottsága	149
5.70. ábra. Elsődlegesen a könyvtártudományhoz kapcsolódó folyóiratok, amelyeket továbbra is referálnak és indexelnek az ISA-ban	150
5.71. ábra. A JASIS-nek egy egész 1999-es száma hiányzik az ISA-ból, ugyanakkor megtalálható a LISA-ban	152

5.72. ábra. A magasan rangsorolt Annual Review of Information Science and Technology (ARIST) átlag alatti feldolgozottsága az ISA-ban	152
5.73. ábra. Az ISA 1999 januári, rendkívül optimista szerkesztőségi cikke	154
5.74. ábra. Az ISA 1999 áprilisi szerkesztőségi cikke annak okáról, miért csökkent a referátumok száma	155
6. 1a. ábra. Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (ProQuest)	158
6. 1b. ábra. Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (RGA)	158
6. 1c. ábra. Ugyanazon cikk bibliográfiai leírásainak összehasonlítása (Gale Group Magazine Database)	158
6.2. ábra. Az Ulrich's Ovid változata közvetlenül kattintható ugrópontot szolgáltat a kiadó és a folyóirat weblapjához.	159
6.3. ábra. Részlet a Gale Group Biography Resource Centerének egy tételéből.	159
6.4. ábra. Részlet az H. W. Wilson cég Biographies Plus Illustrated adatbázisának egy tételéből.	160
6.5. ábra. Egy tétel a Merriam-Webster Biographical Dictionaryből.	160
6.6. ábra. Állandó folyóirat kódok használata a korábbi és a mostani címek megtalálásához.	162
6.7. ábra. A kritikák osztályzatai a Periodical Abstracts adatbázisban.	162
6.8. ábra. Az INSPEC adatbázis cikkeinek téma megközelítési mód kódja.	163
6.9. ábra. Az ERIC adatbázis célközönség indexe az Ovidnál.	163
6.10. ábra. Részlet az ABI/INFORM adatbázis egy rekordjából a DIALOG-nál.	164
6.11. ábra. Részlet az ABI/INFORM adatbázis egy rekordjából az Ovidnál.	164
6.12. ábra. Részlet az ABI/INFORM adatbázis egy rekordjából a ProQuestnél.	165
6.13. ábra. Szöveg plusz grafika formátum a ProQuestnél.	165
6.14. ábra. A kinagyított képi összetevő a ProQuestnél.	166
6.15. ábra. Részlet egy ABI/INFORM rekord képi változatából a ProQuestnél.	166
6.16. ábra. A kinagyított kép a ProQuestnél.	167
6.17. ábra. Részlet az Ovid Mental Health Collectionének egy teljes szövegű rekordjából.	167
6.18. ábra. Bibliográfiai adatok a referátumokkal ellátott bibliográfiai rekordokhoz vezető ugrópontokkal.	168
6.19. ábra. Ugrópont a kritika teljes szövegéhez a BRC adatbázisból.	169
6.20. ábra. Ugrópont a kritika teljes szövegéhez a BP adatbázisból.	169
6.21. ábra. Tételek az évfolyamok, számok és oldalszámok adataival az NCJRS folyóiratindexében.	169
6.22. ábra. Tételek kronológiai-numerikus adatokkal a PaperChem folyóiratindexében.	170
6.23. ábra. Különböző tárgyi kódok szemmel láthatóan ugyanahhoz a tárgyi kifejezéshez kapcsolódóan az ISA indexében.	171
6.24. ábra. Tárgyi kódok és szöveges megfelelőjük a megjelenített rekordokban.	171
7.1. ábra. A leggyakrabban elírt szavak.	175
7.2. ábra. A Rorschach névnek egy helyesen írt és egy elírt változata van a referátumban.	175
7.3. ábra. Helyesen írt változat az eredeti (spanyol) címben és elírt változat az angol címben.	176
7.4. ábra. A Rorschach név egyetlen előfordulása el van írva a referátumban.	176
7.5. ábra. A toxoplasma szó egyetlen előfordulása el van írva a címben.	176
7.6. ábra. Elírások az ISA nyelvi mezőjében.	177

7.7. ábra. A Portuguese elírt és helyesen írt változatainak aránya a nyelvi indexben.	177
7.8. ábra. Az English szó nagyszámú elírása a Mental Health Abstracts adatbázis nyelvi mezőjében.	178
7.9. ábra. A nyelvek teljesen hibátlan névalakjai a LibLit adatbázisban.	179
7.10. ábra. Pontatlan névalakok a forrásdokumentumban, a Serials Review-ban.	180
7.11. ábra. A szerző neve leggyakrabban elírt változatának gyakorisága.	181
7.12. ábra. A Jacsó név elírt és helyesen írt változatai szerzőként és hivatkozott szerzőként a Social SciSearch adatbázisban.	182
7.13. ábra. Hibás státusz információ az Ulrich's 1998 őszi kiadásában.	184
7.14. ábra. Egy argentin levéltári folyóirat valótlannul magas példányszám adatokkal.	184
7.15. ábra. Az ISA-ban indexelt folyóiratok száma az Ulrich's szerint.	185
7.16. ábra. Jelentős különbségek vannak a lektorált folyóiratok szakterületenkénti számában.	186
7.17. ábra. Jelentős különbségek vannak a lektorált folyóiratok országonkénti számában.	186
7.18. ábra. Jelentős különbségek vannak a lektorált folyóiratok kiadónkénti számában.	187
7.19. ábra. Lektorált könyvtár- és információtudományi folyóiratok listája az Ulrich's-ban (részlet)	187
7.20. ábra. Lektorált könyvtár- és információtudományi folyóiratok listája a The Serials Directoryban (részlet)	188
7.21. ábra. Az ACM lektorált folyóiratai, ahogy azokat az ACM meghatározta.	188
7.22. ábra. Az ACM lektorált folyóiratai, ahogy azokat az Ulrich's meghatározta.	189
7.23. ábra. Az ACM lektorált folyóiratai, ahogy azokat a The Serials Directory meghatározta	189
7.24. ábra. Olyan folyóiratok keresése, amelyeket az Excerpta Medica referál és indexel a DIALOG-nál.	190
7.25. ábra. Referáló és indexelő szolgáltatás hibásan kódolt neve az Ulrich's DIALOG-os változatában.	190
7.26. ábra. Pontos kódolt Excerpta Medica az Ulrich's Ovidos változatában.	190
7.27. ábra. Nyilvánvaló pontatlanság a G. K. Saur World Databases adatbázisában.	191
7.28. ábra. E-mail a zavaró kronológiai jelzéssel kapcsolatban.	192
7.29. ábra. E-mail a Sage hibás információiról egyik kiadványuk megjelenési gyakoriságával kapcsolatban.	192
8. 1a. ábra. A nyelvi index kifejezései a Sociological Abstractsben.	195
8. 1b. ábra. A kiadás országa indexkifejezései a Sociological Abstractsben	195
8. 1c. ábra. Deszkriptorok a Sociological Abstracts tezaurusában.	195
8.2. ábra. A Bowker név variációinak tömege egy Bowker adatbázisban.	196
8.3. ábra. Tételek az Ulrich'sban a John Wiley & Sons, Inc. név alatt.	197
8.4. ábra. További tételek százai az Ulrich'sban a Wiley & Sons, Inc. név alatt.	197
8.5. ábra. A U. S. News & World Report címének különböző formái.	198
8.6. ábra. Súlyos következetlenség a LISA szerzői indexében.	199
8.7. ábra. Részlet a példás Getty Name Authority File-ből.	199
8.8. ábra. A Database folyóirat címének változatai egyetlen adatbázisban, a PASCAL-ban.	200
8.9. ábra. Egy hosszabb folyóiratcím változatai a PASCAL-ban (egymáshoz közeli tételek).	200

8. 10a. ábra. Egy hosszabb folyóiratcím változatai a PASCAL-ban (szétszóródott tételek).	200
8. 10b. ábra. Egy hosszabb folyóiratcím változatai a PASCAL-ban [folytatás]	201
8.11. ábra. A találatok számának megoszlása egy folyóirat címének két egymáshoz nem közeli formája között.	201
8.12. ábra. Következetlenül meghatározott dokumentumtípusok ugyanannak a rovatnak a közleményeiben.	202
9.1. ábra. A Kongresszusi Könyvtár osztályozási jelzeteinek nagymértékű kihagyása az Ulrich's Plus adatbázisban.	204
9.2. ábra. A dokumentumtípus mező kihagyása az NTIS adatbázisból minden magyarázat nélkül.	204
9.3. ábra. A dokumentumtípus mező kihagyása a GeoArchive adatbázisból minden magyarázat nélkül.	205
9.4. ábra. A mezők jelenlétének ellenőrzése a könyvtár- és információtudományi adatbázisokban.	205
9.5. ábra. A mezők jelenlétének ellenőrzése a PsycINFO és az MHA adatbázisokban.	206
9.6. ábra. Részlet az MHA adatbázis-leírásából.	206
9.7. ábra. Nincs rekord az English szóra az MHA nyelvi mezőjében.	206
9.8. ábra. Részlet az MHA adatbázis nyelvi indexéből.	207
9.9. ábra. A rekordok teljes számának és két mező jelenlétének meghatározása az Ovidban.	207
9.10. ábra. Az Oceanic Abstracts adatbázis-leírásának információi.	208
9.11. ábra. Hiányzik a dokumentumtípus mező nagy mértékű kihagyására vonatkozó megjegyzés az NTIS-ben.	209
9.12. ábra. A hiányzó adatelemek egyértelmű meghatározása.	210
9.13. ábra. Rekordok meghatározatlan nyelvvel a LibLit-ben.	210
9.14. ábra. Meghatározatlan nemre vonatkozó információk a Marquis Who's Who adatbázisban	211
9.15. ábra. Rengeteg olyan rekord van az ABI címjegyzékben, amely a 9999-es SIC kódot kapta.	211
9.16. ábra. Az Office Depot mint osztályozhatatlan szervezet az ABI címjegyzékben.	212
9.17. ábra. A Walt Disney Company a 9999-es kódot kapta a Moody's-tól.	212
9.18. ábra. Az America Online, American Eagle Airlines és Burger King a 9999-es kódot kapta a Dun & Bradstreettől.	212
9.19. ábra. A 9999-es kód félrevezető használata a Books in Printben	213
9.20. ábra. A rekordok teljes számának keresése az ERIC DIALOG-os változatában.	214
9.21. ábra. A teljességre vonatkozó keresés a Bowker's Complete Video Directoryban.	214
9.22. ábra. A teljességre vonatkozó keresés az Ulrich's Bowker által kiadott CD-ROM változatában.	214
9.23. ábra. Egy kevés értéket tartalmazó index átnézése az Ovidban.	215
9.24. ábra. A kiadványtípus és a nyelvi mező teljességének összehasonlítása a PAIS Ovidos változatában.	216
9.25. ábra. A példányszám mező teljességének meghatározása az Ulrich's DIALOG-os változatában.	216
9.26. ábra. A teljesség tesztelése a Bowker's Complete Video Directory 2000 tavaszi kiadásában.	217
9.27. ábra. A katalógus kihívása és hallgatólágos ígérete - és a valóság.	218

9.28. ábra. A keresés első lépései.	218
9.29. ábra. A keresés szűkítése a vígjáték műfajára.	219
9.30. ábra. Nem kapunk találatokat, maga a keresés válik vígjátékká.	219
9.31. ábra. A sűgő könyvekre utal, de azért ad némi útmutatást.	220
10.1. ábra. Zavaró és hibás hivatkozások a LIBRARY SCHOOL STUDENTS és LIBRARY RELOCATION alatt	230
10.2. ábra. A szakirodalomban nem létező spanyol kifejezések	232
10.3. ábra. Rengeteg kirívó elírás az MHA deskriptor mezőjében	232
10.4. ábra. Rosszul megválasztott angol deskriptor a PASCAL-ban	233
10.5. ábra. A „Searching for Skeletons” cikk rekordjai a PIRA adatbázisban	234
10.6. ábra. A „Searching for Skeletons” cikk rekordjai az ABI/INFORM adatbázisban	234
10.7. ábra. A „Data Transfer” cikk rekordjai az ABI/INFORM adatbázisban	235
10.8. ábra. A „Searching for Skeletons” cikkek rekordjai az INSPEC-ben	235
10.9. ábra. A „Searching for Skeletons” cikkek rekordjai az LISA-ban	235
10.10. ábra. A „Searching for Skeletons” cikk 2. részének rekordja a PASCAL-ban	236
10.11. ábra. A „Data Transfer” cikkek rekordjai a LISA-ban	237
10.12. ábra. A LISA rekordjai redundanciával foglalkozó cikkekhez	237
10.13. ábra. A T&I rekordja a „Negotiating” cikkhez	237
10.14. ábra. A „Searching for Skeletons” cikkek rekordjai a T&I adatbázisban	238
10.15. ábra. A „Searching for Skeletons” cikkek rekordjai a Business Abstractsben.	238
10.16. ábra. A „Quality of Abstracts” cikk rekordja a PIRA adatbázisban	239
10.17. ábra. Egy cikk eredetijének és magyar fordításának rekordjai a LISA-ban	239
10.18. ábra. A „Negotiating” cikk rekordjai az INSPEC-ben.	240
10.19. ábra. A duplikátumok ciklikus mintája az ISA adatbázisban	240
10.20. ábra. Az aktualizálás rekordjainak tökéletesen kerek száma az ISA adatbázisban	241
10.21. ábra. A széles tárgyi kategóriák következetlen meghatározása az ISA duplikátumaiban.	242
11.1. ábra. Korlátozás a DIALOG-ban parancsmódban	249
11.2. ábra. Korlátozás az Ovid haladó menümódjában	249
11.3. ábra. A referátumokat tartalmazó rekordok aránya olyan adatbázisokban, amelyeknek a címében szerepel az Abstracts szó	250
11.4. ábra. A referátumokat tartalmazó rekordok aránya olyan adatbázisokban, amelyeknek a címében nem szerepel az Abstracts szó	250
11.5. ábra. A referátumokban szereplő szavak átlagos száma négy adatbázisban	251
11.6. ábra. Könyvkritikák rekordjainak egyforma „referátumai” az Academic Search Elite-ben	251
11.7. ábra. Könyvkritikák rekordjainak „referátumai” a Periodical Abstract PlusText adatbázisban	252
11.8. ábra. Egy könyvkritika terjedelmes referátuma a Newspaper Abstracts Daily adatbázisban	253
11.9. ábra. A „Bosszú” cikk referátuma az RGA-ban	254
11.10. ábra. A „Bosszú” cikk referátuma az Academic Search Elite-ben	254
11.11. ábra. A „Bosszú” cikk referátuma a Periodicals Abstracts PlusTextben	254
11.12. ábra. A „Harvardi Hamiska” cikk referátuma az Academic Search Elite-ben	255
11.13. ábra. A „Harvardi Hamiska” cikk referátuma a Periodicals Abstracts PlusTextben	255

11.14. ábra. A „Crimson Copycat” cikk referátuma az RGA-ban	255
11.15. ábra. A „Lincoln” cikk referátuma az RGA-ban	256
11.16. ábra. A „Lincoln” cikk referátuma a Periodical Abstracts PlusTextben	256
11.17. ábra. A „Lincoln” cikk referátuma az Academic Search Elite-ben	257
11.18. ábra. Referátum az ABI/INFORM-ból	258
11.19. ábra. Referátum a Periodical Abstracts PlusTextből	259
11.20. ábra. Referátum az H. W. Wilson cég Business Abstracts adatbázisából	259
11.21. ábra. Referátum az ISA-ból	259
11.22. ábra. Referátum a Trade & Industry Database-ből	260
11.23. ábra. Referátum a LISA-ból	260
11.24. ábra. Referátum az ISA-ból	261
11.25. ábra. Referátum az INSPEC-ből	261
11.26. ábra. Referátum az ABI/INFORM-ból	262
11.27. ábra. Referátum a LISA-ból	262
11.28. ábra. Referátum az ISA-ból	263
11.29. ábra. Referátum a Trade & Industry Database-ből	263
11.30. ábra. Referátum a Periodical Abstracts PlusTextből	263
11.31. ábra. Referátum az H. W. Wilson cég Business Abstractséből	264
11.32. ábra. Referátum az ABI/INFORM-ból	264
11.33. ábra. Az ISA referátuma	265
11.34. ábra. Referátum az INSPEC-ből	265
11.35. ábra. Referátum a LISA-ból	266
11.36. ábra. Referátum az ABI/INFORM-ból	266
11.37. ábra. Referátum az INSPEC-ből	267
11.38. ábra. Az első referátum az ISA-ból	267
11.39. ábra. A második, az ERIC-ből importált referátum az ISA-ból.	267
11.40. ábra. Referátum a LISA-ból	268
12.1. ábra. Mintarekord a PubListből	270
12.2. ábra. Mintarekord az Ulrich'sből (1. rész)	270
12.3. ábra. Mintarekord az Ulrich'sből (2. rész)	271
12.4. ábra. Próbakeresés az EContent feldolgozottságáról	271
12.5. ábra. A kifejezések leképezése és a teaurusz megjelenítése az ERIC/AE-ben.	273
12.6. ábra. A Northern Light 1 dolláros ára egy hosszú cikkért	277
12.7. ábra. Disszertációk ingyenes referátumai a Contentville-ben.	277

Bibliográfia

- Ahmad, Nazir. 1991. Newspaper Indexing: An International Overview. *The Indexer*. 17(4): 257-65.
- Ajiferuke, Isola, and Clara M. Chu. 1988. Quality of Indexing in Online Databases: An Alternative Measure For a term Discriminating Index. *Information Processing & Management* 24(5): 599-601.
- Allcock, Harry M. 1997. IFI/Plenum Takes Jacso to Task. *Database* 20(3): 6.
- Amba, S., and M. D. Naresh. 1994. Coverage of Leather Literature in CD-ROM Databases. *Online & CD-ROM Review*. 18(6): 341-346.
- American Libraries*. 1986. News Brief: Just Doing Her Job. 17(2): 115.
- . 1996. News Fronts USA. Chicago PL Takes Heat for Calendar's Errors. 27(1): 25-26.
- Atherton, Pauline Cochrane, and Stella Keenan. 1965. Review of AIP/Documentation Research Project Studies. Prepared for the American Institute of Physics, New York. Report No. AIFPDRP-65-2, March 1965. URL: <http://www.libsci.sc.edu/bob/ISP/cochrane2.htm> Accessed July 15, 2001.
- Ballard, Terry, and Arthur Lifshin. 1992. Prediction of OPAC Spelling Errors through a Keyword Inventory. *Information Technology and Libraries* 11(2): 139-45.
- Basch, Reva. 1990a. May I Help You? Customer Service and Beyond. *Database Searcher* 6(6): 14-18.
- . 1990b. Measuring the Quality of the Data: Report on the Fourth Annual SCOUG [Southern California Online User Group] Retreat; Developing a Framework for Judging the Quality and Reliability of Databases. *Database Searcher* 6 (October): 18-23.
- Bates, Mary Ellen. 2000. TANSTAAFL: In Search Of the Free Lunch and No-Cost/Low-Cost Full Text Archives. *Searcher* 8(6): 55-59.
- Beaubien, Denise M. 1992. Wilson vs. IAC on Tape: A Comparison. *Database* 15(1): 52-56.
- Benning, Susan P., and Susan C. Speer. 1993. Incorrect Citations: A Comparison of Library Literature with Medical Literature. *Bulletin of the Medical Library Association* 81(1): 56-68.
- Boissonnas, Christian M. 1979. The Quality of OCLC Bibliographic Records. *Law Library Journal* 72(1): 80-85.
- Borko, Harold, and Charles L. Bernier. 1975. *Abstracting Concepts and Methods*. New York: Academic Press.
- Bottle, Robert T. and Efthimis Efthimiadis. 1984. Library and Information Science Literature: Authorship and Growth Patterns. *Journal of Information Science: Principles & Practice* 9(3): 107-116.
- Bourne, Charles P. 1977. Frequency and Impact of Spelling Errors in Bibliographic Data Bases. *Information Processing & Management* 13(1): 1-12.
- Boykikeva, I. 1994. Online Access to Japanese Information in Engineering: Comparative Analysis of the JICST-E, INSPEC, and COMPENDEX Databases. *Information Services & Use* 14(1): 25-35.

- Braam, Robert R., and Jeanet Bruil. 1992. Quality of Indexing Information: Author's Views on Indexing of Their Articles in Chemical Abstracts Online CA-File. *Journal of Information Science* 18(5): 339-408.
- Brenner, Everett H. 1989. Should Abstractors Index? *Newsletter of the American Society of Indexers* (91):1.
- Brenner, Sandra H. and Emma Jean McKinn. 1989. CINAHL and MEDLINE: A Comparison of Indexing Practices. *Bulletin of the Medical Library Association*. 77(4): 366-371.
- Briggs, Kim, and, and Ian Crowlesmith. 1995. EMBASE - The Excerpta Medica Database: Quick and Comprehensive Drug Information. *Publishing Research Quarterly* 11(3): 51-60.
- Brooks, Terrence A. 1993. All the Right Descriptors: A Test of the Strategy of Unlimited Aliasing. *Journal of the American Society for Information Science* 44(3): 137-147.
- Buntrock, Robert. 1994. Chemical Compound Registration - Algorithmic or Otherwise (Use of Chemical Abstract registry Numbers in Database Indexes). *Database* 17(1): 108-110.
- . 1995. GOLD [CASRN=7440-57-5]. Is Where You Find It, or Caveats on Finding Chemical Substance Using CASRN. *Database* 18(3): 50-52., 54-55.
- Byrne, Alex. 1983. How to Lose a Nation's Literature: Database Coverage of Australian Research Database. *Database* 6(3): 10-17.
- Cahn, Pamela. 1994. Testing Database Quality. *Database* 17(1): 23-26., 28-30.
- Chall, Miriam, and Terry M. Owen. 1995. Documenting the World's Sociological Literature: Sociological Abstracts. *Publishing Research Quarterly* 11(3): 83-95.
- Chan, Lois Mai. 1989. Inter-Indexer Consistency in Subject Cataloging. *Information Technology and Libraries* 8(4): 349-358.
- Cheeseman, Elanine N. 1995. Patents Preview and Patent Fast-Alert. *Database* 18(4): 65-71.
- Cleveland, Donald B., and Ana D. Cleveland. 2001. *Introduction to Indexing and Abstracting*. Englewood, CO: Libraries Unlimited.
- Cooper, William S. 1969. Is Interindexer Consistency a Hobgoblin? *American Documentation* 20(3): 268-278.
- Dansey, P. 1973. A Bibliometric Survey of Primary and Secondary Information Science Literatures. *ASLIB Proceedings* 25(7): 252-263.
- Dixon, Bernard. 1988. Science and Information Society. *Scholarly Publishing* 20(1): 3-12.
- Duelstgen, Ronald R. 1991. Access to Japanese Technical Information. *Database* 14(2): 105-107.
- Dwyer, Jim. 1991. Invisible College at Work: The Case of Dirty Database Test. *Cataloging & Classification Quarterly* 14(1): 75-82.
- Edwards, Tom. 1976. A Comparative Analysis of the Major Abstracting and Indexing Services for Library and Information Services. *Unesco Bulletin for Libraries* 30(1): 18-25.
- Eldredge, Jonathan D. 1993. Accuracy of Indexing Coverage Information as Reported by Serials Sources; Ulrich's International Periodicals Directory; SERLINE; The Serials directory. *Bulletin of the Medical Library Association (Bull Med Libr Assoc)*. 81 (October): 364-370.

- . 1997. Identifying Peer-review Journals in Clinical Medicine. In the Serials Directory and Ulrich's International Periodicals Directory. *Bulletin of the Medical Library Association (Bull Med Libr Assoc)*. 85 (4): 418-422.
- Ernest, Douglas J., Holley R. Lange, and Della Herring. 1988. An Online Comparison of Three Library Science Databases. *RQ* 28(2): 185-194.
- Esteibar, Belen Altuna, and F. Wilfrid Lancaster. 1992. Ranking of Journals in Library and Information Science by Research and Teaching Relatedness. *The Serials Librarian* 23(1/2): 1-10.
- Ewbank, Bruce W. 1982. Comparison Guide to Selection of Databases and Database Services. *Drexel Library Quarterly* 18(3/4): 189-204.
- Freedman, Bernadette. 1995. Growth and Change in the World's Biological Literature as Reflected in BIOSIS Publications. *Publishing Research Quarterly* 11(3): 61-79.
- Froehlich, Thomas. 1994. User Assumptions about Information Retrieval Systems: Ethical Concerns. In: *Ethics in the Computer Age Conference Proceedings*, edited by Joseph M. Kizza, November 11-13., 1994, Gatlinburg, TN. New York: ACM, 146-150.
- Froom, Paul, and Jack Froom. 1993. Deficiencies in Structured Medical Abstracts. *Journal of Clinical Epidemiology* 46(7): 591-594.
- Fry, Sally A., and Kathy A. Parsons. 1994. Comparative Analysis of IAC, IMU, and Wilson Database Tape Products. *The Serials Librarian* 25 (1/2): 133-144.
- Fugmann, Robert. 1992. Predictability versus Consistency. *International Classification* 19(1): 20-22.
- Funk, M. E., C. A. Reid, and L. S. McGoogan. 1983. Indexing Consistency in MEDLINE. *Bulletin of the Medical Library Association* 71(2): 176-183.
- Gibson, Robert W., and Barbara K. Kunkel. 1980. Japanese Information Network and Bibliographic Control: Scientific and Technical Literature. *Special Libraries* 71(3): 154-162.
- Gilchrist, Alan. 1966. Documentation of Documentation: A Survey of Leading Abstracts Services in Documentation and an Identification of Key Journals. *ASLIB Proceedings* 18(3): 62-80.
- Gilchrist, Alan, and Alexandra Presanis. 1971. Library and Information Science Abstracts: The First Two Years. *ASLIB Proceedings* 23(5): 251-256.
- Giral, Angela, and Arlene G. Taylor. 1993. Indexing Overlap and Consistency Between the Avery Index to Architectural Periodicals and the Architectural Periodicals Index. *Library Resource & Technical Services* 37(11): 19-44.
- Goldberg, Martin. 1992. CD-ROM Periodical Indexes: Better Evaluation Necessary. *The Indexer* 18(1): 11-15.
- Goldstein, Samuel. 1973. Statistical Bibliography and Periodical Literature. 1972 Abstracting, Indexing, and Coverage of Library and Information Science Periodicals, Part 4. *Current Awareness - Library Literature* 2(4): 3-13.
- Griffith, Belder C., Howard D. White, Carl M. Drott, and Jerry D. Saye. 1986. Test of Methods for Evaluating Bibliographic Databases: An Analysis of the National Library of Medicine's Handling of Literatures in the Medical Behavioral Sciences. *Journal of the American Society for Information Science* 37(4): 261-270.

- Grzeszkiewicz, Anna, and Craig A. Hawbaker. 1996. Investigating a Full-Text Journal Database: A Case of Detection. *Database* 19(6): 59-62.
- Harbourt, Anna M., Lou Knecht, and Betsy L. Humphreys. 1995. Structured Abstracts in MEDLINE, 1989-1991. *Bulletin of the Medical Library Association* 83(2): 190-195.
- Harter, Stephen P., Thomas E. Nisonger, and Aiwei Weng. 1993. Semantic Relationships Between Cited and Citing Articles in Library and Information Science Journals. *Journal of the American Society for Information Science* 44(9): 543-552.
- Hartley, James, and Matthew Sydes. 1996. Which Layout Do You Prefer? An Analysis of Readers' Preference for Different Typographic Layouts of Structured Abstracts. *Journal of Information Science* 22(1): 27-37.
- Hartley, James, Matthew Sydes, and Anthony Blurton. 1996. Obtaining Information Accurately and Quickly: Are Structured Abstracts More Efficient? *Journal of Information Science* 22(5): 349-356.
- Hawkins, Donald T. 1999. What is Credible Information? *Online* 23(5): 86-89.
- 2001a. Information Science abstracts: Tracking the Literature of Information Science. Part 1: Definition and Map. *Journal of the American Society for Information Science and Technology* 52(1): 44-53.
- 2001b. Setting the Database record Straight. *Information World Review* No. 166: 12-13.
- Hawkins, Donald T., and Lynn A. Murray. 1998a. New Publishers for ISA. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9807.htm>. Accessed July 15, 2001.
- 1998b. Synergy, New Journals. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9808.htm>. Accessed July 15, 2001.
- 1998c. Currency. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9810.htm>. Accessed July 15, 2001.
- 1998d. Core Journals. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9812.htm>. Accessed July 15, 2001.
- 1999a. Looking Back, Looking Ahead. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9901.htm>. Accessed July 15, 2001.
- 1999b. Currency, relevance and Quality. *Information Today, Inc. Home Page - ISA Editorials*. URL: <http://www.infotoday.com/isa/ed9904.htm>. Accessed July 15, 2001.
- 2000a. ISA for the New century! *Information Today, Inc. Home Page - ISA Editorials* 35(1). URL: <http://www.infotoday.com/isa/ed0001.htm>. Accessed July 15, 2001.
- 2000b. New features added. *Information Today, Inc. Home Page - ISA Editorials* 35(7). URL: <http://www.infotoday.com/isa/edv35n8.htm>. Accessed July 15, 2001.
- 2000c. Special E- Journals Issue! *Information Today, Inc. Home Page - ISA Editorials* 35(8). URL: <http://www.infotoday.com/isa/edv35n8.htm>. Accessed July 15, 2001.
- 2000d. Letter to the Editor. *Online Information Review* 24(1): 93-95.
- Haynes, R. Brian. 1987. A Proposal for More Informative Abstracts of Clinical Articles. *Annals of Internal Medicine* (106): 598-604.
- . 1993. More Informative Abstracts: Current Status and Evaluation. *Journal of Clinical Epidemiology* 46(7): 595-597.

- Haynes, R. Brian, Cynthia D. Mulrow, Edward J. Hutn, Douglas G. Altman, and Martin J. Gardner. 1996. More Informative Abstracts Revisited. *The Cleft Palate-Craniofacial Journal* 33(1): 1-9.
- Heller, M. B. 1990. Structured Abstracts. *Annals of Internal Medicine* 113(9): 722.
- Hern, Elizabeth. 2001. Strong Belief. *Information World Review*. June, No. 110, p.10.
- Hernon, Peter, and Cheryl Metoyer-Duran. 1992. Literature Reviews and Inaccurate Referencing: An Exploratory Study of Academic Librarians. *College & Research Libraries* 53(5): 499-512.
- Hightower, Christy, and Robert Schwarzwald. 1991. A Comprehensive Look at Materials Science Database. *Database* 14(2): 42-53.
- Hitti, Angela. 1995. Secondary Publishing in Changing Times: Profile of Cambridge Scientific Abstracts. *Publishing Research Quarterly* 11(3): 80-82.
- Holt, Janifer, and Karen A. Schmidt,. 1995. CARL UnCover2 or Faxon Finder? A Comparison of Articles and Journals in CARL Uncover2 and Faxon Finder. *Library Resources & Technical Services* 39(3): 221-228.
- Hood, William and Conception S. Wilson. 1994. Indexing Terms in the LISA Database on CD-ROM. *Information Processing & Management* 30(3):327-342.
- Hurst, Jill Ann. 1999. *DIALOGWeb/FT*. Woodstock, GA: Hermograph Press.
- . 2000. DIALOGWeb Under the Microscope. *EContent*, 23(3):35-38.
- Jacsó, Péter. 1989. Directory of Library and Information Professionals on CD-ROM. *Laserdisk Professional* 2(4):63-73.
- . 1991a. Coverage and Accessibility in Ulrich's Plus and EBSCO-CD. *The Serials Librarian* 18(1/2): 1-35.
- . 1991b. Data Transfer Capabilities of CD-ROM Software- Part 1. *CD-ROM Professional* 4(2): 63-66.
- .1991c. Data Transfer Capabilities of CD-ROM Software-Part 2. *CD-ROM Professional* 4(3): 61-66.
- . 1992a. *CD-ROM Software, Dataware, and Hardware: Evaluation, Selection, and Installation*. Enlewood, CO: Libraries Unlimited.
- . 1992b. What Is in a(n) (Up)date? Currency Test Searching of Databases. *Database* 15(3): 28-33.
- . 1993a. A Proposal for Database "Nutrion and Ingredient" Labeling. *Database* 16(1): 7-9.
- . 1993b. Searching for Skeletons in the Database Cupboard: Part 1. Errors of Omission: *Database* 16(1): 38-49.
- . 1993b. Searching for Skeletons in the Database Cupboard: Part 2. Errors of Commission: *Database* 16(2): 30., 32-36.
- . 1995. Testing the Quality of CD-ROM Databases. In: *Electronic Information Delivery: Ensuring Quality and Value*, edited by Reva Basch. Brookfield, VT: Gower, 141-168.
- . 1996. Watching Your Online Bottom Line. *Online* 20(4): 50-51.
- . 1997a. Content Evaluation of Databases. In: *Annual Review of Information Science and Technology*, vol.32., chap.5., edited by Martha E. Williams. Chicago: American Society for Information Science, 231-267.

- . 1997b. Information Science Abstracts. *Link-Up* 14(3): 3, 11.
- . 1997c. Jacsó Drops the Bruno Magli Shoes. *Database* 20(3): 6.
- . 1997d. Péter's Picks and Pans. ASCE's Civil Engineering Database. *Database* 20(5): 80-81., 84-85.
- . 1997e. Péter's Picks and Pans: Information Science Abstracts. *Database* 20(1): 86-87.
- . 1998a. Analyzing the Journal Coverage of Abstracting/Indexing Databases at Variable Aggregate and Analytic Levels. *Library & Information Research* 20(2): 133-151.
- . 1998b. Péter's Picks and Pans: EconLit Index. *Database* 21(6): 70-72.
- . 1998c. Péter's Picks and Pans: Mental Health Abstracts on CD-ROM and Online. *Database* 21(1): 79-82.
- . 1998d. Publishing Textual Databases on the Web. *Information Today* 15(9): 47-48.
- . 1998e. Péter's Picks and Pans: World Databases. *Database* 21(5): 80-82.
- . 1999a. Savvy Searching: Database Currency. *Online Information Review* 23(6): 345-348.
- . 1999b. Savvy Searching Starts with Browsing. *Online & CD-ROM Review* 23(3): 169-172.
- . 1999c. Péter's Picks and Pans: PASCAL. *Database* 22(2): 70.
- . 2000a. Be Savvy! Sometimes the Free Resources Are better. *Computers in Libraries* 20(5): 56-58.
- . 2000b. Endangered Database Species - December 2000. *Information World Review*. No. 166 (December), p. 72-73.
- . 2000c. Savvy Searching: Accuracy. *Online Information Review* 24(1): 90-93.
- . 2000d. Savvy Searching: Database Source Coverage: Myth and Reality. *Online Information Review* 24(6): 450-453.
- . 2000e. Savvy Searching: Digital Journal Lists. *Online Information Review*. 24(4): 337-339.
- . 2000f. Encyclopedia of the Orient. *Gale Group Home-Page- Péter's Digital Reference Shelf*. URL: <http://www.galegroup.com/servlet/HTMLFileServlet?inprint=9999®ion=7&fileName=reference/archive/200007/orient.html>. Accessed July 15, 2001.
- . 2000g. Bowker's Complete Video Directory. *Gale Group Home-Page- Péter's Digital Reference Shelf*. URL: <http://www.galegroup.com/servlet/HTMLFileServlet?inprint=9999®ion=7&fileName=reference/archive/200007/bowker.html>. Accessed July 15, 2001.
- . 2000h. Péter's Picks and Pans: Davis Free Internet Encyclopedia. *EContent* 23(3): 82-85. Lásd még URL: <http://www2.hawaii.edu/~jacso/extra/picks-pans/davis.htm> Accessed July 15, 2001.
- . 2000i. Péter's Picks and Pans: Population Demographics Database. *EContent* 23(4): 83-86. Lásd még URL: <http://www2.hawaii.edu/~jacso/extra/picks-pans/demographics/demographics.htm> Accessed July 15, 2001.
- . 2001a. Awesome Library. *Gale Group Home-Page - Péter's Digital Reference Shelf*. URL: <http://www.galegroup.com/servlet/HTMLFileServlet?inprint=9999®ion=7&fileName=reference/archive/2000103/awesome.html>. Accessed July 15, 2001.

- . 2001b. Péter's Picks and Pans: TheReference. *EContent* 25(3): 89-92. Lásd még URL: http://www2.hawaii.edu/~jacso/extra/picks-pans/thereference/thereference_files/frame.htm Accessed July 15, 2001.
- . 2001c. Scirus Biz 'R Us -Elsevier's Science Search Engine. 18(56): 34-35. *Information Today* 18(6): 34-35. Lásd még URL: <http://www2.hawaii.edu/~jacso/extra/infotoday/scirus/scirus.html> Accessed July 15, 2001.
- Jacsó, Péter, and F. Wilfrid Lancaster. 1999. *Build Your Own Database*. Chicago, IL: American Library Association.
- Jacsó, Péter, and Judit Tiszai. 1995. Now Featuring ... Movie Databases. *Database* 18(1): 24-24., 26-28., 30-32.
- Jaguszewski, Janice M., and Jody L. Kempf. 1995. Four Current Awareness Databases: Coverage and Currency Compared. *Database* 18(1): 33-44.
- Johnson, David H., and David R. Bevan. 1996. Structured Abstracts. *Canadian Journal of Anesthesia* 43(1): 1-3.
- Johnson, Susan. 1999. Using Information Technology to Improve Collection Management. *Library Computing* 18(2): 98-104.
- . 2000. Személyes közlés az ABI/INFORM tezaurusának létrehozásáról.
- Katritzky, Alan R., Stephen J. Cato, and James A. Deyrup. 1993. Comparison of the Scope, Timeliness and Quality of Chemical Abstracts from VINITI and CAS. *Journal of Information Science* 19(3): 199-210.
- Katz, William A. 1987. *Introduction to Reference Work*. New York: McGraw-Hill.
- Killion, Vicki J. 1995. RNdex Top 100: A Quality-filtered Database for Nursing Research. *Medical Reference Services Quarterly* 14(3): 1-11.
- Kister, Kenneth F. (1986). *Best Encyclopedias: Guide to General and Specialized Encyclopedias*. Phoenix, AZ: The Oryx Press.
- Klemmer, E. T., and Gregory R. Lockhead. 1962. Productivity and Errors in Two Keying Tasks: A Field Study. *Journal of Applied Psychology* 46(6): 401-408.
- Koehler, Wallace et al. 2000. A Profile in Statistics of Journal Articles: Fifty Years of American Documentation and the Journal of the American Society for Information Science. *Cybermetrics* 4(1). URL: <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p3.html> Accessed July 15, 2001.
- Kohl, David F., and Charles H. Davis. 1985. Ratings of Journals by ARL Library Directors and Deans of Library and Information Science Schools. *College & Research Libraries* 46(1): 41-47.
- Kulkarni, M. V. K. Gupta, and Viswanathan. 1996. An Automatic Index Generation Retrieval System for Bibliographic Databases Developed at the Indian National Scientific Documentation Center. *Program* 30 (January): 65-72.
- LaBorie, Tim, Michael Halperin, and Howard D. White. 1985. Library and Information Science Abstracting and Indexing Services: Coverage, Overlap, and Context. *Library & Information Science Research* 7(2): 183-195.
- LaGuardia, Cheryl. 1991. Philosopher's Index On Disc. *CD-ROM Professional* 4(6): 119-120.

- Lancaster, F. Wilfrid. 1971. The Evaluation of Published Indexes and Abstract Journals. *Bulletin of Medical Library Association* 59(3): 479-494.
- Lancaster, F. Wilfrid (1979). *Information Retrieval Systems: Characteristics, Testing & Evaluation*. New York: John Wiley & Sons, Inc.
- . 1979. *Information Retrieval Systems: Characteristics, Testing, and Evaluation*. New York: John Wiley.
- . 1991. *Indexing and Abstracting in Theory and Practice*. Champaign, IL: University of Illinois, Graduate School of Library and Information Science.
- . 1998. *Indexing and abstracting in Theory and Practice*. 2d ed. Champaign, IL: University of Illinois at Urbana-Champaign.
- Lavin, M. R. 1998. A Clash of Titans - Comparing America's Most Comprehensive Business Directories. *Database* 21(3): 44-48.
- Lawrence, Barbara, and Tony Lenti. 1995. Application of TQM to the Continuous Improvement of Database Production. In: *Electronic Information Delivery: Ensuring Quality and Value*, edited by Reva Basch. Brookfield, VT: Gower, 69-87.
- Layne, Sara Shatford, 1994. Some Issues in the Indexing of Images. *Journal of the American Society for Information Science*. 45(8): 583-588.
- Leininger, Kurt. 2000. Interindexer Consistency in PsycINFO. *Journal of the American Society for Information Science* 32(1): 4-8.
- Leonard, Lawrence E. 1977. Inter-Indexer Consistency Studies, 1954-1975: A Review of the Literature and Summary of Study Results. *University of Illinois Graduate School of Library Science Occasional papers*, no. 131.
- Markey, Karen. 1984. Interindexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials. *Library & Information Science Research* 6(2): 155-177.
- Martin, Sabine, and Günter Bergerhoff, 1991. Chemical Abstracts Online: A Study of the Quality of Controlled Terms. *Journal of Chemical Information and Computer Science* 31(1): 147-152.
- Martyn, John, and Margaret Slater. 1964. Test on Abstract Journals. *Journal of Documentation* 20(4): 212-235.
- McCain, Katherine W., Howard White, and Belder C. Griffith. 1987. Comparing Retrieval Performance in Online Data Bases. *Information Processing & Management* 23(6): 539-553.
- McCormick, Edith. 1991. AI Aside-Ideas: The Dirty Database Test (Jeffrey Beall). *American Libraries* 22(3): 197.
- Milas-Bracovic, Milica, and Jasenka Zajec. 1989. Author Abstracts of Research Articles Published in Scholarly Journals in Croatia (Yugoslavia): An Evaluation. *Libri* 39(4): 303-318.
- Miller, Uri. 1996. The Sport Database: Some Comments. *Online & CD-ROM Review* 20(2): 67-74.
- Milstead, Jessica L. 1994a. Needs for Research in Indexing. *Journal of the American Society for Information Science* 45(8): 577-582.
- . 1994b. *ASIS Thesaurus of Information Science and Librarianship*. Medford, NJ: Learned Information.

- Mintz, Anne P. 1990. Quality Control and the Zen of Database Production. *Online* 14(6): 15-23.
- . 1995. Quality Issues in Information Retrieval: a Publisher Perspective. In: *Electronic Information Delivery: Ensuring Quality and Value*, edited by Reva Basch. Brookfield, VT: Gower, 47-58.
- Moed, H. F., and M. Vriens. 1989. Possible Inaccuracies Occuring in Citation Analysis. *Journal of Information Science: Principles & Practice* 15(2): 95-107.
- Narine, L., D. S. Yee, T. R. Einarson, and A. L. Ilersich. 1991. Quality of Abstracts of Original research Articles in CMAJ in 1989. *CMAJ* 144(4): 449-453.
- Nicholls, Paul. 1998. Great Expectations: A Tale of Two or Three Directories. *Searcher* 6(8): 66-75.
- Norton, Nancy Prothro. 1981. Dirty Data: A Call for Quality Control. *Online* 5(1): 40-41.
- Ojala, MaryDee. 1992. The Prestige Factor in Electronic Business Journals. *Database* 15(5): 89-92.
- O'Neill, Edward T., and Diane Vizine-Goetz. 1988. Quality Control in Online Databases. In: *Annual Review of Information Science and Technology*, vol. 23, edited by Martha E. Williams. Amsterdam, The Netherlands: Elsevier Science Publishers for the American Society for Information Science, 125-156.
- Orenstein, Ruth M. 1989. The Fullness of Full Text; Survey of Coverage, Currency and Enhancement of 13 Services. *Database Searcher* 5(September): 21-27.
- . 1993. 'How Full is Full' Revisited: A Status Report on Searching Full-Text Periodicals. *Database* 16(5): 14-23.
- Pagell, Ruth A. 1987. Searching Full-Text Periodicals: How Full Is Full? *Database* 10(5): 33-36.
- Pandit, Idrisa 1993. Citation Errors in Library Literature: A Study of Five Library Science Journals. *Library & Information Science Research* 15(2): 185-198.
- Pao, Miranda Lee. 1989. Importance of Quality Data for Bibliometric Research. In: *Proceedings of the 10th National Online Meeting*, edited by Martha E. Williams, May 9-11. 1989. Medford, NJ: Learned Information. 321-327.
- Pedersen, Martin. 1992. Texas Schoolbook Massacre: 5200 Errors Found in 10 History Books. *Publishers Weekly* 239 (March 2): 11.
- Pemberton, Jeff. 1983. The Linear File: The Dark Side of Online Information -- Dirty Data. *Database* 6(4): 6-8.
- Perry, Stephen, and Lutishoor Salisbury. 1995. Access to Information in Both CitaDel and FirstSearch: A Comparative Study of Dissertation Coverage. *Information Technology and Libraries* 14(1): 17-29.
- Pollock, J. J., and A. Zamora. 1975. Automatic Abstracting Research at Chemical Abstracts Service. *Journal of Chemical Information and Computer Sciences* 15(4): 226-232.
- Quint, Barbara. 1989. Caveat Searcher: Liars, Damned Liars, and Statisticians. *Database Searcher* 5(9): 36-37.
- . 1995. Better Searching Through Better Searcher. In: *Electronic Information Delivery: Ensuring Quality and Value*, edited by Reva Basch. Brookfield, VT: Gower, 99-116.

- Reich, Phyllis, and Erik J. Biever. 1991. Indexing Consistency: The Input/Output Function of Thesauri. *College & Research Libraries* 52(4): 336-342.
- Salager-Meyer, Francoise. 1991. Medical English Abstracts: How Well Are They Structured? *Journal of the American Society for Information Science* 42(7): 528-531.
- Sievert, MaryEllen C., and Donald E. Sievert. 1991. Online Searching in Philosophy: A Comparison of Philosopher's Index and FRANCIS. *Online Review* 15(2): 63-76.
- Sievert, MaryEllen C., and Mark J. Andrews. 1991. Indexing Consistency in Information Science Abstracts. *Journal of the American Society for Information Science* 42(1): 1-6.
- Smith, Linda C. 1981. Citation Analysis. *Library Trends* 30(1): 83-106.
- Snow, Bonnie. 1998. Alternative Medicine Information Sources. *Database* 21(3): 18-19.
- Sodha, R. J. 1993. Trends in Biomedical Publications: US and Japanese Authors in US Journals and European Journals. *Journal of Information Science* 19(1): 71-73.
- Soergel, Dagobert. 1994. Indexing and Retrieval Performance: The Logical Evidence. *Journal of the American Society for Information Science* 45(8): 589-599.
- Soremark, Gun. 1990. MEDLINE Versus EMBASE: Comparing Search Quality. *Database* 13(6): 66-67.
- Sparck Jones, Karen. 1973. Does Indexing Exhaustivity Matter? *Journal of the American Society for Information Science* 24(6): 313-316.
- Sparck Jones, Karen, and C. J. Van Rijsbergen. 1976. Information Retrieval Test Collections. *Journal of Documentation* 32(1): 59-75.
- Stieg, Margaret, and Joan L. Atkinson. 1988. Librarianship Online: Old Problems, No New Solutions. *Library Journal* 113(16): 48-59.
- Svenonius, Elaine, and Dorothy McGarry. 1993. Objectivity in Evaluating Subject Heading Assignment. *Cataloging & Classification Quarterly* 16(2): 5-40.
- Sweetland, James H. 1989. Errors in Bibliographic Citations: A Continuing Problem. *The Library Quarterly* 59(4): 291-304.
- Taddio, Anna, Tapas Pain, Frank F. Fassos, Heater Boon, A. Lane Ilersich, and Thomas R. Einarson. 1994. Quality of Non-structured and Structured Abstracts of Original Research Articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*. *Canadian Medical Association Journal* 150(10): 1611-1615.
- Tenopir, Carol. 1982. Evaluation of Database Coverage: A Comparison of Two Methodologies. *Online Review* 6(5): 423-441.
- . 1992. Evaluation Criteria for Online, CD-ROM. *Library Journal* 117(4): 66-67.
- . 1995. Priorities of Quality. In: *Electronic Information Delivery: Ensuring Quality and Value*, edited by Reva Basch. Brookfield, VT: Gower, 119-139.
- . 1997. Reading Vendor Literature. *Library Journal* 122(June 1): 35-36.
- Tenopir, Carol, and Gerald Lundeen. 1988. *Managing Your Information*. New York, NY: Neal-Schuman Publishers.
- Tenopir, Carol, and Péter Jacsó. 1993. Quality of Abstracts. *Online* 17(3): 44, 46-48.

- Tenopir, Carol, and Ralf Neufang. 1995. Electronic Reference Options: Tracking the Changes. *Online* 19(4): 67-73.
- Thomas, Sarah E. 1990. Bibliographic Control and Agriculture. *Library Trends* 38(3): 542-561.
- Tjoumas, Renee, and Virgil L. P. Blake. 1992. Faculty Perceptions of the Professional Journal Literature: Quo Vadis? *Journal of Education for Library and Information Science* 33(3): 173-194.
- Tonta, Yasar (1991). A Study of Indexing Consistency between Library of Congress and British Library Catalogers. *Library Resources & Technical Services* 35(2): 177-185.
- Turner, James M. 1995. Comparing User-Assigned Terms with Indexer-Assigned Terms for Storage and Retrieval of Moving Images: Research Result. In: *Proceedings of the 58th Annual Meeting of the American Society for Information Science*. Vol. 32, edited by Tom Kinney. Medford, NJ: Information Today, 9-12.
- Turtle, Mary R., and William Robinson. 1974. The Relationship between Time Lag and Place of Publication in Library and Information Science Abstracts and Library Literature. *RQ* 14(1): 28-31.
- Ubico, Rafael E., John A. Baily, and Pamela J. Weaver. 1995. Statistical Analysis of the TULSA Database, 1965-1994. *Publishing Research Quarterly* 11(3): 138-144.
- Van Camp, Anne J. 1993. Online Sources for Alternative Medicine Information. *Database* 16(5): 100-103.
- Watson, Maureen Martin, and Richard Perrin. 1994. A Comparison of CINAHL and MEDLINE CD-ROM in Four Allied Health Areas. *Bulletin of Medical Library Association* 82(2): 214-216.
- Way, Harold E. 1988. The BRS 1988 Annual Meeting. *Database Searcher* 4(5): 15-22.
- Weston, E. Paige, and Diane S. Lauderdale. 1988. How Do We Learn What a Database Includes? A Case Study Using Psychology Dissertation. *RQ* 28(1): 35-41.
- Whitney, Gretchen. 1990. *Language Distribution in Databases: An Analysis and Evaluation*. Metuchen, NJ: Scarecrow Press; 379.
- . 1992. Access to Third World Science in International Scientific and Technical Bibliographic Databases. *Scientometrics* 23(1): 201-219.
- Whitney, Gretchen. 1993. Patterns of Authorship in Major Bibliographic Databases: The European Region. *Scientometrics* 26(2): 275-292.
- Williams, Martha E. 1990. Highlights of the Online Database Industry and the Quality of Information and Data. In: *Proceedings of the 11th National Online Meetin*, ; edited by Martha E. Williams, 1990 May 1-3. New York, NY. Medford, NJ: Learned Information, Inc.: 1-4.
- Williams, Martha E., and Laurence Lannom. 1981. Lack of Standardization of the Journal Title Data Element in Databases. *Journal of the American Society for Information Science* 32(3): 229-233.
- Yannakoudakis, E. J., and D. Fawthrop. 1983. The Rules of Spelling Errors. *Information Processing & Management* 19(2): 87-99.