

31. évf. 9. sz. 1984. szeptember

Tudományos és Műszaki Tájékoztatás

GONDOLATOK ÉS JAVASLATOK AZ OMIKK INFORMÁCIÓKERESŐ NYELVEINEK FEJLESZTÉSÉRŐL

Vajda Erik

Országos Műszaki Információs
Központ és Könyvtár

Ungváry Rudolf

Országos Széchényi
Könyvtár

I. Előzmények és magyarázatok

Az alábbi publikáció kétségtelenül sajátos, mind létrejöttének körülményeit, mind témáját és annak kifejtését, mind stílusát és szerkezetét, mind pedig a társszerzők viszonyát illetően. Ezért néhány előzetes megjegyzést igényel.

Előzmények

Az Országos Műszaki Könyvtár igazgatójához javaslatot nyújtottak be a könyvtár szakkatalógusainak, információkeresési eszközeinek és gyakorlatának reformja érdekében. E javaslatot az OMK igazgatója véleményezésre átadta több, az információkereső nyelvekkel elméletben, illetve gyakorlatban többé-kevésbé intenzíven foglalkozó kollégának, köztük e cikk szerzőinek is. Egyikünk (Vajda Erik) írásban közölte véleményét a javaslatról, de az alkalmat egyben felhasználta arra (és írásának nagy részét ez tette ki), hogy állást foglaljon az OMIKK információkereső nyelveinek fejlesztéséről, mind elviekben, mind a gyakorlatban, jóllehet nem teljességre törekedve. Másikünk (Ungváry Rudolf) már kizárólag az utóbbi kérdéssel (tehát az OMIKK információkereső nyelveinek fejlesztésével) foglalkozott, többé-kevésbé szorosan kapcsolódva Vajda Erik e tárgyban leírt gondolataihoz és egyáltalán nem foglalkozva az eredeti javaslattal.

Utóbb mindketten úgy ítéltük meg, hogy gondolatainknak az a része, amely (a kifejtésükhöz „ürügyet” támasztó javaslattól függetlenül) az OMIKK információ-

kereső nyelveinek fejlesztésére vonatkozik, szélesebb körű érdeklődésre is számot tarthat.

Magát a kiindulópontot jelentő javaslatot mindketten elvetendőnek tartottuk. A javaslatból semmit át nem vettünk, sőt az alábbiak nem tartalmazzák írásainknak azt a részét, amely magával a javaslattal foglalkozott, azt cáfolta, illetve értékelte. Ezért felmentve érezzük magunkat az alól, hogy a továbbiakban e javaslatra hivatkozunk.

A téma megéri a nyilvánosságot!

Az információkereső nyelvek problémaköre világszerte is, Magyarországon pedig különösen igen aktuális. Ez az aktualitás nem azért áll fenn, mintha az információkereső nyelvek fontossága a korábbiakhoz képest növekedett volna, sokkal inkább azért, mert a korábbiaknál sokkal nagyobb a zűrzavar az információkereső nyelvek fontosságának és problematikájának megítélésében; sokkal nagyobb a szakadék a szükségletek és az e szükségletek kielégítésére alkalmas, vagy ehhez hozzásegítő erőfeszítések, szándékok között. Erről bővebben nem is mondunk bevezetőben: reméljük, hogy a cikk önmagáért beszél. Egyetlen intézménynek e problémájáról szóló cikkünket (sőt még inkább valamiféle csonka *vitairatot* és javaslatokat) azért ajánlottuk közreadásra, mert ez az intézmény a hazai információs rendszerben elfoglalt helye, az általa feldolgozott információk mennyisége és az általa követett módszerek, használt

eszközök hatása szempontjából nem egyszerűen egy a sok közül, hanem tudományos—műszaki—gazdasági téren kiemelkedő jelentőségű, másfelől ezeket a gondolatokat kár lett volna nem közkinccsá tenni, hiszen (az OMIKK útján vagy az OMIKK-tól függetlenül) lényegesen szélesebb kört, kisebb-nagyobb mértékben egész tájékoztatásügyünket és könyvtárügyünket érintik.

Stílus és szerkezet

E cikk sem szerkezetében, sem stílusában nem nagyon hasonlít ahhoz, amit a folyóirat cikként közölni szokott, vagy akár ahhoz, amit a szerzők cikként eddigi tevékenységük során közreadtak. Az indok a már leírt előzményekben rejlik; a cikk egyik része egy vitairat kapcsán papírra vetett gondolatok sora, másik része e gondolatokhoz csatlakozó és azokat kiegészítő javaslatok rögzítése. Úgy véltük azonban, hogy a rendelkezésre álló szöveg átalakítása, átépítése éppen frissességét, hatását veszélyeztette volna, nem beszélve az erre fordított felesleges munkáról és a „cikk-stílus” szinte kötelező redundanciájáról. Ezért csak kissé „fésültük meg” eredeti szövegünket; nem változtattunk szerkezetén és csak néhány olyan utalást, megjegyzést hagytunk ki, ami hivatali előterjesztésében, vagy éppenséggel vitairatban helyénvaló lenne, de publikációban kevéssé tűri a nyomdafestéket.

Egy cikk vagy két cikk?

Az olvasó két, eredetileg különálló írást tart a kezében (a jelenlegi II. részt: Vajda Erikától, a III. részt Ungváry Rudolftól), jóllehet mindketten kisebb-nagyobb mértékben kijavítottuk saját írásunkat a „szerzőtárssal” folytatott megbeszélés után, no meg eddigi szakmai életünkben nem egyszer volt módunk arra, hogy e kérdésekben kicseréljük, vagy éppenséggel ütköztessük nézeteinket. Amikor mégis úgy döntöttünk, hogy nem két különálló, hanem egyetlen, társszerzőként szignált cikket adunk közre, akkor abból indultunk ki, hogy elveink, gondolataink, szándékaink az itt tárgyalt kérdésben többnyire azonosak, vagy csak árnyalataikban, a megközelítés módjában eltérők. Néhány részletkérdésben fennálló ellentétes véleményünk ezen nem sokat változtat, így az olvasó elnézését is közösen kérjük e cikk rendhagyó voltaért.

II. Szükségletek, problémák, lehetőségek

Minek kellene az OMIKK-nak információkereső nyelvek?

Az OMIKK funkcióinak jelentős része vagy feltételezi az információkeresést, vagy éppenséggel információkere-

sési funkció, vagy pedig a használatnak kívánja lehetővé tenni az információkeresést. Azok az adatbázisok, amelyeket információkeresési célokra használunk, részben külsők (vagyis más intézménytől feldolgozva, osztályozva, indexelve kerülnek hozzánk), részben pedig belsők, tehát létrehozásukkor az OMIKK-nak kell és lehet gondoskodnia az információkeresés lehetőségeiről. Mivel a külső adatbázisok „átindexelése” aligha célszerű és gazdaságos, az információkereső nyelvek OMIKK-on belüli fejlesztésekor alighanem a saját magunk által előállított adatbázisokra kell szorítkoznunk.

Szabad szövegszavas keresés és információkereső nyelvek

Az információkeresési műveletek (legalábbis jelentős részben) a párbeszédés üzemmódban (online) végrehajtott információkeresés felé tartanak. Ez esetben különösen, de a kötegelte eljárással végzett információkeresés esetében is egyre inkább tért hódít a szabad szövegszavakkal végzett információkeresés (helytelenül: szabad tárgyszavas információkeresés), amely kizárni látszik az információkereső nyelvek használatát. Nyomatékosan szeretném hangsúlyozni, hogy ez csak látszat, mert

- minél több módszerrel és minél több irányú megközelítéssel tudják végrehajtani az információkeresést (ide értve a különféle jellegű, kötött információkereső nyelvekkel végzett indexelést és az e nyelveket alkalmazó információkeresést is), annál inkább tudunk alkalmazkodni a mindenkor kérdés jellegéhez és annál inkább tudjuk javítani a keresés eredményét, paramétereit (teljesség, pontosság);
- a szabad szövegszavas keresőprofilok megfogalmazásának méltatlanul mellőzött, elengedhetetlen segédeszköze a fogalmak választékát és összefüggésrendszerét tartalmazó teaurusz, vagy más hasonló „információkereső nyelvi szótár”;
- végül, de nem utolsósorban: szabad szövegszavas információkeresés csak akkor végezhető megbízhatóan, ha van, és megfelelő minőségű a keresés alapjául szolgáló „szabad szöveg”, márpedig korántsem mindig ez a helyzet.

Amit a szabad szövegszavas információkeresésről a fentiekben mondtunk — mutatis mutandis — érvényes az automatizált indexelésre és az automatizált indexelés során „menet közben” készített információkereső nyelvekre is.

Mindez azt jelenti, hogy a jövő aligha bízható kizárólag a szabad szövegszavas információkeresésre és/vagy az automatikus indexelésre.

Egyetlen vagy több nyelv?

Határok és követelmények

Részben az elmondottakból következik, de egyéb — említendő — tényezők miatt is reménytelen vállalko-

zás egyetlen olyan információkereső nyelvet keresni, amely megoldaná az OMIKK összes ilyen problémáját. Ha mégis lenne ilyen nyelv (vagy ilyen nyelvek viszonylag könnyen adaptálható rendszere), annak rendkívül rugalmasnak kellene lennie. Semmiképpen sem létezhet azonban egyetlen olyan nyelv, amely módosítások nélkül alkalmas lehetne minden célra. Az információkereső nyelv jellegét, struktúráját, tényleges tartalmát több tényező is befolyásolja, köztük pl.

- az információkereső rendszer elvi és technikai realizálása: cédulakatalógus, kötet (lista), permutált index vagy ezek bármelyike COM mikrofilmlapon, vagy pedig számítógépben tárolt adatok és automatizált keresés;
- a számítógépben tárolt állomány szekvenciális vagy invertált elrendezése;
- a számítógépes információkeresés kötegelt vagy párbeszédéses üzemmódja;
- a feldolgozott dokumentumok jellege (szélső példákat megnevezve: egyfelől könyvek, sőt kézikönyvek, másfelől folyóiratcikkek, sőt rövidebb közlemények, pl. hírek vagy akár faktografikus adatok);
- a szakterület jellege és összetétele; fogalmi rendszere, valamint „információs népsége (populációja)”;
- a keresési igények jellege és – ebből adódóan – a keresett témák szűkebb vagy tágabb, egyszerűbb vagy összetettebb volta;
- a keresésre alkalmassá tett állomány tematikai szélessége;
- a rendszer használatának gyakorisága és „sürgőssége”.

Mindezt még bonyolítja, hogy nem egy esetben ugyanazt az állományt – lehetőleg ugyanabban a munkamenetben – többcélúan kell indexelni. Így pl. a referáló folyóiratokban (szakirodalmi tájékoztatókban) közölt tételek (referátumok stb.) indexelésének lehetővé kell tennie a füzetenkénti és éves – hagyományos – tárgymutatók készítését, de távlatilag alkalmasnak kell lennie számítógépes – kötegelt vagy párbeszédéses üzemmódban végzett – információkeresésre is, ha nem is magában, hanem más információkereső nyelvekkel, illetve módszerekkel (szabad szövegszavas keresés) kiegészítve. Másik példa: más-más követelmények merülnek fel a könyvtári állomány indexelésével szemben (még ugyanazon dokumentumfajta esetében is), ha az indexelés egyfelől számítógépben tárolt és online lekérdezhető „szakkatalógus” céljaira készül, és másfelől ki kell elégítenie egy COM-katalógusban szisztematikusan elrendezendő állomány osztályozási-elrendezési igényeit is.

Az indexelendő állományok

Bár az OMIKK-ban elvileg még számos állomány (adatbázis) információkeresési célokra végzendő indexelése, osztályozása képzelhető el, jelenleg – figyelmen

kívül hagyva az Országos Kutatásnyilvántartást – két lényeges területről van szó:

- a) a könyvtári állományról, ezen belül
 - könyvekről (mégpedig igen különböző típusú könyvekről), szükség esetén ezek részleteiről (analitikus feldolgozás);
 - időszaki kiadványokról (csak összevont szinten);
 - kutatási jelentésekről (ezek, legalábbis eddig, nagyrészt amerikai kutatási jelentések, amelyek angol nyelvű indexelési eredményei rendelkezésre állnak);
 - fordításokról (vagyis végeredményben cikkekről, könyvrészletekről, szabadalmi dokumentumokról, szabványokról, kutatási jelentésekről, prospektusokról, gépkönyvekről és más termékismertetőkről);
- b) az ún. „szakirodalmi bázisról”, vagy „magyar nyelvű referátumbázisról”, vagyis különféle dokumentumok (elsősorban cikkek, de még 6–7 féle dokumentum) referátumairól, esetleg annotált bibliográfiai leírásairól, vagy pusztán bibliográfiai leírásairól, minden esetben címfordítással kiegészítve.

Mint látható, az információkeresésre számításba jövő két nagy állomány egymáshoz képest is és önmagában is heterogén. A kettő közül (a növekedés dinamikáját is figyelembe véve) az utóbbi a lényegesen nagyobb, hiszen évi 150–200 ezer dokumentummal gyarapszik, míg az előbbi gyarapodása aligha éri el a 70 ezer egységet, és különösen nem érné el, ha a cikkk fordítások állományát a második helyen említett állománnyal egyesítenők (ami nagyon is indokolt lenne).

Az Országos Műszaki Könyvtár információkereső nyelvei

A fent elsőként említett (könyvtári) állomány „információkeresési jövője” (ezúttal mereven és voltaképpen indokolatlanul elkülönítve a téma szerinti információkeresést az egyéb ismérvek alapján végzett információkereséstől, és csak az előbbivel foglalkozva) előreláthatólag úgy alakul, hogy a könyvtári szakkatalógus (a cédulakatalógus lezárása és/vagy párhuzamos fenntartása mellett) átalakul, egyfelől számítógépben tárolt és számítógépes (vélelmezhetően online, interaktív) keresést lehetővé tevő katalógussá, másfelől a cédulakatalógushoz képest információkeresési szempontból kevés elvi eltérést jelentő mikrofilmlapos (COM) katalógussá.

A számítógépben tárolt katalógussal szemben az igény nyilván az, hogy – a katalogizált dokumentum jellegétől is függően és egyéb információkereső nyelveket sem kizárva – speciális témák szerinti információkeresést is lehetővé tegyen, és éljen azokkal a lehetőségekkel, amelyek a számítógépes realizálás révén a koordinált indexelés (vagyis egy dokumentum tartalmának számos deskriptorral való jellemzése) terén megnyílnak. Ugyanakkor a COM-katalógusnak alighanem valamilyen szisztematika szerint kell tartalmaznia a dokumentumok leírá-

sait, bibliográfiai tételeit, legalábbis a COM-katalógus tárgyköri visszakeresésére alkalmas változatában (a COM szakkatalógusban). A COM-katalógus esetében elvben két szélsőség között lehet választani: az ETO (vagy a helyébe léptetett bármely egyéb hierarchikus rendszer) fenntartása az egyik oldalról és a mechanikus felépítésű (vagyis a tételeket a tárgyszavak betűrendjében tartalmazó) tárgyszókatalógus a másik oldalról.

A tárgyszókatalógusról tudni kell azt, hogy sem a cédulakatalógus, sem a COM-katalógus nem enged meg a számítógépes realizáláshoz hasonlóan részletező, koordinált indexelést, hanem ennél átfogóbban osztályozó tárgyköri csoportosítást kíván. Ez esetben a variációs lehetőségek csupán a tárgyköri csoportalkotás mélysége, lineáris vagy fazettás jellege, különösképpen pedig a tárgyszavak (targyköri csoportok) mechanikus vagy egyéb elrendezése terén állnak fenn. Minden olyan elképzelés, amely – a számítógépben realizált katalógus céljaira végzett indexelés eredményeként a dokumentumokhoz hozzárendelt – összes deskriptorhoz „le akarja rakni” a megfelelő katalóguscédulákat, naiv, és aligha jelent mást, mint kevés gyakorlati értékkel bíró kísérletet, még akkor is, ha az azonos deskriptorral jellemzett bibliográfiai tételeket egymáshoz képest a további deskriptorok betűrendjében helyezzük el a katalógusban (a ciklikusan permutált indexek módszerét követve).

A könyvtári állomány szakkatalógusai jelenleg részben ETO szerinti, részben tárgyszavasak. Az ETO szerinti katalógussal itt nem kell bővebben foglalkozni. A tárgyszókatalógust (legalábbis annak rendszerét) nem ismerem kellőképpen ahhoz, hogy állást tudjak foglalni akár arról, hogy jelen formájában van-e jövője (az előbb vázolt perspektívák szempontjából), vagy akár alapjául szolgálhat-e további tárgyszórendszer-alkotásnak, vagy éppenséggel teaurusz-munkálatoknak. Kételyeim mindenestre erősek a kezembe kerülő katalóguskártyákon talált tárgyszavak, illetve a tárgyszavazás színvonalának ismeretében. Ettől függetlenül, ha a jelenlegi tárgyszórendszert bármilyen jövőbeli fejlesztéshez fel akarjuk használni, akkor ennek alapfeltétele a tárgyszóállomány kategóriák és „szakcsoportok” szerinti elrendezése és a tárgyszavak közötti teaurusz-összefüggések (szemantikai összefüggések) kimunkálása. Ennek eredménye nem lesz szükségképpen az, hogy a tárgyszókatalógus teaurusz-szerűen vagy teaurusz kiinduló alapjaként váljon használhatóvá, de legalább meg lehet állapítani, hogy e célokra mennyire alkalmas. Az egyébirányú fejlesztési lehetőségeket lásd a továbbiakban.

A referátum-állomány indexelése

Ami a másik nagy állományt – a továbbiakban: referátumállomány – illeti, itt is alkalmazzuk az ETO szerinti osztályozást. Ennek létjogosultságát három szempont adja meg:

- a referáló folyóiratok „legalsó szintű” fejezetein belül a tételeket ETO szerinti sorrendbe rakjuk és ez ideig nem találtunk ki semmi jobbat;
- a referátum-állományban végzett visszamenőleges információkeresésre (ha eltekintünk a szakirodalmi tájékoztatók éves, kumulált tárgymutatóitól) egyetlen eszköz áll rendelkezésre: a volt figyelőkarton-szolgálat anyagából, majd a szakirodalmi tájékoztatók anyagából szerkesztett ún. információk karterek, vagy másik néven szakirodalmi adattár, ez pedig ETO szerint van elrendezve és ez idő szerint nem is rendezhető el más információkereső nyelv alapján;
- bármi is legyen a jövő, az ETO (legalábbis egyelőre úgy tűnik) kiegészítő (egyes esetekben elsődleges) információkereső nyelvként használható marad (lásd előbb).

A fentiek ellenére is nyilvánvaló, hogy a referátum-állomány géppel olvasható formájának létrehozása és az így kialakuló adatbázis visszamenőleges információkeresésre alkalmassá tétele az ETO és a szabad szövegszavas információkeresés mellett (az utóbbi lehetőségei ez esetben aránylag kedvezőek, ha a referátum szövegét, illetve a magyar – értelmező – címfordítást is géppel olvasható adattárban kezeljük) más megoldást is feltételez, és e megoldás legalább tárgyszórendszer(ek), vagy – még inkább – teaurusz(ok) alkalmazása kell, hogy legyen. Mint már korábban említettem, e megoldás visszahat a szabad szövegszavas információkeresés eredményességére is.

A tárgyszórendszer(ek) vagy – távlatilag – teaurusz(ok) létrehozására tettünk bizonyos bátortalan és következtelen, de annál gyakorlatibb lépéseket, amikor a szakirodalmi tájékoztatók füzetenkénti tárgymutatójának készítésére és ennek érdekében az egyes referátumok (tételek) tárgyszavas indexelésére tértünk át a korábbi, jellegzetesen „subject heading”-es éves mutatók készítéséről. A tárgyszavazók menet közben építik fel (vagy egyre inkább építették fel, ami a tárgyszavak törzsállományát illeti) az egyes szakirodalmi tájékoztatók tárgyszórendszerét, mégpedig (azokban a ritka esetekben, amikor kellő rugalmassággal, de alaposan követték hézagos instrukcióinkat) lényegében olyan elvek alapján, mintha teaurusz szóállományát építették volna fel az indexelt irodalomból. Ez azt jelentette, hogy javasoltuk a fogalmak (targyszavak) időnkénti kategóriás és szakcsoportos elrendezését és (részben ennek alapján e célra használható szövegösszefüggések alapján, részben pedig saját ismereteik alapján) „lásd még” típusú utalók beépítését a teauruszok alá- és fölrendeltségi, illetve egyéb szemantikai relációinak képviselőjében. Hasonlóképpen a teaurusz-készítéskor követett elveket igyekszünk megkövetelni a „nem-targyszavak” („nemdeskriptorok”) meghatározásakor, vagy – „tudományosabbban” – az ekvivalencia-osztályok összeállításakor és ezek reprezentánsainak kijelölésekor.

Másfelől viszont

- a tárgyszóként felhasznált fogalmak (terminusok) jellege és „fogalomszervezési szintje” (vagyis prekoordinációjának mértéke) erősen igazodik a mindenkori szakterülethez, és – ami a nagyobb probléma – ahhoz a tényhez, hogy a tárgyszavazás nem számítógépen realizált információkereső rendszer, hanem hagyományos tárgymutató céljait követi, minél pragmatikusabban;
- a különböző szakirodalmi tájékoztatók tárgyszó-állománya egymástól függetlenül, elszigetelten fejlődik ki, igazodván e szakterület terminológiájához is, és mellőzve mindennemű horizontális terminológiai egyeztetést;
- mind a tárgyszavazás minősége és szemlélete, mind a tárgyszórendszer fejlesztésének minősége és szemlélete rendkívül heterogén, a tartalom és a követett módszerek tekintetében egyaránt, más szóval, függ a tárgyszavazó szerkesztő személyétől, képességeitől, képzettségétől és szándékától.

A fentiek értelmében a szakirodalmi tájékoztatók tárgyszójegyzékai nyilvánvalóan felhasználhatók – jóllehet csak részben és erős kritikával – valamiféle OMIKK-tezaurusz(ok) egyik alapjaként. Arra is gondolni kell, hogy az OMIKK és a szakirodalmi tájékoztatók „gyűjtőköré”, tematikája nem teljesen azonos; a szakirodalmi tájékoztatókban tudatosan mellőzünk az OMIKK főgyűjtőkörébe tartozó olyan területeket, amelyek még az átlagshoz képest sem vagyunk versenyképesek a nagy külföldi szakirodalmi adatbázisokkal, illetve, amely területeken egyszerűen nem engedhetjük meg magunknak a valóban reprezentatív erejű referálást.

Következtetések és kérdések

Minden fenti megfontolásból következik, hogy alighanem a könyvtári információkeresés céljaira is, de a referátum-állományban végzett információkereséshez mindenképpen (további feladatokról nem is beszélve) célszerű lenne tezaurusz vagy tezauruszok fejlesztése az OMIKK-ban. E sommás megállapítás azonban nagyjából az utolsó, amit határozottan ki lehet mondani. A többi: a következőkben hézagosan megfogalmazott kételyek vagy variánsok együttese.

a) Egyetlen „általános műszaki” vagy még inkább „általános OMIKK” tezauruszra van szükség, vagy szaktezauruszok sorozatára, vagy éppenséggel szaktezauruszok rendszerére? Személyesen úgy vélem, hogy egyetlen „OMIKK tezaurusz” még akkor is kilátástalan vállalkozás, ha eltekintünk az adott szakterületen vagy szakterületeken használandó tezaurusz(ok) eltérő használati célokból adódó eltéréseitől. Az OMIKK tevékenységi területe túl széles ahhoz, hogy akár a már említett

„fogalomszervezési szint” egysége szempontjából, akár a homonímiából adódó problémák megnyugató megoldása szempontjából el tudnám képzelni egyetlen tezaurusz felépítését. Ennek egyébiránt ellentmond az is, hogy tevékenységi területünk egyes részletein az információsűrűség oly csekély, hogy az illető terület „tezaurusz-szelete” jószerivel egyetlen (és mint ilyen csúcs-) deskriptorból és nemdeskriptorok tetszőleges, de mindenképpen meglehetősen nagy halmazából állna, vagy legalábbis célszerűen ebből kellene állnia. Az egymástól független szaktezauruszok alkalmazása (ha ezek szerkezeti elveik és készítési, valamint használati metodikájuk tekintetében többé-kevésbé egységesek) vonzó megoldás, de már eleve feltételezi a teljes állomány előzetes, sekély osztályozását annak érdekében, hogy eldöntessék, melyik tételt melyik szaktezauruszsal, vagy melyik szaktezauruszokkal kell indexelni, vagy éppenséggel egyáltalában nem kell indexelni, mert az adott szakterületbe sorolás (annak csekély információsűrűsége miatt) elegendő.

Az „előosztályozás” viszont valamiféle (jóllehet a felsorolandóknál lényegesen sekélyebb) BSO-t [1] vagy rubrikátort, vagy éppenséggel „osztaurusz” [2] tételez fel, mégpedig olyant, amely alkalmazkodik az OMIKK feladatköréhez, tematikájához és feldolgozott dokumentumaihoz egyaránt. Ilyen pedig legalább annyira nincs, mint amennyire nincsenek megfelelő és egységes szemléltető szaktezauruszok. A jövő mégiscsak e felé a megoldás felé kacsint, kiegészítve azzal, hogy a szaktezauruszok ne legyenek egymástól teljesen elszigeteltek, hanem legalábbis a terminológiai kontroll és a kölcsönösen, illetve általánosan felhasználható szemantikai összefüggések érvényesítése tekintetében legyenek egy egységes rendszer részei.

b) Ha valamely csoda folytán (mindjárt szó lesz róla, hogy ehhez miért kellene csoda) létrejönne az OMIKK szaktezauruszainak rendszere és az információs tételeket a szaktezauruszhoz irányító, egyes területeken a tezauruszt pótló sekély osztályozórendszer, akkor is kérdéses marad, hogy az így kialakuló rendszer, illetve annak tagjai hogyan lesznek alkalmasak a korábbiakban már többször említett párhuzamos, a követelmények szempontjából gyakran ellentmondó indexelési követelmények teljesítésére. Ehhez a tematikusan eltérő („szak-”) tezauruszoknak még változatait is elő kellene állítani, pl. a COM-katalógus, illetve a számítógépes könyvtári katalógus, vagy a szakirodalmi tájékoztatók mutatói és a referátum-állomány adatbázisának céljaira. Ez viszont egyfelől igen nagy és bonyolult munkát jelent (nem mintha az eddig említettek valami mást jelentenének), másfelől feltétlenül szükségessé teszi magának a tezaurusz-építésnek, karbantartásnak és alkalmazásnak a számítógépesítését is. Erre pedig hol a kapacitás, a szoftver és – a legkevésbé sem utolsósorban – a szakértelem?

c) És a fentiekben van a kutya elásva! Míg más országokban kiváló elmék sokasága munkálkodik nem

csak az osztályozásmélet művelésén, hanem a jobb vagy rosszabb teauruszok készítésén is, addig Magyarországon e törekvések elszigeteltek; míg a szocialista országokban (ha különböző eredményekkel is) különálló intézetek, illetve nagy információs intézmények jelentős osztályai és csoportjai foglalkoznak az információkereső nyelvek metodikájának fejlesztésével és egyes konkrét keresőnyelvek (teauruszok stb.) készítésével, illetve a teaurusz-készítés koordinációjával, addig Magyarországon egy-két, az elméletet és gyakorlatot egyaránt (jóllehet egy ember kivételével nem főfoglalkozásban) művelő ember pusztába kiáltott szava képviseli az egyetlen – koncepciójában alátámasztott, rendszerszemléletű és ugyanakkor kellően gyakorlatias – odafigyelésre érdemes szemléletet, nem is szólva az OMIKK nyomorúságáról, amely korábbi (6–8 évvel ezelőtti és azt megelőző) jobb és rosszabb kísérletek után az információkereső nyelvek ügyét egyetlen csomagban (a szervezeti alapokkal együtt) kidobta az ablakon, és (jó egy éve) is csak az NTMIR megfelelő munkacsoportjában való hazai részvétel koordinációját hozta vissza, mérsékelt eredménnyel.

d) Világosan látjuk tehát, hogy az OMIKK információkereső nyelveinek fejlesztése terén sok mindent kell tenni, ha az OMIKK teljesíteni kívánja információkeresési funkcióit a házon, illetve országon belül felépülő adatbázisok terén, és ha egyben segíteni akarja a műszaki tájékoztatási egységek és műszaki szakkönyvtárak hasonló feladatainak végrehajtását. Kevésbé világosan, de hozzávetőlegesen azt is látjuk, hogy milyen célokat kell elérni, illetve milyen úton kell megindulni e célok realizálása felé. Mindezeknél világosabb azonban, hogy csak akkor érhetjük el vagy közelíthetjük meg e célokat, sőt csak akkor hozható be valami hosszú és súlyos lemaradásunkból, ha ennek anyagi, személyi és mindezekelőtt tervezési-szervezési feltételeit megteremtik az erre illetékesek.

III. ... és egy lehetőség

Amiből kiindulhatunk

Abból indulhatunk ki, hogy

1. az OMIKK-ban (a) előbb-utóbb elkerülhetetlenül be kell vezetni az online üzemmódban folytatott információkeresést, amelyhez szükség lesz

- mind a szabad szövegszavakkal (tehát ellenőrzött szótár segítségével nélkül) végezhető információkeresésre,
- mind pedig ellenőrzött szótárral (végső formájában teauruszsal) végezhető információfeltárással és -kereséssel;

és (b) ugyanakkor továbbra is szükség lesz a „hagyományos” hierarchikus osztályozási rendszer használatára is

(melyhez fejlett hazai kultúrája következtében továbbra is az ETO a legalkalmasabb) [3]. Az ETO-n alapuló katalógust még az automatizálást követően is indokolt kézzel kezelhető formában – katalóguscédulákon – fenntartani;

2. az ellenőrzött szótár(ak) kidolgozásához – noha az online üzemmódot az utcáról betérő felhasználó a közeljövőben még aligha használhatja (holott bevezetésének elsősorban ez volna az értelme) – már most hozzá lehet kezdeni. Ezt – a későbbiekben javasolt szakaszos megvalósíthatóság mellett – az is indokolja, hogy ily módon a megfelelő gépi feltételek megvalósulásakor egyrészt rendelkezésre áll(nak) kiinduló ellenőrzött szótár(ak), másrészt pedig lesznek már némi tapasztalatok a koordinált indexelés alkalmazásában. Az automatizált eszközöket tehát üzembehelyezésüktől kezdve használni lehet az információkeresésre. Más szóval: amortizációjuk idővesztés nélkül megkezdődhet.

A kiinduló ellenőrzött szótár(ak)nak erre az időre nem kell teljesen kész – teaurusz – formában meglennie. A gépi információkereső rendszerekben használt teauruszokat ugyanis a mindenkori információkereső programcsomag [4] lehetőségeinek megfelelően öntik végleges formába. Előre kialakítható azonban a tervezett teaurusz(ok) szókincsének jelentős része, e szókincs előre rendezhető (fogalmi kategóriák és alkalmas szakcsoportok szerint), megállapíthatók e szókincsen belül a legfontosabb – még ha nem is feltétlenül végleges és teljes – szemantikai relációk, sőt: kiemelhetők e szókincsből azok a kifejezések, melyek az OMIKK teauruszának a „csúcát” (makroteauruszát), s szükségszerűen a legállandóbb részét alkotják.

E munkák már önmagukban sok időt igényelnek, számtalan – és elkerülhetetlen – buktatóval, az előzetes elképzelések helyesbítésével, s nem utolsósorban hasznos tapasztalatokkal járnak. Mindez azonban az automatizálás megvalósulásáig még eltelő időre esnek, ami az online üzemmód gyakorlati bevezetését illetően tiszta időnyereség: nem attól számítva kell az ellenőrzött szótár készítésével elkerülhetetlenül együtt járó tanulópénzt megfizetni.

Ráadásul – lévén, hogy az online üzemmód bevezetésére minden bizonnyal még várni kell – ezek a munkák nyugodt, sürgetésmentes körülmények között volnának elvégezhetőek, aminek a megfelelő szellemi infrastruktúra hiányai miatt nem elhanyagolható a jelentősége;

3. az OMIKK ellenőrzött szótárral (végső soron: teauruszsal) végzendő információkeresési igényeit

- nem lehet egyetlen óriásteauruszsal kielégíteni, mivel ennek elkészítéséhez elméletileg sem értek meg a feltételek,

hanem

- több, egymástól a gyakorlatban függetlenül használt speciális teauruszsal kell kielégíteni, melyek szerke-

zetüket, készítésüket és használatukat illetően egységes elveken alapulnak.

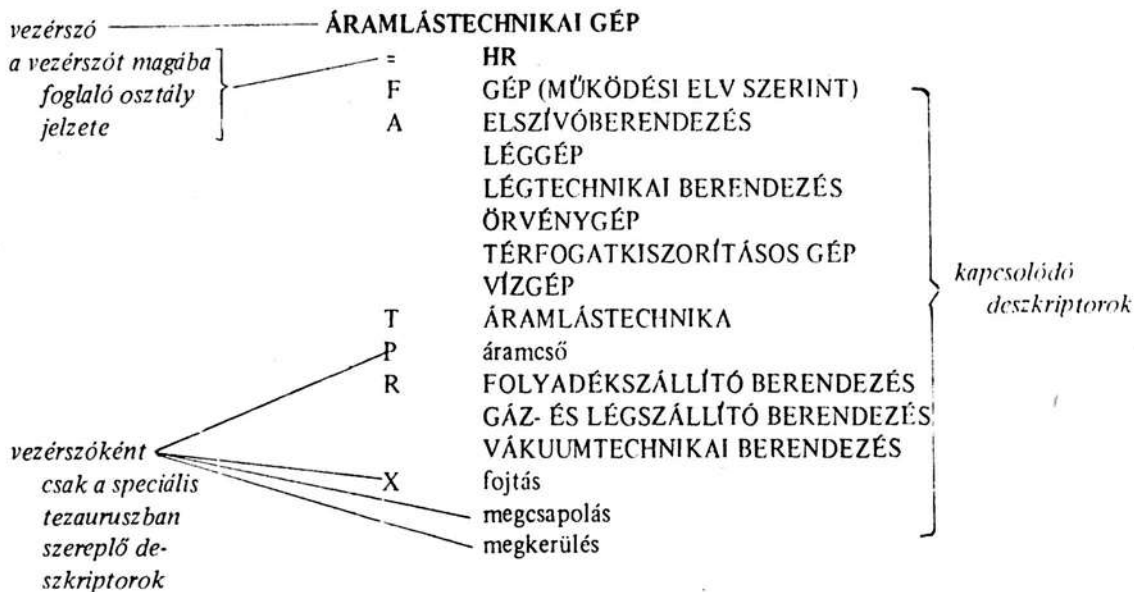
E speciális teauruszok összehangolása és a koordinált dokumentumfeltárás érdekében szükség van olyan csúcsteauruszra (makroteauruszra), mellyel a feltárandó dokumentumok teljes állománya átfogóan, általánosabb szinten („sekély” módon) előosztályozható, s így módon az egyes speciális, OMIKK-on belüli információs rendszerekhez irányíthatók a dokumentumok, ahol a speciális teauruszokkal megtörténhet a mindenkori igények (online üzemmód vagy COM-katalógus, vagy cédulakatalógus stb.) szerint a finomfeltárás.

Az e célra készítendő átfogó teauruszt nevezem „osztaurusznak”, mivel segítségével előosztályoznak

(ezért „osz-”), ugyanakkor ezt az előosztályozást egyrészt koordináltan, mellérendelő módon valósítják meg, másrészt osztályaihoz a teaurusz-elv alapján kialakított szemantikai kapcsolatokat (szóbokrokat) is hozzárendelnek (ezért „(tez)aurusz”). Ilyen osztaurusz készült el 1983–84-ben az OSZK és az OMIKK együttműködésében (az OMIKK számítógépén), melyet a két intézményben közösen használnak a külföldi folyóiratok egységes tartalmi feltárására [5].

Az osztaurusz osztályai olyan deskriptorokhoz kapcsolódnak, melyek a speciális teauruszokban az átfogóbb deskriptorokat képviselik. Belőlük a nem-faj, egész-rész, irányultság-függőség és rokonsági relációk alapján speciálisabb jelentésű illetve jelentőségű deskriptorokhoz lehet a mikrofeltáráskor jutni.

Egy ilyen osztauruszcikk látható az alábbiakban:



A HR az „áramlástechnika” osztály jelzete, mellyel minden olyan dokumentumtartalmat jelölni kell, melyet az **ÁRAMLÁSTECHNIKAI GÉP** fogalom jellemez.* Az összes többi deskriptor a vezérszó által képviselt fogalom közvetlen kapcsolatairól tájékoztatja a feldolgozót illetve a keresőt, annak érdekében, hogy további, szóba-jöhető asszociációkat is figyelembe vehessen.

Az átfogó (elő)osztályozáskor egyszerre több – egymással közvetlen szemantikai kapcsolatban nem álló – osztályba is besorolható a dokumentumtétel. Az osztályokat úgy kell a tervezett osztauruszban kialakítani, hogy elvezessenek az OMIKK-ban működő speciális információkereső rendszerekhez.

* Még számos más deskriptor (pl. szivattyú, szellőztetés, dugattyúgyűrű, turbina, nyomómagasság, kavitáció) esetében is ugyanezzel az osztályozási jelzettel kell átfogóan osztályozni, azaz mindezeknek a deskriptoroknak a deskriptorcikkében ott szerepel e jelzet.

Az osztaurusz használatának lényege, hogy az egyes rendszerekben a dokumentumokat az osztauruszban megadott jelzetek szerint (tehát nemcsak a speciális teauruszok deskriptorai szerint) is feltárják. Az így feltárt dokumentumok eljuttathatók ennek segítségével mindazokhoz a speciális információkereső rendszerekhez, melyek a szóban forgó osztaurusz-osztályokba tartozó dokumentumokra igényt tartanak.

Egy-egy osztályra egynél több rendszerben is igényt tarthatnak. Ebből, valamint a már említett tényből, hogy egy dokumentum egyszerre több osztaurusz-osztályba is besorolható, következik, hogy az egyes dokumentumok az OMIKK több speciális információkereső rendszerében is megjelenhetnek (pl. a FOLYADÉKSZÁLLÍTÓ BERENDEZÉSt tárgyaló szakcikk az „anyagmozgatási”, a „gépgyártástechnológiai” és a „vízügyi” referáló folyóiratokban).

A speciális teauruszokban ugyancsak megjelennek az osztauruszban szereplő deskriptorok és osztályozási

jelzeteik; a hozzájuk tartozó szöbökroknak már nem kell egyformáknak lenniük minden tezauszban. A speciális tezauszok szókészletének túlnyomó része viszont nem szerepel az osztauruszban, és nem is kapcsolódnék hozzájuk osztályozási jelzet, mivel specifikus jelentőségük ezt már nem indokolja.

A párbeszédés – online – üzemmódban a kereső vagy definiálja az információkereső rendszert, és akkor annak speciális tezauszát használja, vagy nem definiálja, és akkor az osztauruszt használhatja (és megelégszik az átfogóbb kérdésfeltevéssel).

Mi valósítható meg az OMIKK információkereső rendszereinek gépesítéséig?

A megvalósíthatóságot szakaszonként célszerű vizsgálni. Mindegyik szakasz önmagában zárt, és terméke haszonnal alkalmazható, függetlenül attól, hogy a következő szakaszokra sor kerül-e vagy sem.

I. szakasz

1. Összegyűjthetők az OMIKK különböző tematikájú ill. rendeltetésű rendszereiben jelenleg – egyelőre ellenőrzött szótár nélkül – használt kifejezések (ilyenek például azok, melyeket a jelenlegi „tárgyszókatalógusban”, a referáló kiadványokban, a fordítónyilvántartásban stb. használnak).
2. Elkészíthető az összegyűjtött kifejezések kumulált betűrendes jegyzéke.
3. Morfológiailag (a szóalakok tekintetében) rendezhetők a kumulált jegyzékben szereplő kifejezések.

Mindehhez szükség van:

- *szógyűjtő, rendező és módosító programrendszerre, mellyel mind betűrend szerint, mind a gyűjtés forrásai szerint jegyzékek nyomtathatók;*
- *a morfológiai egységesítés elveit összefoglaló szabályzatra.*

Az első szakasz eredményeként keletkező jegyzékek az egyes rendszerekben segédletként már a gépesítés bevezetése előtt is felhasználhatók. Célszerű arról is gondoskodni, hogy a jegyzékben nem szereplő, de a gyakorlati munka során felvett újabb kifejezések szintén folyamatosan bekerüljenek a kumulált jegyzékbe.

Mindebből következik, hogy már a munka eme szakaszában állandó szerkesztőséget kell felállítani, mely a szavak folyamatos begyűjtéséről, kumulációjáról, morfológiai karbantartásáról és a számítógépes feldolgozás megszervezéséről gondoskodik.

II. szakasz

4. Az összegyűjtött kifejezések fogalmi kategóriák és pragmatikus szempontok szerint kialakított szakterületek – szakcsoportok – szerint rendezhetők és kinyomtathatók.

Mindehhez szükség van:

- *katagóriák, szakcsoportok szerint rendező, válogató és kinyomtató programrendszerre;*
- *alkalmas szakcsoportos rendszerre a fogalmak besorolására és a fogalmi kategóriák besorolási rendszerére.*

A szakcsoportosítás és a kategorizálás a szükséges feltétele annak, hogy a kifejezések között megállapíthatók mind a valódi szinonímákat (és az OMIKK minden speciális információkereső rendszerében kváziszinonímának tekinthető kifejezéseket), mind pedig a deskriptorok közötti legfontosabb szemantikai relációkat.

A második szakasz eredményeként keletkező szakcsoportok alapján egyben már olyan jegyzékek nyomtathatók ki, melyek pontosabban, válogatva tartalmazzák a speciális rendszerekben szükséges kifejezéseket.

Ez a szakasz már jóval nagyobb munkát ró a szerkesztőségre. Célszerű bevonni külső munkatársakat, hogy a szerkesztőség elsősorban az ellenőrző és koordináló munkát végezhesse.

III. szakasz

5. Kiválogathatók az OMIKK osztauruszába való általánosabb fogalmak.
6. Kidolgozhatóak az osztauruszcikkek: megállapíthatók az osztaurusz nemdeskriptorai, azaz az OMIKK minden rendszerében érvényes szinonímák és kváziszinonímák, továbbá az osztaurusz deskriptorai között szükséges szemantikai kapcsolatok, valamint a deskriptorok kapcsolatai, az osztaurusz-osztályok megfelelő jelzeteivel.

Mindehhez szükség van:

- *az osztaurusz-osztályok megfelelő rendszerére;*
- *a tezausz-cikkek felépítésére, kezelésére, módosítására, rendezésére és kinyomtatására alkalmas programrendszerre;*
- *az OMIKK-ban érvényes tezausz-szerkesztési irányelvekre.*

A III. szakasz eredményeként keletkező osztaurusz segítségével próbafeldolgozások végezhetőek, s maga az osztaurusz ennek alapján tökéletesíthető. Másrészt kipróbálhatóak azok a körülmények, ahol az osztaurusz alkalmazása szóba jöhet (adott esetben akár manuális rendszerben is).

Az osztaurusz szókészletét áttekinthető, néhány ezer szavas nagyságrenden célszerű tartani. A deskriptorok és az osztályozási jelzetek közötti kapcsolatot az OMIKK szükségletei határozzák meg elsősorban, és csak másodsorban tudományos-elméleti szempontok. A szemantikai kapcsolatok megállapítása ezzel szemben kizárólag szakmai kérdés, melyet külső szakemberek bevonásával célszerű elvégeztetni.

IV. szakasz

7. Megállapíthatók a további, az osztauruszba nem kerülő kifejezések közötti szemantikai relációk. Ezeknek

a relációknak egy része az osztaurusz készítéséből már eleve adódik, más részük fokozatosan állapítható meg, párhuzamosan azzal, ahogy az OMIKK egyes rendszereiben kialakulnak a speciális teauruszok a II. szakaszban rendelkezésre bocsátott jegyzékek alapján.

Mindehhez szükség van:

- az I–III. szakaszban leírt programrendszerre;
- a III. szakaszban említett teaurusz-szerkesztési irányelvekre.

A központi szerkesztőség az egységes programrendszer segítségével nyilvántartja és kinyomtatja mind az osztauruszt (melyet elvileg minden speciális információkereső rendszer az OMIKK-ban makroteaurusznak fogad el), mind pedig a speciális teauruszokhoz szükséges kifejezéseket, a rájuk vonatkozó, már megállapított adatokkal együtt.

V. szakasz

8. Az ajánlásként használt jegyzékek alapján a speciális rendszerekben az egységes teauruszpépítést segítő programrendszer segítségével fokozatosan kidolgozhatják a nekik megfelelő ellenőrzött szótárt (amelynek nem kell feltétlenül elméie a teaurusznak megfelelő érettségi szintet). E szerkesztéskor azonban figyelembe kell venni az OMIKK osztauruszát és az egységes szerkesztési irányelveket.

A központi szerkesztőség gondoskodik az osztaurusz karbantartásáról, a teljes szójegyzék folyamatos kumulálásáról (esetleg a kifejezésekhez kapcsolódó összes adat tárolásáról), és szükség esetén befolyhat a speciális ellenőrzött szótárak szerkesztésébe illetve karbantartásába.

Milyen eszközök állnak rendelkezésre jelenleg a munkához?

Meglepőnek tűnhet, de valójában a felsorolt eszközök szinte mind a rendelkezésre állanak. Így rendelkezésre áll

- egy a gyakorlatban már többszörösen kipróbált teauruszpépítést segítő programrendszer (TSPR), melyet többek között az OMIKK számítógépén is használnak a Nemzeti Periodika Adatbázis és az OMIKK közös folyóirat-teauruszána szerkesztésére és karbantartására [5]. (Az OMIKK jelenlegi „tárgyszavainak” kezelésére is készült egy – egyelőre dokumentálatlan – egyszerű program. Ennek fejlettségi szintje azonban távol van attól, hogy teauruszok építésére is fel lehessen használni. A TSPR léte – és az, hogy az OMIKK gépén is operatív – céltalanná teszi a továbbfejlesztését; ez a munka ugyanis korántsem olyan egyszerű.)
- számos hazai teaurusz (pl. építésügyi, számítástechnikai, általános műszaki, kohó- és gépipari, növényvédelmi stb.), melyekből hasznos információ szerezh-

hető a szemantikai kapcsolatok megállapítására, hiszen az OMIKK-ban feldolgozott dokumentumok köre mindezekre a szakterületekre kiterjed [7].

- egy átfogó, a műszaki–tudományos és gazdasági élet fogalmaira kiterjedő fogalomosztályozási és kategorizációs rendszer [8], mely az OMIKK igényeinek megfelelően nagyobb ráfordítás nélkül módosítható.
- egy – folyóiratok osztályozására használt – osztaurusz, mely eleve az OMIKK igényeit figyelembe véve készült [5], és ugyancsak nagyobb ráfordítás nélkül módosítható.

Amire nem tértünk ki

Nem tértünk ki – eltekintve a gépesítés és az adatbáziskezelő programsomag kérdésétől – a szakértelemre és a munkaerőkapacitásra.

Abból indulunk ki – ha egyáltalán lehetünk derűlátók –, hogy mindkettőhöz fokozatosan, az előzőekben felsorolt szakaszok mentén lehet közelíteni. Az első szakasz talán már a jelenlegi lehetőségek mellett megvalósítható. A következő szakaszra felhalmozódnak némi tapasztalatok is, és az egyre összetettebbé váló feladatok megoldására fokozatosan bevonhatók további szakemberek.

A vázolt fejlesztés megvalósítása sok időt igényel, de ez lenne a helyzet minden más, komolyan vehető javaslattal is. Az OMIKK információkereső nyelveit – e nyelvek rendkívüli „nagyága” és a készítési tapasztalatok szűkössége miatt – csak évek hosszú során lehet kialakítani. Minden más, rövidtávú elképzelés illúziókon alapszik. Maga az osztaurusz elkészítése – tehát a III. szakasz végéig terjedő munka – legalább 2–3 évet igényel, a IV. szakasz további éveket, az V. szakaszról pedig általánosságokon kívül még nem sokat mondhatunk, annyira módosítani fogják az eredeti elképzeléseket a menetközben szerzett tapasztalatok.

A vázolt fejlesztési elképzelés előnye, hogy meglévő eszközökre (a TSPR-re, a folyóirat-osztauruszra, a már meglévő szókészletekre) épít, és fokozatosan valósítható meg, az egyes szakaszok végén már önmagában is használható végtermék keletkezik, a felhalmozódó tapasztalatok pedig módot nyújtanak az elkerülhetetlen helyesbítésekre.

Az OMIKK információkereső nyelveinek kialakítása a felvetett problémák fényében aligha lesz diadalmenet. Ez így lesz akkor is, ha a javasolt fejlesztési változatnak létezik jobb alternatívája. A legjárhatóbb út megtalálásához minden javaslatot érdemes tanulmányozni, amely elbírja a nyilvános közzétételt és az osztályozási és információkeresési szakma ugyancsak nyilvános kritikáját.

*

Megjegyzések és hivatkozások

1. BSO (Broad System of Ordering). Schedule and Index. Third Revision. Prepared by the FID/BSO Panel, Eric Coates, Geoffrey Lloyd, Dusan Simandl. — The Hague, Paris: FID, UNESCO, 1978. XIV. 102, 82 p. (FID Publication 564) továbbá magyarul:

DIENES Gedeonné: Az átfogó tárgyköri osztályozás = Könyvtári Figyelő, 30. köt. 3. sz. 1984. p. 285–293.

2. Bővebben lásd a III. részben!

3. A tezauszok terjedésével a hagyományos osztályozási rendszerek nem váltak feleslegessé. A korszerű dokumentációs intézetekben és gépesített könyvtárakban terjedőben van a természetes nyelven alapuló ellenőrzött szótárak és a hierarchikus, mesterséges nyelven alapuló osztályozási rendszerek (pl. az ETO, Dewey Tizedes Osztályozása) párhuzamos használata. A legújabb helyzetről tartalmaz felmérést és statisztikai adatokat egy német vizsgálat: BURKART, M., WERSIG, G.: Nutzung von Klassifikationssystemen in deutschen IuD-Stellen. Ergebnisse einer Umfrage. (Osztályozási rendszerek alkalmazása az NSZK információs és dokumentációs intézményekben. Egy körkérdés tapasztalatai.) — Berlin: Progris, 1981, október, 125 p.

4. Az információkereső rendszerekhez szükséges automatizált adatbáziskezelő programcsomagok nem tévesztendőek össze a tezauszépítést támogató programcsomagokkal. Az előbbieket illetően az OMIKK két olyan szoftvertermék — a CDS ISIS és az NTMIK DIALOG — birtokosa, melyek információkereső nyelvet kezelő képessége sajnos már ma elavultnak tekinthető, s még inkább az lesz, mire az online üzemmód a felhasználók számára is általánosan hozzáférhető lesz (ennek az üzemmódnak igazán csak akkor van értelme, ha a felhasználónak is a rendelkezésére áll). Az ISIS és a DIALOG ugyanis nem teszik lehetővé

- se a szintaktikai relációk deszkriptorláncon belüli használatát, ami pedig előfeltétele annak, hogy az OMIKK információkereső nyelveinek szókincsét ne duzzasszák fel az áttekinthetelenségig összetett kifejezésekkel;

- se az automatikus nemdeszkriptor-deszkriptor utalás „ki-és bekapcsolását” információkeresés közben (az ISIS semmiféle reláció automatikus kezelésére nem képes), ami pedig előfeltétele annak, hogy a kváziszinonimák okozta zajt szükség esetén kiszűrjék;

- se az automatikus deszkriptor-deszkriptor utalást, ami pedig előfeltétele az ún. böngésző keresésnek és nem utolsósorban a gépidő takarékos felhasználásának.

E hiányok véleményem szerint még sok keszvet okoznak majd az OMIKK szótárainak készítésekor.* Részletesebben lásd:

UNGVÁRY R.: A tezauszok automatizálása. A gépi tezausz-készítés és használat néhány kérdése = Tudományos és Műszaki Tájékoztató, 30. köt. 10. sz. 1983. p. 385–395.

5. Periodika osztályozási rendszer és tezausz. Szerk.: Ungváry R., Budapest, 1984. január. Országos Széchényi Könyvtár Fejlesztési Csoport. 40 p. (Kézirat)

6. SCHLANGER L.: A tezauszépítés számítógépes segítése. Budapest, OSZK Könyvtártudományi és Módszertani Központ, 1983. 220 p.

UNGVÁRY R.: Tezausz-technológia. Az információkereső tezauszok készítésének folyamata. Budapest, OSZK Könyvtártudományi és Módszertani Központ, 1979. 277 p.

7. Az elkészült hazai tezauszok felsorolása megtalálható: UNGVÁRY R.: A Magyarországon készülő tezauszok a nemzetközi szabványosítás tükrében = Tudományos és Műszaki Tájékoztató, 30. köt. 8/9. sz. 1983. p. 305–330.

8. Fogalomosztályozási Rendszer. Szerk.: Ungváry R. Budapest, KG INFORMATIK, 1978. 357 p.

- * Az OMIKK tervezi a DIALOG programcsomag olyan korszerűsített változatának az implementálását 1985-ben, amely — a rendelkezésre álló információk szerint — lehetőséget ad a felsorolt hiányosságok megszüntetésére. (— a Szerk.)

VAJDA Erik—UNGVÁRY Rudolf: Gondolatok és javaslatok az OMIKK információkereső nyelveinek fejlesztéséről

Az OMIKK-ban tervezett gépesítéshez szükség van az önerőből létrehozott adatbázisok információkereső nyelveinek fejlesztésére is. A jövő nem bízható kizárólag a szabad szövegszavas információkeresésre vagy az automatikus indexelésre. Ellenőrzött szótárakkal rendelkező információkereső nyelvekre is szükség van. E nyelveket azonban számos tényező befolyásolja; bonyolítja a helyzetet, hogy gyakran ugyanazt az állományt többcélúan kell indexelni. Aligha létezik egyetlen olyan nyelv, amely az OMIKK összes problémáját megoldhatná, hanem több, összehangolt nyelvre van szükség: részben természetes nyelvre alapozott nyelvekre, részben pedig szintetizáló osztályozó rendszerekre. A legfontosabb két indexelendő állomány közül a könyvtári a tervezett

COM-katalógushoz az eddig is használt ETO-t, az online-katalógushoz pedig koordinált indexelésre alkalmas nyelvet igényel. A referátumállományban végzett információbeszerzéshez az ETO és a szabad szövegszavas keresés mellett több szaktezauszra van szükség. Mindezeket a szótárakat „tartalmilag” az „Osztaurusznak” nevezett csúcstezausz koordinálná, „formailag” pedig az, hogy már kipróbált hazai tezausz-készítést segítő programrendszert (a TSPR-t) használnák fel a kidolgozásukra és karbantartásukra. A fejlesztési javaslat lényege a fokozatosság: az egyes kidolgozási szakaszok végén mindig olyan termékek (egyszerű szójegyzékek — deszkriptorjegyzékek szinonimarelációkkal — csúcstezausz — tezauszok) keletkeznek, melyek önmagukban is felhasználhatók már az egyes részrendszerekben.

* * *

VAJDA, E.—UNGVÁRY, R.: Ideas and proposals for the development of retrieval languages in OMIKK

Automation plans in the National Technical Information Centre and Library (Országos Műszaki Információs Központ és Könyvtár) and the establishment of its databases require development of retrieval languages. Free-text search and automatic indexing are not sufficient to cover all kind of information retrieval tasks: retrieval languages having controlled vocabulary are also necessary. Various factors have influence on these languages, multiple purpose indexing of collections being one reason for difficulties. An attempt to meet all needs by a single complex retrieval language seems to be a failure, one needs more mutually harmonized languages, that is, a mixture of natural language based systems and classification schemes. The library needs ongoing use of UDC for the planned COM-catalogue and a natural language based system for the online catalogue. The file of abstracts requires use of a set of special thesauri, in addition to free text search and UDC-based search facilities. As far as their contents is concerned, all these thesauri should be coordinated by a common roof thesaurus called „Classaurus” (in Hungarian: *Osztaurusz*); as to structure and methodology of thesaurus establishment, they are to be coordinated by the common use of a well-proved Hungarian software package for thesaurus construction and maintenance (TSPR). The work is to be done step by step. All individual development phases would result in well-defined products (simple lists of keywords — lists of descriptors with equivalence relations — roof thesaurus — special thesauri) available in and applicable to various system-elements as retrieval languages.

* * *

ВАЙДА, Э. — УНГВАРИ, Р.: Представления и предложения по развитию информационно-поисковых языков в ОМИКК

Планы по автоматизации в Государственном информационном центре и библиотеке по технике (Országos Műszaki Információs Központ és Könyvtár), ОМИКК требуют развития информационно-поисковых языков (ИПЯ) для создания и использования собственных баз данных. Поиск, основанный на свободных ключевых словах из текста, а также автоматическое индексирование не являются достаточными. Необходимо также

применение языков с контролируемым словарем. На желательные свойства этих языков влияют разные факторы, и положение усложняется с потребностью в многократном индексировании одних и тех же массивов. Попытка решения всех задач с помощью единственного комплексного ИПЯ сама по себе обречена на провал; требуются разные взаимогармонизирующие языки. Путь развития ведет к комбинации ИПЯ, основанных на естественном (венгерском) языке и классификационных системах. Для каталогов библиотеки в форме микрофильма на выходе из ЭВМ (СОМ) применение УДК является желательным и в дальнейшем; между тем каталоги библиотеки, доступные через терминальную связь, требуют ИПЯ на основе естественного языка. Поиск в массиве рефератов требует кроме УДК и возможностей поиска свободными ключевыми словами также тезаурусов разных тематических областей. Эти тезаурусы надо координировать: по содержанию, с помощью общего „кровельного” тезауруса (под названием „классaurus”, по-венгерски „*Osztaurusz*”) и по форме, путем применения венгерского пакета прикладных программ для создания и ведения тезаурусов: TSPR. Эти работы могут быть проведены постепенно, и таким образом в конце отдельных этапов развития создаются разные средства (простые списки ключевых слов, списки дескрипторов с указанием отношения эквивалентности — „кровельный” тезаурус — специальные тематические тезаурусы) сами по себе применяемые в разных подсистемах.

* * *

VAJDA, E.—UNGVÁRY, R.: Gedanken und Forderungen zur Entwicklung der Informations-recherchesprachen in OMIKK

Die geplante Automatisierung im Nationalen Informationszentrum und Bibliothek für Technik (Országos Műszaki Információs Központ és Könyvtár; OMIKK) erfordert zu den betriebseigenen Datenbasen auch die Entwicklung von Informationsrecherchesprachen. Die Aufgaben können nicht ausschliesslich durch Freitext-recherche oder automatisches Indexieren gelöst werden, es sind auch Sprachen mit kontrolliertem Vokabular notwendig, die jedoch durch sehr unterschiedlichen

Faktoren beeinflusst werden; die Lage wird noch durch die Forderung des Mehrzweckindexierens derselben Bestände erschwert. Der Versuch, allen Ansprüchen durch eine einzige, komplexe Sprache zu entsprechen, dürfte wohl zum Scheitern verurteilt sein; es sind mehrere, miteinander harmonisierende Sprachen erforderlich. Die Entwicklung sollte in Richtung einer Mischung von natürlichsprachigen und Klassifikationssystemen führen. Der Bibliotheksbestand erfordert für den geplanten COM-Katalog weiterhin die UDK und für den online-Katalog ein natürlichsprachiges System. Der Referatebestand erfordert neben UDK und Freitextsuche mehrere Thesauri. All diese Thesauri sollten „inhaltlich“ durch einen gemeinsamen – als „Klassaurus“ bezeichneten

Dachthesaurus, „formell“ durch die gemeinsame Verwendung eines schon erprobten Programmsystem ungarischer Herkunft für die Thesauruskonstruktion (TSPR) miteinander koordiniert werden. Die Durchführung dieser Arbeiten könnte durch ein stufenweises Vorgehen gekennzeichnet werden: am Ende jeder Entwicklungsphase entstünden Produkte (einfache Stichwortlisten – Deskriptorenlisten mit Äquivalenzrelationen – Dachthesaurus – Thesauri), die schon in sich in den verschiedenen Teilsystemen als Sprachen verwendet werden können.

*

Kedves Könyvtáros Kollégák!

Nyílt levélben fordulunk Önökhöz, azt remélve, hogy ennek több eredménye lesz, mint hagyományos üzleti propagandának. A cél továbbra is változatlan: egy kiadványunkra szeretnénk figyelmüket nyomtatékosan felhívni.

1982 – a Magyar Nemzeti Bibliográfia. Időszaki kiadványok repertóriumai decentralizálása – óta az OMIKK referáló folyóiratai (előbb műszaki lapszemlék majd szakirodalmi tájékoztatók) mellékleteként adja közre a műszaki szakterületek repertóriumát.

1983-ban a Magyar Könyvtárosok Egyesülete illetékes fórumain, több szakember részéről, és ennek nyomán az Országos Könyvtárügyi Tanács ülésén is, kemény bírálatot kaptunk, mert – úgymond – a könyvtáraknak oly fontos repertóriumhoz csak a külföldi szakirodalmat is referáló, „drága” referáló folyóiratok mellékleteként lehet hozzájutni, jóllehet szakmai csoportosításban, de szerzői mutató nélkül.

Megszívtük a bírálatot és 1984 januárjától közreadjuk a *Műszaki cikkek a magyar időszak kiadványokban c. repertóriumot*, amely szakcsoportos szabadlappos, szétvágható; osztályozott tételekből áll és amelyhez éves szerzői mutató készül. Ezek után az évi 300,- Ft-ért előfizethető, 12-szer megjelenő kiadványból könyvtáraink (ismertetőt és megrendelő levelet mintegy 2000 címre küldtünk) 78 azaz *hetvennyolc példányt* fizettek elő!

Ez a tény bátorított fel minket arra, hogy Önökhöz forduljunk. A továbbiakat bölcs megítélésükre bízunk, de hozzátesszük, hogy 150 megrendelt példány alatt a kiadványt 1985-től nem tartjuk fent!

A kiadvány megrendelhető:

1984. januárjától folyamatosan (1985-re is)

1984. júliusától folyamatosan (1985-re is)

1985. januárjától

Cím: OMIKK Kereskedelmi Főosztály
Értékesítési csoport
1428 Budapest, Pf. 12.

üdvözlettel

OMIKK
Szerkesztési főosztály
Referáló kiadványok szerkesztősége