

Subject and Information Analysis

edited by **Eleanor D. Dym**

Subject and information analysis. Ed. by Eleonor D. Dym. New York, Basel, Dekker, cop. 1985. IX. 498 p. (Book in library and information science. 47.)

Az információkeresés új megközelítésben

A Dekker könyvkiadó publikációira mindig érdemes odafigyelnünk. A kiadó könyvtartudományi és informatikai sorozatának 47. köteteként jelent meg ez az új szemléletű mű, Eleonor D. Dym, a pittsburghi Egyetem Könyvtartudományi és informatikai tanszékének professzora szerkesztésében. Dym professzor szándékának megfelelően egy tankönyvet pótló olvasókönyvet állított össze, amikor a hazánkban is jól ismert 23 amerikai és angol szakíró tanulmányát egybeszerkesztette. A szerzők között olyan ismert neveket találunk mint D.J. Foskett, J.E. Daily, A.A. Kent, B.C. Vickery stb. A szakértő olvasó számára a tanulmányok egyike-másika nem lesz ismeretlen, ezek ugyanis már más gyűjteményekben is megjelentek. A válogatás forrásai általában számítástechnikai, könyvtartudományi és informatikai szakencikopédiák, a válogatás célja pedig az volt, hogy a tartalmi feltárás és

az információkeresés új szemléletű megközelítését mutassák be. Az információkeresés szempontjából a bibliográfiai tételek formai és tartalmi ismérveit egyaránt vizsgálják.

Az egyes tanulmányokat lényegében hat nagyobb témakörbe csoportosították. Ezek: leíró katalogizálás, természetes nyelvű szövegek gépi elemzése, indexelés, terminológiai ellenőrzés, osztályozás és kivonatok készítése.

A hat témakör elé M.S. Stevens és B.C. Vickery bevezetői nyújtják azt az áttekintést, amely alapokat ad az egyes témák megértéséhez. Ezt követi a leíró katalogizálást tárgyaló fejezet J.E. Daily tollából. Magam részéről úgy vélem, hogy helyesebb lett volna ezt a kérdést a terminológiai ellenőrzés témakörénél szerepeltetni, ugyanis a terminológiai ellenőrzés elsőként a leíró katalogizálás során valósul meg, amikor a dokumentumok adat-elemeit és formáját egységes alakban rögzítik.

A második témakör a természetes nyelvű szövegek feldolgozását, gépi elemzését tárgyalja. Az egyes fejezetek az alábbi témákat tárgyalják: tartalomelemzés, gépi fordítás, szövegfeldolgozás, automatikus gépi elemzés. E témakör legterjedelmesebb fejezete a *Tartalomelemzés* (57–109.p.) címet viseli. Közismert, hogy a tartalomelemzés a dokumentumokban foglalt közlések meghatározott célú elemzésére irányul. Nagyon leegyszerűsítve a tartalomelemzés bizonyos szimbolumok és kategóriák összeszámlálását és interpretációját jelenti. A számítógép számára „olvashatóvá” vált dokumentumok elemzése úgy történik, hogy a szintaktikailag kódolt szöveg összetételre kerül a gép memóriájában tárolt kategó-

ria-szótárral. E művelet részletei után érdeklődőknek C.E. Hicks, J.E. Rush és S.M. Strong számítógépes szakemberek tanulmányának elolvasását ajánljuk.

A tartalomelemzést azért is tartom izgalmas problémának, mert úgy vélem, hogy az ezen a területen már alkalmazásra került módszerek az információkeresésben is hasznosíthatók lehetnek. Az itt elérhető eredmények megmutatják, hogy milyen új lehetőségeink lesznek a dokumentumok tartalmának reprezentációjára és visszakeresésére. Ugyanis azazal, hogy feltártuk és részben már kiaknáztuk a kulcsszavak közötti szemantikai kapcsolatok adta lehetőségeket, elméletileg egy határhoz jutottunk el. Ezen a határon való túllépés az lehet, amikor a számítógép képes lesz a dokumentum tartalmának megértésére, azaz a közös összetevők felismerésére, egymástól nagyon távoli dolgok összekapcsolására. (Egyébként ebbe az irányba mutatnak a „mesterséges intelligenciával” kapcsolatos kutatások is.) Ezért talán nem túl nagy merészség leírni: az információkeresésben egy újabb „forradalom” következik be majd akkor, amikor az információkereső rendszerek nem egyszerűen bibliográfiai adatokat keresnek vissza, hanem a dokumentumok teljes szövegének tartalmát elemezve keresnek választ a felhasználó kérdéseire.

A gépi fordítással két tanulmány is foglalkozik (5. és 6. fejezetek). Ezek a fejezetek ismételtelen felhívják figyelmüket a nyelvészet és a dokumentáció kapcsolatára. A tanulmányokat olvasva elégedetten nyugtázhatjuk, hogy a nyelvfeldolgozás és a dokumentáció kapcsolatát országunkban is elég korán felismertük. (Gondoljunk csak itt *Petőfi* S. János, *Varga* Dénes és mások munkáira.) Elégedettségünkbe azonban némi keserűség is vegyül, amikor arra gondolunk, hogy a hatvanas évek végének, a hetvenes évek elejének lendülete megtört és a várt folytatás elmaradt.

Az indexelésről S. *Artandi*, Ch. L. *Bernier*, B. C. *Landry*, J.E. *Rush*, M.E. *Stevens* tanulmányai (9–12. fejezetek) adnak áttekintést. Ezek a tanulmányok ma már jól ismert kérdéseket tárgyalnak mint, pl.: az indexelő technikákat, számítógéppel előállítható permutált indexek problémáit továbbá az automatikus indexelés kérdéseit.

A *Terminológiai ellenőrzés* címet viselő témacsoportba sorolták be a szerzői nevek egységesített formájának amerikai gyakorlatát tárgyaló fejezetet (szerző: J.E. *Daily*), a tárgyszavazásról szóló két fejezet (szerzők: J.E. *Daily* és M.L. *Manheimer*), továbbá a tezaurusokkal foglalkozó tanulmányt (szerző: D.J. *Foskett*). Ez a csoportosítás a „terminológiai ellenőrzést” az input felől közelíti meg. Hiányoljuk, hogy a tanulmányok között egyetlen olyan sem akad, amely ugyanezt a problémát az output oldaláról vizsgálná, azaz a felhasználók visszakeresési stratégiáját venné szemügyre. Nyilvánvaló, hogy a felhasználók terminológiája előre nem határozható meg. Mindig lesznek olyan felhasználók akik a tárgyat átfogó módon közelítik meg, míg mások specifikus szempontból nézik. Ugyancsak változik maga a terminológia is. Ezeknek a kérdéseknek a vizsgálatát a terminológiai ellenőrzés szempontjából is fontosnak tartjuk, ezért van hiányérzetünk e blokk olvasásakor. E témacsoport egyik fejezete sem keltette fel különösebben figyelmünket. Magyar olvasó számára a tárgyszavazás amerikai gyakorlatának ilyen szintű ismertetése legfeljebb érdekességnek számít, azonban gyakorlati haszna viszonylag kevés. *Foskett* írása szemléletes bevezetés a tezaurus-problémába és minden bizonnyal érdeklődéssel forgatják majd a tanulmány lapjait az angolszász nyelvterület egyetemi hallgatói. A magyar olvasó számára azonban hasznosabb lesz ha kiváló hazai szerzőink (*Horváth* Tibor, *Molnár* Imre, *Ungváry*

Rudolf, Varga Dénes és mások) munkáiból meríti a tezaszuszokra vonatkozó ismereteket.

Jól és érdekesen megírt fejezeteket gyűjtöttek össze az *Osztályozás* címet viselő blokkba. Az osztályozásról írt tanulmányok szerzői: J.E. Daily, Karen Sparck Jones, M.L. Mainheimer, N. Neelameghan és M.E. Stevens. Ezek a tanulmányok mind a természetes mind a mesterséges nyelven alapuló osztályozással foglalkoznak.

A mű befejező három fejezete a referátumok, kivonatok típusaival és készítésükkel foglalkozik a manuálistól a gépi módszerekig.

Befejezőként ismételten megállapíthatjuk: a kötet jó bevezető olvasmányokat tartalmaz az osztályozás, az információkeresés és általában a tartalmi feltárás témaköréhez. Mindezt interdiszciplináris megközelítésben teszi, amely ismét felhívja figyelmünket az egyes szakterületek egymást kiegészítő kölcsönhatásaira.

RÓNAI Tamás



A Bok og Bibl 1984. 5.sz. humorából.