

csatorna, amelyen keresztül a hatás átadódik. De mi is az, ami átadódik?

Ha impulzus, akkor tisztán fizikai (és biológiai) folyamattal van dolgunk, aminek nincs semmi köze az információs folyamatokhoz.

Ha jel, akkor a csatorna végén léteznie kell egy olyan interpretáló rendszernek, amely rendelkezik ismeretekkel, amelyek segítségével az impulzust jellé alakítja. De mi ez az ismeret a különböző természetű objektumokban? A legegyszerűbb válasz,

hogy ez az információ. De vajon mindenki elégedett lesz-e ezzel a válasszal?

/BLÜMENAU, D. I.: Információ: miféle real'nost'? = Naučno-technická informácia, Ser. 2. 2. sz. 1985. p. 1–4./

(Környei Márta)

A szókészlet ellenőrzése és az információs technológia

A brit könyvtáros egyesület katalogizálási és indexelési szekciójának 1983. áprilisi *chesteri konferenciáján* tartott előadás azt a – főként *H. Borko* által képviselt – álláspontot járja körül, miszerint a természetes nyelvű dokumentumokat feldolgozó automatikus indexelő rendszerek feleslegessé tehetik az indexelő szakember munkáját, s az általa használt tezaurszt is, mivel a használó által megfogalmazott kérdésben szereplő szavak az információs technológia gyors fejlődése révén közvetlenül kereshetővé válnak a dokumentumok teljes szövegét (vagy referátumát) tartalmazó adatbázisokban. *F. W. Lancaster* – a fenti nézet másik képviselője – ugyan elismeri, hogy így sok nem releváns tételt is kaphatunk, de az automatikus indexelést ezzel együtt is olcsóbbnak tartja.

Borko az 1983. évi *augsburgi FID-szemináriumon* fejtette ki elképzeléseit a jövő automatizált könyvtáráról. A katalógust kizárólag számítógéppel kezelt MARC-rekordok alkotják, a rekordmezők invertált fájlokhoz vannak hozzárendelve, lehetővé téve a szerzők, címek stb. szerinti online keresést. (Több tájékoztató intézményben ez már valóság.) A dokumentum egészét, részét, referátumát stb. számítógépre viszik, az automatikusan képezi a kulcsszavakat, s az invertált fájlok révén ezeket is visszakereshetővé teszi. (Könyvtárban még nem alkalmazzák, de néhány tájékoztató intézményben igen.) A szabadpolc átnézését a terminál előtti "böngészés" váltja fel, ami véget vet a szabadpolc központú osztályozásnak, a könyveket pedig beszerzésük sorrendjében, nagyság, esetleg csak igen tág tematikus csoportok szerint lesz célszerű tárolni. A tárgy szerinti keresés abból áll majd, hogy a használó kérésében előforduló kifejezéseket a számítógép összehasonlítja az automatikus indexelés során alkotott kulcsszavakkal. Ez lehetővé teszi, hogy az olvasó több ezer tétel közül a terminál képernyőjén meg-

kapja az őt érdeklő néhányat, mindezt másodpercen belül. Borko szerint ez ugyanolyan hatékony lesz, mint a mai "manuális" módszer, de olcsóbb annál.

A gazdaságossággal való érvelés azonban megtévesztő. Igaz, hogy ha a számítógép mellett nem alkalmaznak indexelő szakembert, az csökkenti az intézmény kiadásait, viszont növekedhet a keresés költsége, a terminál előtt eltöltendő idő, a sok irreleváns tétel átnézése pedig bosszúságot okozhat.

Néhány ilyen kudarc tényezővel találkozott a *Middlesex-i Műszaki Főiskola* egyik könyvtárosa, amikor a SDC Orbit rendszerben az egyik oktató kérésére a közös piaci országokban alkalmazott világítástechnikai előírásokra vonatkozó dokumentumokat próbálta meg visszakeresni. A természetes (zsargon-) kifejezés ("yellow light") nem eredményezett találatot. A lehetséges szavak csonkolt, illetve teljes alakú Boole-operátorokkal való kombinálása révén az automatikus indexelő rendszer 32 találatot írt ki, de ezek között egyetlen releváns sem akadt. A kudarcok háromféle okra vezethetők vissza, mindhárom a szókészlet ellenőrzése (vocabulary control) témakörébe tartozik.

Az első ok a tárgy megragadásával kapcsolatos: a számításba jövő kifejezések közül az indexelő szakember ki tudja választani a megfelelőket, míg a gép csak a szavak formális kezelésére képes, nem tudja a szöveggörnyezetből kikövetkeztetni, hogy az adott szavakat melyik értelmükben használják. Ez pl. az idiomatikus kifejezések terén okoz gondot ("a ... fényében" kifejezést a gép felveszi, mert nem tudja, hogy itt nem a fizikai értelmű fényről van szó). A jövőben a számítógép is megszerezheti ezt a képességet, de ma a szakember megbízhatóbb.

A második ok: a szakember a fogalmak közötti tartalmi viszonyt is pillanatokon belül képes átlátni – pl. hogy (az egyik találat szövegében felbukkanó két

szó közül) a világítás szolgálja-e a jogot (pl. a börtönök éjszakai biztonságával összefüggésben) vagy a jog a világítást (amire a keresés eredetileg irányult). A számítógép ennek a "szerepvárnak" a feloldására is csak távlatilag lesz alkalmas, míg az olyan létező indexelő szolgáltatások, mint a PRECIS vagy a BLAISE, sikerrel birkóznak meg vele a prekoordinált indexelés és a Boole-operátorok révén. Az összekapcsolás (pl. "attitűd" és "nők" önmagában ugyan még nem tárja fel a finomabb összefüggéseket (a nők attitűdjé vagy a nőkkel kapcsolatos attitűd), de erre a megkülönböztetésre lehetőség van a PRECIS számszámai követő szöveges "jellemzők" (verbal features) gyors átfutása révén.

A harmadik ok: az azonos alakú szavak összekeverhetősége (pl. a "light" főnévként fényt, melléknévként könnyűt jelent – az egyik példában ráadásul a jelzöt több szó is elválasztja a jelzett szótól). Egy nemzetközi (ISO 2788) és egy brit szabvány (BS 5723) az önálló mellénevek deskriptorként való használata ellen foglal állást (megengedve használatukat összetett kifejezésekben). Az ajánlott főbb szabályok: elsősorban főneveket kell felvenni, a fogalmakat következetesen kell használni, az azonos értelmű kifejezések közül egyet kell deskriptorként használni, s a többit utalni. A megszámlálható fogalmakat kifejező főnevek esetében a többes szám, a megszámlálhatatlanok esetében az egyes szám alkalmazása az ésszerűbb. Az összetett szavakat és szókapcsolatokat minél kisebb (de önálló jelentésű) elemre kell felbontani, s a teljes kifejezést utólag összerakni (pl. repülőgépmotorok → repülőgépek; motorok). A bonyolultabb esetek megoldását egyszerű lépésekből álló algoritmus segíti, ez kiterjed az alap- (motorok) és a megkülönböztető jegy (repülőgép-) kiválasztására, az elemek közötti logi-

kai viszony feltárására (a motor része a repülőgépnek) stb.

Az ellenőrzött szókészlet használata egyszerűbbé teszi az indexnyelvet, növeli a megbízhatóságát, hogy egy adott tárgy hol és milyen formában fordulhat elő. Növeli az egy rendszeren belüli következetességet, és reményt nyújt a különböző rendszerek közötti összhang megteremtésére. A többnyelvű tezausz az ellenőrzött szókészlet egyik esete: a nemzeti nyelven kiválasztott keresőkifejezések átalakíthatók a közös keresőnyelvnek a számítógépi fájlban tárolt megfelelőire. Ez olyan terület, ahol az automatikus indexelés semmit sem tud nyújtani, míg a szókészlet-ellenőrzés általános, nemzetközi rendszerré válhat. Ez a folyamat hátrányosan érinthet néhány ismert indexelő rendszert, valamint az egyes nemzetek indexelő hagyományait, ami már az "indexelési politika" körébe vág. Mindenesetre az idő ma nem kedvez egy saját, nem szabványos rendszer kifejlesztésének. Nehézségek támadhatnak egyes használók részéről is, akik pl. a gyakran használt komplex kifejezést nem az elemeknél keresnek. Ezeknek egyrészt a gazdag utalórendszer jelentheti a kompromisszumot (ahol az összetevőkre az összetett alakokról is van utalás), másrészt figyelembe kell venni, hogy minden igényt egyszerre kielégítő, tökéletes indexnyelv elvileg nem lehetséges, az itt körvonalazott rendszer viszont működőképes.

/AUSTIN, D.: *Vocabulary control and information technology*. = *Aslib Proceedings*, 38. köt. 1. sz. 1986. p. 1–15./

(Mándy Gábor)

Az információtechnika trendjeinek áttekintése 1970–2000-ig

Az 1985. decemberi *nemzetközi online információs kongresszus* (London) plenáris ülése az információtechnika fejlődésével és várható trendjeinek áttekintésével foglalkozott. Az ott elhangzott három előadásról számolunk be az alábbiakban.

1. Az 1985-ig terjedő időszak fő jellemzői

Az 1960-as évek végén kezdődött el az elektronikus adatfeldolgozó és -tároló rendszerek alkalmazása nagy mennyiségű, szöveges információ kezelé-

sére. 1965-ben jött létre az USA-ban a *Lockheed* vállalat interaktív elektronikus információkereső rendszere. Az online információszolgáltatás életképessége csak néhány évvel később vált bizonyossá, amikor 1972-ben megalakult a *Lockheed Dialog* nevű információszolgáltató vállalata (a ma már önálló *Dialog* vállalat), és 1973-ban beindult a *Mead Data Central* vállalat a *LEXIS* adatbázissal. Ettől kezdve a nyilvános online szolgáltatóközpontok száza jöttek létre, amelyek az adatbázisok ezreinek online elérését teszik lehetővé.