

elő bármikor a konkrét feldolgozott szövegekben (bár az is igaz, hogy ezek a szóalakok messze nem azonos valószínűséggel fordulnak elő). A Microsoft Word magyar helyesírás-ellenőrzőjének néha elég furcsa javaslataiból is tapasztalhatjuk, hogy nem könnyű még ez a nyelvelírás szempontjából viszonylag egyszerű feladat sem. A *szintaktikai elemzésben* is nehézséget jelent a magyar nyelv szabad szórendje az igencsak kötött angol szórendhez képest. Ezért bizonyos fentebb felsorolt technikák nehezen, vagy nem alkalmazhatók az agglutinatív nyelvekre.

Vannak olyan magyar vonatkozású háttér munkálatok, amelyekre keveset és kevesen (vagy éppen egyáltalán nem) hivatkoznak. Ilyenek a Papp Ferenc által szerkesztett *Szóvéghmutató szótár* (Bp., Akadémiai Kiadó, 1968), a Füredi M.–Kelemen J.: *A mai magyar szépróza gyakorisági szótára* (Bp., Akadémiai Kiadó, 1989), s e kettőnek számítógépes adatbázisai (Debreceni tezaurusz, GyakNagy), amelyeknek jó időben „public domain”-nek nyilvánított, több száz emberévnyi munkával elért eredményeit messzemenően bedolgozták termékeikbe a magyar helyesírás-ellenőrző programok készítői. Ezek az 1960-as, 1970-es években manuális kódolással elvégzett alaputatók és számítógépes feldolgozásaik adtak megbízható háttérrel a mai magyar nyelv szótári szerkezetének, és a nagyobb lélegzetű irodalmi szövegek statisztikai, morfológiai, esetenként szintaktikai és szemantikai sajátosságainak megértéséhez [5].

Bár a nyolcvanas évek elején teljesen megszűnt a kvantitatív nyelvészeti kutatások támogatása, ma már remény van arra, hogy a Nyelvtudományi Intézetben összegyűlt beszélt nyelvi anyagok, valamint az Akadémiai Nagyszótár céljaira kódolt anyagok alapján – felhasználva az azóta is fejlődő nyelvészeti elemző technikákat – különböző műfajokra, közöttük a tudományos és tudomány-népszerűsítő szövegekre is lehessen következte-

téseket levonni, ezzel is segítve a könyvtári dokumentum-visszakereső szolgáltatásokat.

Irodalom

- [1] LUHN, H. P.: A statistical approach to mechanised encoding and searching of literature information. = IBM Journal Research and Development, 1. köt. 4. sz. 1957.
LUHN, H. P.: The automatic creation of literature abstracts. = IBM Journal Research and Development, 2. köt. 2. sz. 1958. p. 159–165.
STEVENS, M. E.: Automatic indexing – A state of the art report. National Bureau of Standards, Washington, D. C., 1965.
FANGMEYER, H.: Semi-automatic indexing: state of the art. Paris, 1974.
- [2] GUIRAUD, P.: Les caractères statistiques du vocabulaire. Essai de méthodologie. Presses Universitaires, Paris, 1954.
HERDAN, G.: Type/token mathematics. 's Gravenhage, 1960.
HERDAN, G.: Quantitative linguistics. London, 1964.
NAMBIAR, K. K.: Theory of search engines. 2001. január 8.
SHANNON, C. E.: The mathematical theory of communication. University of Illinois Press, Chicago, 1963.
- [3] Dublin Core metadata element set reference description, version 1.1. 1999. 07. 02. http://purl.org/dc/documents/proposed_recommendations/pr-dces-19990702.htm
- [4] HUSSAM, A. et al.: Semantic highlighting: enhancing search engine display and Web document interactivity. = Human-Computer Interaction – INTERACT – 1999. II. köt. Szerk.: S. Brewster–A. Cawsey–G. Cockton. The British Computer Society. IFIP TC. 1.3. 1999. p. 85–86.
- [5] FÜREDI M.: Összesített adatok a szépróza gyakorisági szótáráról. Igék és igezármazékok. = Magyar Nyelv, 82. köt. 2. sz. 1986. p. 190–198.

(Füredi Mihály)

Nyomtatott és elektronikus folyóiratok kezelési költségeinek összehasonlító vizsgálata

Sok felsőoktatási és tudományos könyvtár van azon az úton, hogy folyóirat-gyűjteményét nyomtatott formátumról elektronikusra váltsa. Ennek következtében – kivéve a legnagyobb könyvtárakat – az elektronikus folyóirat-gyűjtemények mérete

messze meghaladja a nyomtatottnál valaha is tapasztalt nagyságot. Ez természetszerűen kihat a folyóiratokkal végzett könyvtári műveletekre, és a személyzettel szemben támasztott követelményekre is. Rendszerszerű megközelítésben különös

súlyt nyer az elektronikus archiválás megoldatlansága.

Az *Electronic-Archiving Initiative* – amelyet a JSTOR 2003-ban indított útjára – felmérést kezdeményezett 11 kis, közepes és nagy könyvtárban azzal a céllal, hogy az elektronikus forrásokra történő váltás milyen hatással van a könyvtári költségekre, különösen az előfizetési díjak mellett jelentkező kiadásokra (nevezzük ezeket kezelési költségeknek). A kutatás eredményeiről részletesebben a *D-LIB Magazine* 2003. 1. száma ad hírt, a teljes jelentés 2004-ben jelenik meg *Schonfeld, R. C.–King, D. W.–Okerson, A.–Fenton, E. G.* tollából *The nonsubscription side of periodicals: changes in library operations and costs between print and electronic formats* címen.

A vizsgált 11 felsőoktatási könyvtárban 100-tól 23 000-ig terjed az előfizetett nyomtatott folyóiratok száma; az elektronikusoké 5000-tól 20 000-ig. Az előfizetési költségek alsó értéke 250 000 dollár, felső határa 2,5 millió. Ezen összegen belül elektronikus folyóiratokra 15–80%-ot költenek.

A vizsgálat során az ún. életciklusmódszert alkalmazták: a hosszú távra szóló pénzügyi kötelezettségvállalást számították ki egy adott folyóiratcímre vonatkozóan egy adott év beszerzései alapján. E műveletet külön végezték el a nyomtatott és az elektronikus folyóiratokra nézve, így képesek voltak összehasonlítani a különböző formátumok kezelési költségeit is.

Megállapították, hogy az elektronikus formátummal kapcsolatos hosszú távú pénzügyi kötelezettség kevesebb, különösen a kisebb intézményekben. Bár különféle helyi tényezőket is figyelembe kell

venni, számos könyvtár számára valószínűleg csökkennek a kezelési költségek az elektronikus formátumra való áttérés következtében. Mivel azonban – különösen a kis és közepes könyvtárakban – feltehetően növekedni fog a gyűjtemény nagysága, a kezelési költségekben mutatkozó nyereséget felemészti a folyóiratcímek megnövekedett száma.

Habár a kezelési költségek nem csökkennek annyira, amennyire egyes optimista előrejelzések jósolták, a trend mégis az, hogy a kezelésre fordított kiadások csökkennek, ahogy a könyvtárak elektronikus folyóiratokra váltanak át. Mindenesetre számolni kell azzal, hogy bizonyos költségek, amelyeket a könyvtárak, illetve a kiadók viseltek, helyet cserélhetnek a két fél között (pl. a szerveren való tárolás).

Az archiválás megoldatlansága határozottan hátráltatja az elektronikus folyóiratokra való áttérést. Remélhető, hogy az a megtakarítás, amely a tanulmány szerint a kezelési költségekben mutatkozik, hozzá fog járulni az archiválás megvalósításához. Válaszra várnak azonban a következő kérdések: Elegendő lesz-e a várható megtakarítás általánosan elfogadható archiválási eljárások kidolgozására, vagy más forrásokból kell fedezetet találni erre? Késlelteti-e az archiválás megoldatlansága az áttérést az elektronikus formátumra? Vagy felgyorsul az áttérés, és az archiválás problémájától eltekintenek?

/SCHONFELD, Roger C.–FENTON, Eileen Gifford:
Library periodicals expenses: comparison of nonsubscription costs of print and electronic formats on a life-cycle basis. = CLIR issues, 39. sz. 2004. május–június, p. 1–4./

(Papp István)

Adatbányászat a könyvtárban

A *bibliomining* vagy könyvtárak számára végzett adatbányászat a könyvtári szolgáltatásból adatokat előállító adatbányászati és bibliometriai alkalmazás. (A bibliomining kifejezést a szerzők, S. Nicholson és J. Santon egy korábbi publikációjukban kívánták bevezetni az adatbányászat könyvtári alkalmazására. A cikkben a továbbiakban *könyvtári adatbányászatot* fogunk említeni.) A könyvtári adatbányászat a könyvtári folyamatokkal kapcsolatos adathalmaz feltárásának statisztikai és mintázatfelismerési eszköze, melynek célja a dön-

tés-előkészítés, illetve a szolgáltatások igazolása. (Az adatbányászat az ipari és kereskedelmi szektorban az utóbbi tíz évben bevetté vált tevékenység; a szervezet jelentős folyamatait elektronikus eszközökkel lefedő rendszerekben tárolt adatokat, illetve a közöttük lévő összefüggéseket tárja fel. Ennek célja lehet a tervezés, döntés-előkészítés, értékelés, kontrolling, minőségirányítás stb. Az ilyen rendszerek lehetővé teszik, hogy a vállalkozások azonosítsák a szolgáltatásukat igénybe venni nem kívánók – a lemorzsolódók – csoportját.