

Az említett számítógépes rendszerekhez hasonlóan modern információhordozók lehetővé teszik nagy mennyiségű adat kis helyen történő tárolását. A mikroformátumokat és az egyéb hagyományos információhordozókat egyre jobban háttérbe szorítja a lézerlemez (CD-ROM, CD-Audio, CDI, CDV), a videolemez, az optikai leolvasó (DON), és még folytathatnám.

Mindezek az eszközök, módszerek, eljárások nem ismeretlenek nálunk, de széles körű elterjedésük feltehetőleg még sokáig várat magára. A távolságot segíti áthidalni az évről évre növekvő számú ösztöndíjas képzése, a kutatómunka, tanulmányút, melyből most egyet én magam is sikerrel végigküzdöttem. Számomra mindenképpen megérte.

Bruckner Nóra

(Országos Műszaki Információs Központ és Könyvtár)

A könyvtár- és tájékoztatástudományi témák szóródása a bibliográfiai adatbázisokban

A könyvtár- és tájékoztatástudomány irodalmának jó része található más tudományágak adatbázisaiban elrejtve. Pl. az idős emberek könyvtári igényeiről szóló cikk ugyanolyan valószínűséggel fordulhat elő egy geriátriai, mint egy könyvtári folyóiratban. Ha tehát csak egy-két adatbázisra szorítkozik a kereső, releváns tételektől eshet el.

Egy kutatás során adatgyűjtést folytattunk annak kiderítésére, hogy a *könyvtár- és tájékoztatástudományi témák* milyen módon fordulnak elő a nem könyvtártudományi gyűjtőkörű adatbázisokban. A vizsgálat a téma szétszórt voltára összpontosított, nem pedig a feldolgozott folyóiratok körének átfedésére.

A LISA 1985-ös évfolyamából 200, véletlenszerűen kiválasztott referátum első tárgyszavát és annak tágabb kategóriáját (azaz a tartalomjegyzék fejezetcímét) vettük kiindulópontként. A duplumok kiszűrése után 128 kifejezéssel dolgoztunk a következő hét kategóriában: könyvtárügy, könyvtárak, könyvtári dokumentumok, szervezés és vezetés, olvasószolgálat, feldolgozó munka, információ tárolás és -keresés.

A LISA és a teljes szövegű adatbázisok kihagyásával 55 elemezhető adatbázis maradt. A BRS ún. CROSS-szolgáltatását felhasználva minden egyes tárgyszóra adatbázisonként megkaptuk a találatok számát, ami együttesen megközelítette a 11 milliót.

Az 55 adatbázisban és 7 témakörben kapott találatsszámok mátrixát a *várható találati arány* mátrixává alakítottuk át a következő képlet szerint:

$$E = (p \cdot tp) / (rt \cdot ct),$$

ahol E a várható találati arány, p a találatsszám, tp a találatsszámok összege, rt a sorösszeg, ct az oszlopösszeg. E értéke mindenütt 1 lenne, ha a várható találati arány minden adatbázisra és minden témakörre egyforma lenne.

Az E értékek mátrixát klaszteranalízisnek vetettük alá. A klaszteranalízis olyan matematikai technika, amely a vizsgált objektumokat (esetünkben az adatbázisokat) bizonyos hasonlósági kritériumok (esetünkben az egyes témakörök várható találati arányai) szerint csoportokba rendezi.

Az elemzés eredményeképpen 45 adatbázis jól besorolható volt egyik vagy másik csoportba, és csak 10 adatbázis volt besorolhatatlan.

A klaszteranalízis eredményét az adatbázisokban végzett tényleges keresésekkel is ellenőriztük. Mivel az ötletszerűen kiválasztott 128 keresőkifejezéssel értelmes keresés nem végezhető, egyszerűsített eljárást alkalmaztunk. A hét kategória közül kettőt (könyvtárügy, könyvtárak) egyesítve kiemeltünk fő kategóriaként. A többire kategóriánként VAGY operátorral kapcsoltuk össze a keresőkifejezéseket, majd az így kapott összetett kifejezést ÉS operátorral kapcsoltuk a fő kategória hasonlóan képzett kifejezéséhez. Ezekkel a logikai kifejezésekkel minden adatbázisban és mind az öt megmaradt témakörben elvégezve a keresést, ismét egy találatsszámmátrixot kaptunk (az eredeti mátrix részhalmazát), és ezt is a várható találati arányok mátrixává alakítottuk. Így egymással összehasonlítható két mátrixot kaptunk. A két eredmény az öt témakör közül négyben jól megközelítette egymást.

Végző lépésként a különböző adatbázisokban kapott találatok relevanciáját vizsgáltuk meg. Anyagi korlátok miatt ezt csak (statisztikailag nem releváns) viszonylag kis mintán végezhattük. A kb. 169 ezer találat kb. 5%-át nyomtattuk ki és elemeztük, adatbázisonként legalább 10, és legfeljebb 100 referátumot, összesen 2655-öt. Három szempont, a hely, a tevékenység és a tárgy szerint 0-tól 5-ig pontoztuk a dokumentumokat, összesen maximálisan 15 pontot érhetett el egy-egy dokumentum. A hely és a tevékenység szempontját úgy vizsgáltuk, hogy bizonyos szótövek előfordulását kerestük a címben vagy a referátumban, a tárgy relevanciájáról pedig szubjektíve döntöttünk. Az 5 pontnál többre értékelt dokumentumokat tekintettük relevánsnak. Ilyen értékelési mód mellett a kis mintában kapott találatoknak nagyjából a fele bizonyult relevánsnak.

A várákosoknak megfelelően az ERIC (oktatásügyi), az NTIS (műszaki) és az MESH (orvostudományi) adatbázis bizonyult a legzárdaabbnak könyvtártudományi anyagban. A könyvtártudományi kutatók figyelmére méltó még az EMED (biológiai-orvostudományi), a BIOL (biológiai), az INSP (fizikai),

az LLBA (nyelvtudományi), a SOCA (szociológiai) és az NCMH (pszichoterápiái) adatbázis. (Ezek 9–69 ezer találatot adtak a tényleges keresések során.) Néhány adatbázis kevesebb könyvtártudományi anyagot tartalmaz ugyan (pl. a DRSC gyógyszer-tudományi, az NOOZ hírlapcikk-, és az INFO üzleti adatbázis), ezek viszont magas pontossági értékekkel tűntek ki.

Sok könyvtár- és tájékoztatástudományi anyag szerepel tehát "elszórva" a nem könyvtártudományi adatbázisokban. Ha a CROSS-kereséssel kapott 10 millió tételnek csak 0,25%-a releváns, akkor is több

mint 25 ezer ilyen tétel van a BRS adatbázisaiban. Ezek közül sok lehet a LISA adatbázisban is, az átfedés vizsgálata azonban egy további kutatás tárgya lehet majd.

/YERKEY, N.–GLOGOWSKI, M.: Scatter of library and information science topics among bibliographic databases. = Journal of the American Society for Information Science, 41. köt. 4. sz. 1990. p. 245–253./

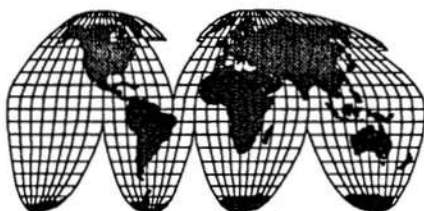
(Hegyközi Ilona)

ALEPH

AUTOMATED LIBRARY
EXPANDABLE PROGRAM

EX LIBRIS

AUTOMATES THE



of **LIBRARIES**
& **INFORMATION CENTERS**

SYSTEMS IN:

BRASIL
CZECH
DENMARK
GERMANY
GREECE
HUNGARY
ISRAEL
ITALY
PERU
POLAND
SLOVAKIA
SPAIN
SWEDEN
SWITZERLAND
TURKEY
USA

**NATIONAL- UNIVERSITY- MUNICIPAL
MUSEUMS - CORPORATE - SPECIAL**

UNIX: DEC, HP, IBM, ICL, SUN
VMS: DEC

DISTRIBUTED IN HUNGARY BY:

IBR GENERAL Ltd

1539 BUDAPEST, BOX 658
TEL/FAX: +36-1-156 5062

Ex Libris S.A. B.P. 1163, L-2210 Luxembourg