

Folyamatábra tézaurusz alapján történi indexeléshez

HANS WELLISCH

A szerző - a műszaki irodalomból vett példán - a tézaurusz használatának általános érvényű módszerét ismerteti szemléletesen. Az eredeti cikk címe: A flow chart for indexing with a thesaurus (Journal of the American Society for Information Science, 23, vol. 1972. 3. no. 185-194. p.). A tömörítés HORVÁTH Magda munkája.

Szinte valamennyi indexelési munka - nagy, általános bibliográfiai szolgáltatások vagy szakkönyvtárak, tájékoztató központok céljaira - tézauruszon (szervezett, ellenőrzött szókészleten) vagy tárgyszójegyzéken alapul. A tézauruszok szerkesztése és naprakészen tartása ma már olyan szintű apparátussal történik, hogy a deskriptorok különböző elrendezésű jegyzékeit (display) bonyolult számítógép-programok segítségével elő lehet állítani, mielőtt a deskriptorok alapjegyzékének kiválasztása s hierarchikus vagy fogalmi relációik megállapítása megtörtént.

Az eredmény a tézaurusz, ez az erősen strukturált és formalizált indexelési segédeszköz, amelyet az indexelés során szigorú szabályok szerint kell felhasználni. A 2. ábraként bemutatott folyamatábra kísérlet a tézauruszokban rejlő valamennyi lehetőség teljes, módszeres kihasználásához szükséges lépések bemutatására. A folyamatábrát nagyrészt az Engineers' Joint Council Thesaurus of engineering and scientific terms (TEST) alapján alakítottuk ki. Választásunk azért esett erre, mert a TEST általános profilu, átfogó tézaurusz (ma még nincs sok ilyen), s az USA-ban erősen használják a természettudományi és műszaki dokumentumok indexelésére. De hangsúlyozni kívánjuk, hogy a folyamatábrát eredményesen lehet alkalmazni bármilyen más, jól strukturált tézaurusz vagy tárgyszójegyzék használatához is, pl. a Library of Congress rendszere szerinti tárgyszavazáshoz - bár a tárgyszórendszerekből természetesen hiányoznak a tézauruszok olyan elemei, mint pl. az NT és az RT megkülönböztetése; ezeket mind "lásd még" hivatkozással jelöljük. (A rövidítések feloldását lásd az 1. ábrán.)

1. ábra. A folyamatábrán és a szövegben használt kifejezések definíciója, a rövidítések feloldása.

DESKRIPTOR = egy tézauruszban jegyzékbe vett, engedélyezett kifejezés

AZONOSÍTÓ = egy tézauruszban jegyzékbe nem vett, engedélyezett kifejezés (pl. földrajzi, rendszertani, vállalati név)

TERM = kifejezés = fogalom jelölésére szolgáló szó (szavak)

BT = broader term = átfogó kifejezés (amely az NT-hez képest generikus)

DT = document term = dokumentumkifejezés (a dokumentum szövegében használt kifejezés)

IT = index term = indexkifejezés (a dokumentum indexelésére használt kifejezés)

NT = narrower term = szűkebb kifejezés (a BT-hez képest specifikus)

RT = related term = kapcsolt kifejezés

USE = "használd" - a hivatkozásoknál alkalmazott "lásd"

UF = use for = a "lásd" reciprok megfelelője, tehát annak jelzése, hogy az IT milyen kifejezés helyett használatos.

A legtöbb gyakorlott indexelő többé-kevésbé rutinszerűen, a folyamatábrán szemléltetett módon használja tézauruszát, de esetleg mégsem tesz meg minden logikailag lehetséges lépést, s így előfordulhat, hogy nem talál rá a megfelelő IT-re, egyszerűen azért, mert nem vizsgálta át módszeresen minden kifejezést - a BT-eket, az NT-eket és az RT-eket. Csak egy pillantás a folyamatábrán látható lépésekre hasznos kontrollt jelenthet, különösen amikor a dokumentumban foglalt tartalom megfelelő és teljes reprezentálására nehéz megfelelő IT-t találni. A kezdő indexelő számára a folyamatábra begyakoroltató segédeszköz lehet, amely bevezeti őt a tézaurusz mélyreható és teljes használatába. Ha elég sok dokumentumot indexelt már ennek segítségével, akkor a tézaurusz használatának majdnem valamennyi mozzanata feltételes reflexszé válhat. Ezzel nem állítjuk azt, hogy a folyamatábra használata révén az indexelés teljesen mechanikus rutinmunkává lesz; csak annyit, hogy segítséget nyújthat egy jól felépített tézaurusz valamennyi lehetőségének kihasználásához, még akkor is, ha a tézaurusz nem tartalmazza az összes megfelelő IT-t. Itt is, mint több kulcspontnál, az indexelő egyéni ítélt- és döntési képességére van szükség.

AZ INDEXELÉSRE KERÜLŐ DOKUMENTUMOK ÉRTÉKELÉSE, KIVÁLASZTÁSA

Az indexelő folyamat megkezdése előtt meg kell állapítani - a dokumentum futólagos, néha felületes átnézésével - hogy az illető dokumentumot fel-

2. ábra. A boxok feliratai

1. Start.
2. Tanulmányozd a dokumentumot.
3. Értékes-e a dokumentum a gyűjtemény számára?
4. Selejtezd ki.
5. Katalogizáld a bibliográfiai adatokat.
6. Olvasd el a dokumentumot, húzd alá a címben, a kivonatban és/vagy a szövegben azokat a DT-eket, amelyek szignifikáns fogalmakat jelölnek.
7. Nézd meg az első aláhuzott kifejezést a tézauruszban.
8. DESKRIPTOR-e a DT?
9. Jegyezd fel a kifejezést IT-ként.
10. Van-e a DESKRIPTORnak BT-je?
11. Nézd át a BT-k jegyzékét.
12. Van-e a DESKRIPTORnak NT-je?
13. Nézd át az NT-k jegyzékét.
14. Van-e a DESKRIPTORnak RT-je?
15. Nézd át az RT-k jegyzékét.
16. Van (mindhárom jegyzékben) IT-nek használható kifejezés?
17. A jegyzékben az utolsó kifejezés NT volt?
18. A jegyzékben az utolsó kifejezés RT volt?
19. DESKRIPTOR-e az utolsó aláhuzott kifejezés?
20. Vége.
21. AZONOSÍTÓ-e a DT?
22. Van-e "lásd" utalás a DT-ről szinonim DESKRIPTOR(OK)ra?
23. A DT-vel jelölt fogalom kifejezhető-e két vagy több DESKRIPTOR kombinációjával?
24. Készíts USE utalást a kifejezésnél a DESKRIPTOR(OK)ra.
25. Keress a DT-nek megfelelő szinonimá(ka)t szótárakban, segéd-könyvekben stb.
26. Találtál valamilyen szinonimát?
27. Hasonlítsd be a szinonim kifejezést (kifejezéseket) a tézauruszba.
28. DESKRIPTOR-e a szinonim kifejezés?
29. Keress a hierarchikus mutatóban egy lehetséges BT-t a kifejezés-hez.
30. Találtál bármilyen megfelelő BT-t?
31. Legyen a DT-ből IT?
32. Jegyezd fel a következő tézaurusz-revizio számára.

vesszük-e a gyűjteménybe, tehát indexeljük-e a későbbi visszakeresés céljaira. Az "érték" itt természetesen nem mindig a dokumentum belső értékét jelenti, hanem attól függ, hogy a dokumentumot egy általános referáló vagy bibliográfiai szolgáltatás (átfogó tárgyorientációjú, nagy számú klienturával), vagy pontosan körülírt információigényű, specializált használócsoport számára, pontosan definiált szakterület gyűjteményébe vesszük-e fel, pl. egy szakkönyvtárba vagy kutatást szolgáló tájékoztató központba. Ez a differenciálás kihat a dokumentum indexelésére szánt IT-k fajtájára és számára is, sőt még a folyamatára használatára is, ahogy majd később látni fogjuk. Így pl. az Engineering Indexbe felvételre érdemesnek ("értékes"-nek) és indexelésre méltónak ítélnénk egy, a műszaki szakemberek széles körét érdeklő dokumentumot. Lehet ez pl. egy cikk a hidépítésről vagy a cinkoxid-kinyerésről vagy bármilyen más műszaki témáról, amennyiben valami új információt tartalmaz és szakszerűen van megfogalmazva. Viszont egy szakgyűjteménybe, pl. egy kulturmérnöki gyűjteménybe csak a hidépítésről szóló cikket vesszük fel, a másikat "kiselejtezzük". Ezen nem azt kell érteni, hogy a dokumentumot megsemmisítjük, csak nem vesszük fel az indexelés folyamatába, vagy átadjuk egy másik intézménynek.

AZ INDEXKIFEJEZÉSEK MEGVÁLASZTÁSA

A dokumentum fizikai lokalizációjához elengedhetetlen bibliográfiai adatok rögzítése után a dokumentumot "elolvassuk" és elemezzük fogalmi tartalmát. A folyamatában 6-tal jelölt boxban látható "elolvadás" bővebb magyarázatra szorul. A dokumentumot legtöbbször nem kell betűről betűre végig olvasni. Sokszor autoreferátum kíséri; a cím, az összefoglalás, a bevezetés és a végkövetkeztetések legtöbbször elegendő támpontot nyújtanak az indexelőnek a dokumentum tartalmát kitevő fogalmak megállapításához. Ugyanebben a boxban áll még az is: "huzd alá a DT-eket". Ez sem betűszerint értendő. Az "uniterm" indexelés előző korszakában ugyan az volt az előírás, hogy "az indexelő készítsen jegyzéket a szöveg alapján a dokumentum tartalmát reprezentáló szavakból". A "szignifikáns" vagy "fontos" főnevek, mellénevek és igék pusztán mechanikus aláhúzása aligha vezet el a fogalmak teljes jegyzékéhez a későbbi visszakeresés számára. A KWIC-indexelés is csak a szerző esetleges szóhasználatán alapul, s ez nem szükségképpen fedi a dokumentumban foglalt eszméket és fogalmakat. Az indexelhető fogalmak és az ezeknek megfelelő IT-k megtalálása sokkal összetettebb folyamat. A téma bőséges irodalmából Oliver és szerzőtársai vizsgálatát ragadjuk ki, amelynek során szintén folyamatában szemléltetik az elemzés és indexelés műveleteit. Folyamatábrájukat be lehet helyettesíteni a jelenlegi folyamatába bevezető stádiumainak sürített ábrázolásaiba (1-7). Az Oliver-féle folyamatára egyik utolsó operáció-felirata így hangzik: "Vizsgáld meg a deskriptorokat, hogy beilleszthetők-e a rendszerbe", s a szükséges lépések ezen pontján kezdődik a jelenlegi elemzés. Éppenséggel nem akarunk olyan szellemi folyamatokat szuggerálni, amelyek egy referátum (abstract, rövid kivonat) megfogalmazásához vezetnének, de nem zárjuk ki, hogy valaki - esetleg nem maga az indexelő - elkészíti a dokumentumban megtestesült eszmék és fogalmak

elsőfoku kivonatát, amelyből az indexelés céljaira történő másodfoku sűrités a leginkább szignifikáns szavak aláhuzásában áll, annak jelzéseként, hogy az illető szavak esetleg bekerülhetnek a végleges IT-jegyzékbe (a dokumentumnak adott összes IT "csomagjába").

PÉLDA A FOLYAMATÁBRA HASZNÁLATÁRA

Az IT-jelölteket be kell hasonlítani a helyileg használt tézauruszba. Különböző szintű részletezés lehetséges; a használók igényei döntően befolyásolják a kifejezések megválasztását. Egy általános indexelő szolgáltatás átfogóbb, általánosabb (de nem szükségképpen kevesebb) kifejezéssel fog élni, míg egy specializált tájékoztató központnak olyan "szűkebb kifejezések"-et kell majd használnia, amelyek explicite nincsenek meg a dokumentumban vagy kivonatában, de amelyeket az indexelő hasznosaknak vélt a specializált klientura számára a dokumentum visszakereséséhez. A jelenlegi példában a színhely egy szakterületi kutatással és gyakorlati alkalmazással megbízott természettudományi és műszaki szakembereket kiszolgáló tájékoztató központ. A folyamatábra kijelöli a szükséges lépéseket a valamennyi (vagy majdnem) potenciálisan hasznos kifejezés megtalálásához egy tézauruszban, ahol a kifejezések összes hierarchikus és korrelatív kapcsolatokkal együtt fel vannak tüntetve. A központi cél egy kifejezésnek IT-ként való kijelölése. Eddig a pontig az összes utak visszafelé vezetnek, különböző lehetőségek megvizsgálása után ezeknek az IT-knek a megtalálásáért.

A konkrét példát taláalomra választottuk az Engineering Index 1967. decemberi számából. A TEST megjelenési éve ez, így joggal feltételezhetjük, hogy a TEST terminológiája megegyezik egy ugyanabban az időszakban publikált műszaki cikk terminológiájával.

Az elemzett szöveg (a dokumentum címleírása és kivonata) így hangzik:

Computer model of slag-fuming process for recovery of zinc oxide, H.H. KELLOGG. Met. Soc. of AIME-Trans. v. 239, no. 9, Sept. 1967, p. 1429-49. Model of slag fuming process for recovery of zinc oxide from lead blast furnace slags, adapted to solution by digital computer, is presented; behavior of model is compared with operation of American Smelting and Refining Company's furnace at El Paso, Tex.; model is used to predict effects on performance of changes in coal rate, temperature of secondary air, and composition of coal.

(Magyar fordításban: Cinkoxid kinyerésére szolgáló salakgázosítási folyamat számítógépes modellje... Bemutatja digitális számítógépes megoldásra adaptált modelljét olyan salakgázosítási folyamatnak, amely cinkoxid kinyerésére szolgál ólom-nagyolvasztó salakjából; a modell viselkedését összeveti az ASRC kohójának (El Paso, Texas) működésével; a modell segítségével előre jelzi azon hatásokat, amelyeket a szén arányában, a szekunder levegő hőmérsékletében és a szén összetételében végrehajtott változások a teljesítményben előidéznek.)

A fenti cimleírás és kivonat megfelelő elsőfoku sűrítésnek bizonyult; a bennük aláhuzott szavakat választottuk ki a téma meghatározása szempontjából leginkább szignifikáns indexelési és visszakeresési kiindulópontokul. A választás természetesen az indexelő döntése nyomán jött létre, de joggal feltehető, hogy lényegbevágó eltérések nem fognak adódni, ha az indexelés alapjául szolgáló kifejezéseket egy rövid indikatív kivonatból vesszük, amelyben viszont a kifejezések magából a dokumentumból származnak, tehát DT-k. Amint a későbbiek során látni fogjuk, a tézaurusszal történő egybevetés után a DT-k nem mindig bizonyulnak a legjobb IT-knek, valamint akkor sem, ha az IT-k száma előre korlátozott.

Az aláhuzott kifejezéseket egyenként összevetjük a tézaurusz kifejezéseivel (alakutani eltérések, pl. egyszám, többesszám, helyesírásvariánsok stb. figyelmen kívül maradnak, természetesen a tézauruszban lévő alakok használandók). Az első kifejezés a (cimben szereplő) COMPUTER MODEL. Ez komplex kifejezés, hiszen egynél több szóból áll, mindegyik jelölhet egyetlen külön fogalmat is. Valóban, ilyen komplex formában nem szerepel a TEST-ben deskriptorként, de még azonosítóként sem. A következő lépés tehát a kifejezés szinonimájának megtalálása pl. egy általános értelmező vagy szakszótárban, ahol a kifejezések definíciójánál gyakran használnak szinonimákat. A komplex kifejezéseknél lehetséges módszer még az, hogy az összetevő kifejezéseket egyenként keressük meg a tézauruszban s ezeknek a kombinációi között rábukkanhatunk a jegyzékben nem szereplő kifejezés szinonimáira. A példában ezt az alternatívát választva, kezdjük a COMPUTER-nél. Az első két kombináció: COMPUTER COMPONENTS és COMPUTER DRIVEN PUNCHES nem megfelelő, de a következő deskriptor: COMPUTERIZED SIMULATION látszik a "Computer models" legközelebbi megfelelőjének és ezért indexkifejezésként feljegyezzük. A kifejezésnek két BT-je van (MATHEMATICAL MODELS és SIMULATION), amelyeket más rendszerekben automatikusan IT-knek vennék fel, azt a szabályt követve, hogy a "felsőbbrendű" generikus kapcsolatokat mindig kiírják. Így járnak el pl. az Euratom indexelői, ahol a legtöbb IT tágabb generikus BT-je is kiírásra kerül. Ha ezt az eljárást akarjuk elfogadni, akkor a folyamatábrát módosítani kell úgy, hogy közvetlen összeköttetés jöjjön létre az ábrán 10-zel jelölt ("Van-e a deskriptornak BT-je") döntés és a 9-cel jelölt operáció ("Jegyezd fel a kifejezést IT-ként") között, figyelmen kívül hagyva a folyamatábrán látható értékelő mozzanatot (11). Megjegyzendő: a BT alatti automatikus IT-kiírást csak azzal a feltétellel lehet alkalmazni, hogy az összes BT valóságosan generikus kifejezés a deskriptorhoz képest, azaz genus/species reláción alapuló hierarchiában fölérendelt szintet reprezentál; az egész/rész, ok/okozat vagy cél/eszköz kapcsolatokat sohasem volna szabad BT-nek feltüntetni, csakis RT-nek. Sajnos nem minden tézaurusz veszi figyelembe ezt a szabályt, maga a TEST sem. A cikk példáiiban azonban minden BT-t értékeltünk abból a szempontból, hogy a potenciális használóknak értéke-e a felvételük IT-ként, ahogyan az NT-k (13) és az RT-k (15) esetében az értékelés és választás elkerülhetetlen. A konkrét példában a COMPUTERIZED SIMULATION alatt talált két BT közül csak a SIMULATION-t vettük figyelembe.

3. ábra. A szövegben elemzett kifejezések (angol betűrendben)

DT-k (a kivonat szövegében szereplő kifejezések)	Engi- neering Index	IT-k (indexkifejezések) a tézaurusz módszer használatával származtatva		a végleges indexbe kiválasztva
1	2	3	4	
Blast furnace (nagyolvasztó)		Blast furnaces (nagyolvasztók)	BLAST FURNACES	
		Burning rates (égési sebesség)	Burning rates	
Coal (szén)		Coal gasification (széngázosítás)	Coal gasifica- tion	
Coal rates (szénarány)		Coal + Rates (szén + arány)		
Composition (összetétel)		Composition (összetétel)		
Computer model (számítógépmodell)		Computerized simulation (szá- mitógépes szí- muláció)	COMPUTER- IZED SIMULATION	
		Computers (számítógépek)		
Digital computer (digitális számi- tógép)		Digital computers (digitális számi- tógépek)		
		Digital simulation (digitális szimulá- ció)		
		Furnaces (Kemencék)		
		Heat of vapor- ization (elgő- zölgesi hő)	Heat of vapor- ization	
Lead (ólom)	Lead	Lead (ólom)	LEAD	
		Materials recovery (anyagok kinyeré- se)		

3. ábra (folyt.)

DT-k (a kivonat szövegében szereplő kifejezések)	Engi- neering Index	IT-k (indexkifejezések) a tézaurusz mód- a végleges szeres használatá- indexbe val származtatva kiválasztva	
1	2	3	4
Performance (teljesítmény)		Metallurgical furnaces (kohóke- mencék)	
		Models (modellek)	
		Performance (teljesítmény)	
		Production models (produkciós model- lek)	Production models
Recovery (kinyerés)			
Secondary air (szekundér levegő)		Simulation (szimuláció)	
Slag fuming (salagfűzés)		Slags + Vapor- izing (Salak + elgőzölögtetés)	SLAGS + VAPORIZING
	Smelting (olvasztás)		
Temperature (hőmérséklet)		Thermal re- covery methods (termális ki- nyerési módsze- rek)	Thermal re- covery methods
		Vaporizing (elgőzölögtetés)	Vaporizing
Zinc oxide (cinkoxid)		Zinc oxides (cinkoxidok)	ZINC OXIDES
ASRC ^x El Paso, Tex.		ASRC El Paso, Tex. US.	ASRC
15	2	25	12

A COMPUTERIZED SIMULATION NT-i között találjuk a DIGITAL SIMULATION-t is, amelynek az első szava a kivonatban mint "Digital computer" fordul elő, ezért mint lehetséges IT feljegyzésre került. Az RT-k között van a COMPUTERS is. Előálló aposztróf mutatja, hogy ennek a kifejezésnek megvannak a maga BT-i, NT-i és RT-i is. Ha a tézauruszban megnézzük a COMPUTERS deskriptort, ezt látjuk; BT DATA PROCESSING EQUIPMENT adatfeldolgozó berendezés, amely itt nyilván nem felel meg, és egy NT DIGITAL COMPUTERS, amely megegyezik a kivonatban található kifejezéssel; a többi NT-k és RT-k nem megfelelőek.

Az elemzést a cimben talált SLAG FUMING kifejezéssel folytatjuk; megint csak komplex kifejezés, s szintén hiányzik a TEST-ből. Ezért alkatrészeire bontottuk a fentebb vázolt módszer szerint. A SLAGS kifejezés megtalálható hivatkozással; RT METALLURGICAL FURNACES. Ez a kifejezés nincs bent szószertint a kivonatban, mégis alkalmasnak látszik - eléggé általános szinten ugyan - IT-nek. Ezt is aposztróffal látták el a TEST-ben, mert van egy BT-je: BT FURNACES és NT-i közül az egyik NT, BLAST FURNACES, a kivonatban is feltűnik.

Itt jegyezzük meg, hogy a TEST tézaurusz külön aposztróffal jelöli azokat az NT-ként vagy RT-ként felsorolt deskriptorokat, amelyek maguk is tagjai egy hierarchiának saját BT-kkel, NT-kkel és RT-kkel. (Sajnos, a TEST-hez fűzött magyarázatban az áll, hogy előálló aposztróf egy kifejezésnél csupán további NT-kre utal, ezzel szemben ténylegesen a legtöbb aposztrófozott kifejezésnek vannak BT-i, NT-i és RT-i, amelyek közül bármelyik hasznos lehet mint deskriptor.) Ha ez a jelölésmód hiányoznék a TEST-ből (mint sok más tézauruszból), akkor a BLAST FURNACES kifejezést később mint aláhuzott kifejezést találtuk volna csak meg, de a SIMULATION kifejezés (a COMPUTERIZED SIMULATION BT-je) kiesett volna a lehetőségek közül.

A FUMING (a cimben talált SLAG FUMING kifejezés második eleme) nem található meg a TEST-ben. Specifikus kifejezés, amelyet csaknem kizárólag a kohászatban használnak és még specifikusabb a "slag"-hez kapcsolatan a cikkben tárgyalt cinkoxid-kinyerés módszerének vonatkozásában. A legközelebb álló kifejezés a "vaporization" elgőzöltetés lenne, amely meg nem engedett kifejezésnek található meg a TEST-ben, utalva a VAPORIZING kifejezésre; ennél van egy NT COAL GASIFICATION, amely egyik lehetséges IT-jelölt, hasonlóan a még alkalmasabb RT HEAT OF VAPORIZATION (mivel a kivonat tartalmazza a hőmérsékletet, amelynél a folyamat lezajlik). A VAPORIZING-nek, mint a FUMING kvázi-szinonimájának használata szükségessé teszi, hogy az indexelő saját tézauruszmunkapéldányán feljegyezze ezt az indexelési döntést:

Fuming
USE Vaporizing
Vaporizing
UF Fuming

(lásd a 24. boxot).

A RECOVERY deskriptorként ugyan szerepel a TEST-ben, de a felhasználására vonatkozó megjegyzés (scope note) arra utasítja az indexelőt, hogy lehetőleg specifikusabb kifejezést használjon. Ezek között van a THERMAL RECOVERY METHODS és a MATERIALS RECOVERY, mindkettő megfelelő IT a dokumentum számára, bár ilyen formában nem szerepelnek benne.

ZINC OXIDES és LEAD lényegében azonosítók, bár a TEST-ben deskriptorként szerepelnek. Egyik NT-jük vagy RT-jük sem használható itt, BLAST FURNACES a következő aláhuzott kifejezés, már feljegyeztük mint a METALLURGICAL FURNACES NT-jét. A SLAGS kifejezés is elintézetést nyert már fentebb. Más rendszerekben a LEAD és a SLAGS kapcsolatjelölőkkel (links) való kapcsolása (linking) ajánlatos lenne, különösen mivel a COAL is lehetséges IT, s így a "Coal slags" téves egybeesés adódhat az olyan keresésnél, amely a COAL, LEAD, SLAGS és ZINC OXIDE deskriptorokat tartalmazza. Ez csak a megfelelő kapcsolatjelölőkkel kerülhető el. A tézauruszok azonban még nem rendelkeznek két vagy több egymáshoz viszonyuló deskriptor kapcsolatjelölő (linking) módszereivel (ez ui. teljesen a használt visszakereső rendszer jellemzőinek függvénye).

A következő aláhuzott kifejezés, DIGITAL COMPUTERS, már a korábbiak során feljegyzésre került.

Az AMERICAN SMELTING AND REFINING CO. azonosítható, ugyanigy EL PASO, TEXAS, amelyhez még hozzá lehetne tenni az ország nevét, tehát az USA-t, mivel minden település elvileg kereshető az állam, az ország neve alatt is. Más szavakkal: az azonosítók BT-it is automatikusan fel kellene jegyezni (különösen a földrajzi és rendszertani nevek esetében).

A PERFORMANCE kifejezést is aláhúzták s valójában deskriptor is, de IT-ként való alkalmazása megkérdőjelezhető általánossága és "üressége" miatt. Mivel mind a tézauruszban, mind a kivonatban előfordul, feljegyzést nyer mint lehetséges IT.

A COAL RATE nem található meg deskriptorként, itt is a már ismert módon a komplex kifejezés lehetséges kombinációját vesszük: COAL + RATE(S). Mindkettő meg is található deskriptorként: RATES (PER TIME) alatt megfelelő NT is adva van; BURNING RATE. (Megjegyzendő, hogy a magyarázó "per time" kiegészítés a homonim RATES (COSTS) ár kifejezés miatt szükséges.) Ha ezt a kombinációt a jövőben IT-ként és visszakeresési pontként használjuk, akkor az indexelőnek ismét hivatkozást kell erről készítenie tézaurusz-munkapéldányán:

Coal rate

USE COAL + Rates (per time)

A TEST-ben van már ilyen analógia: "Dissolved oxygen" USE Dissolved gases+Oxygen cseppfolyósított gázok + oxigén . Az ilyen kiegészítő hivatkozások felvétele a nyomtatott tézaurusz későbbi revideált formájában

fontos szerkesztési és kiadói feladat. Erősen hangsúlyozandó, hogy az indexelők módszeresen rögzítsék a tézaurusz által nem fedett indexelési döntéseket és az ebből adódó USE és USED FOR reciprok hivatkozásokat tézaurusz-munkapéldányaikon.

A TEMPERATURE kifejezés, bár érvényes deskriptor, mégis túl általános, de van egy RT-je: HEAT OF VAPORIZATION, amely kapcsolódik a már korábban levezetett VAPORIZING kifejezéshez. (Itt nem hagyható említés nélkül a TEST következtelen eljárása, amely összefügg a tézaurusz több terminológiai hibájával; ha ui. a "Heat of vaporization" érvényes deskriptor, akkor miért meg nem engedett kifejezés a "Vaporization", s miért kell helyette a "Vaporizing" kifejezést használni, ami nem is egészen azonos értelmű?)

A SECONDARY AIR nem deskriptor a TEST-ben, más deskriptorok kombinációjával sem lehet kifejezni. A folyamatábra jobb oldalán jelzett műveletekre és döntésekre kerül most sor. A kohászati szakszótárban egy lehetséges szinonima utáni keresés eredménytelen maradt. Pedig kell hogy legyenek megfelelő generikus kifejezések a hierarchikus mutatóban (Hierarchical Index), ahol a deskriptorok szisztematikus csoportokba rendezettek, mert végül is minden kifejezés beosztható valamely átfogóbb hierarchikus kifejezés alá. Ezért áll a 29. sz. boxban az ábrán ez a szöveg: "Keress ... egy lehetséges BT-t a kifejezéshez", mivel itt már nincs kilátás NT-k vagy RT-k megtalálására hanem bármilyen megfelelő kifejezés csakis átfogóbb generikus kifejezés lehet valamilyen szisztematikus hierarchiában; bár ugyanakkor a tézauruszba fel nem vett DT "fö-lött" a lehető legalacsonyabb hierarchikus fokon álló kifejezés is túl általános lesz, ahogyan a konkrét eset is mutatja.

A szekundér levegő a Furnaces egy jellemzője, amelyet ha megnézünk a hierarchikus mutatóban, eljutunk a METALLURGICAL FURNACES és a BLAST FURNACES sokkal specifikusabb kifejezésekhez; de más kifejezések nem találhatóak, amelyek alá helyesen be lehetne osztani a "szekundér levegő"-t. Kérdés, hogy fel kell-e venni IT-nek a SECONDARY AIR-t? A pozitív döntés alapja lehet: a) az indexelő itélete, aki ismeri használóinak visszakeresési szokásait, terminológiáját, vagy b) a kifejezés előfordulásának gyakorisága más, rokon témájú dokumentumokban. Ez utóbbi lehet különösen az új kifejezések felvételének kritériuma. Azt a döntést hozták (ezt egyébként nem mutatja a folyamatábra), hogy a kifejezést függőben tartják, amíg későbbi előfordulásai el nem érnek egy bizonyos küszöb-számot, mondjuk tizet. Miután ez a kifejezés tizszer megjelent a különféle dokumentumokban, érvényes IT-nek veszik fel és bejegyzik az indexelő tézaurusz-munkapéldányába. Ha a tézauruszt módszeresen naprakész állapotban tartják (a kurrens kiadást használó indexelők döntéseinek egyesítésével), akkor a kifejezés felvételéről értesíteni kell az új tézaurusz szerkesztőit. Konkrét példánkban feltételezzük, hogy a SECONDARY AIR nem igényelt, keresett kifejezés. Az indexelési eljárás a COMPOSITION kifejezéssel folytatódik. Ez deskriptor ilyen formán: COMPOSITION (PROPERTY), mindenféle alkalmas NT-k és RT-k nélkül, de a jelen dokumentum indexelésére megkérdőjelezhető általánosságá miatt: ha

egyáltalán, akkor kapcsolni kellene a COAL-hoz, amelyik az utolsó aláhuzott kifejezés s amelyet már felvettünk az IT-k jegyzékébe egy másik kifejezés-kombinációban. Itt véget ér a tézauruszba behasonlitandó, aláhuzott kifejezések sora.

A VÉGLEGES INDEXKIFEJEZÉSEK MEGVÁLASZTÁSA

Az előző fejezetben ismertetett indexelési eljárás eredményeit a 3. ábra 3. oszlopában rögzítettük. A 15 eredeti aláhuzott DT-hez még 10 kiegészítő kifejezés járult, amelyekből 9 deskriptor, 1 pedig azonosító. A legtöbb visszakereső rendszerben ez túl nagy számú kifejezés lenne olyannyira, hogy a téves egybeesések valószínűsége is igen magas lenne, még akkor is, ha a nem kívánatos korrelációkat megfelelő kapcsolatjelölőkkel kiküszöbölnék. A legtöbb indexelő eljárásnál kialakult a felhasznált kifejezések számának korlátozása, továbbá az a gyakorlat, hogy különválasztanak egy bizonyos számú fő vagy primér kifejezést, amelyeket a publikált indexekben tüntetnek fel, a fennmaradó szekundárnak minősített kifejezéseket komputer-memóriában tárolják a gépi visszakeresés céljaira. Ennek közismert példája a MEDLARS gyakorlata.

Tegyük fel, hogy a jelen visszakeresési rendszer vezetősége mondjuk 12 indexkifejezésben limitálta a megadható kifejezések számát, ezek közül 5 lehet fő kifejezés. Ebben az esetben választani kell az ismertetett indexelési eljárással nyert 25 IT közül, s ez a választás - mint már említettük - nagy mértékben függ a szolgáltató funkciójától. A példa kapcsán feltételeztük, hogy az indexelést egy kohászati tájékoztató központ számára végezték, ahol a használókat főleg a színes fémek érdeklik. Az ábra 4. oszlopában felsorolt IT-eket éppen e szempontok figyelembevételével választottuk ki.

Nem állítjuk, hogy az IT-k végleges jegyzéke a legjobb vagy az egyedül lehetséges. Mégis tényszerűen megállapítható, hogy a tézaurusz módszeres használata nyomán nyert legtöbb kifejezés specifikusabb a kivonat kifejezéseinél. A BLAST FURNACES, LEAD és ZINC OXIDES (ez utóbbi kettő lényegében azonosító) kifejezéseken kívül az összes többi pontosabb: COMPUTERIZED SIMULATION jobb mint Computer models, mivel egyértelmű (egy komputermodell lehet egy komputernek a modellje is). COAL GASIFICATION tömörebb, mint Coal vagy Coal rate, THERMAL RECOVERY METHODS pontosabb, mint az egyszerű Recovery, stb. Az Engineering Index a cikk megírása idején a "minden cikket egy kifejezés alatt" indexelési koncepciót követte, amelyet legfeljebb csak egy altárgyszó erejéig módosítottak. Nem felelne meg a játékszabályoknak, ha ezt a gyakorlatot egybevetnénk a tézaurusz alapján történő részletező indexeléssel, mégis érdekes megjegyezni, hogy az Engineering Index által adott SMELTING tárgyszó (a 3. ábra 2. oszlopában) csupán a cég nevének részeként szerepel a kivonatban, s a tézauruszban sem található meg még asszociatív kifejezéseként sem a kivonatban szereplő IT-k bármelyikéhez kapcsolódóan. S enyhén szólva megkérdőjelezhető az is, miért kell a cink kinyeréséről szóló cikket egyedül "Lead" alatt indexelni.

A szerző még további két cikk indexelését mutatja be szövegesen is, ábrákon is az elsőhöz hasonlóan. Ezek közül néhány érdekesség: a második cikknél a TEST-ből hiányzó kifejezést (TRAPEZOIDAL) az elemzés után új deskriptorként vették fel. A harmadik cikknél pedig a tézaurusz használata részben tulságosan redundáns, részben tulságosan összetett, mondat szerű deskriptorokhoz vezetett (pl. SOLID PROPELLED ROCKET ENGINES = Szilárd meghajtású rakétamotorok), amelyek alapelemeikre bontva nagyobb visszakeresési értéket eredményeztek volna. Ezeknek kapcsán ismét rámutat a TEST néhány hiányosságára és következetlenségére, némely deskriptor kétértelműségére. Az indexelőknak feljegyzéseket kellene ezekről készíteniük a tézaurusz következő javított kiadása számára. Például olyan kifejezések kiküszöbölése lenne kívánatos, amelyek azonos témák esetében szinonimaszerű kifejezések alatt szórják szét az információt. Világosan kimondva a tézauruszhasználat rögzített lépései csak bővítésekhez, új deskriptorok vagy hivatkozások megalkotásához vezetnek, hanem a nem kívánatos deskriptorok, valamint olyan deskriptorok törléséhez is, amelyeket már nem használnak (jó példa erre a "Laser" kifejezés behelyettesítése az "Optical maser" helyébe), s általában a terminológia pontosításához, a kifejezések szűkítéséhez.

A TÉZAURUSZ KÜLÖNBÖZŐ ELRENDEZÉSŰ JEGYZÉKEINEK (DISPLAY) HASZNÁLATA

Első látásra a folyamatábrán bemutatott munkafolyamat valószínűleg igen terhesnek, időtrablónak látszik. De még a legtapasztaltabb indexelő is legtöbbször öntudatlanul ugyanezt az eljárást követi, hogy a lehető legalkalmasabb IT-re rátaláljon. Az indexelési segédeszközként használt folyamatábra nem csak a tézaurusz következetes használatát biztosítja, hanem a tézaurusz különböző elrendezésű jegyzékeinek jó használatát is. Kérdésessé tették már, vajjon kellene-e egyáltalán ezek a különböző elrendezésű jegyzékek, vagy csak egyszerűen azért vannak, mert olyan könnyen lehet őket számítógéppel előállítani, ha az alapdeskriptorokat BT-ikkel, NT-ikkel és RT-ikkel már ugyanis betáplálták a komputerbe. A folyamatábra azt is bemutatja, hogy - legalábbis a TEST esetében - valahányszor is szempontru jegyzék hasznos, bár kétségtelenül a deskriptorok betűrendes jegyzéke a legkedveltebb. A hierarchikus mutatóra akkor van szükség, ha a tézauruszban nem található kifejezéssel állunk szemben. A permutált mutató elvezeti az indexelőt az összetett deskriptorok második, harmadik stb. szavához. A permutált mutatót jól használhatnák a keresők is a visszakeresések alkalmával, bár tudomásunk szerint ma még nem veszik igénybe. Nem valószínű ugyanis, hogy egy elfoglalt kohászati szakember szívesen babrálna egy ilyen visszakeresési eszközzel, hiszen nem azonnal vezeti el a kívánt dokumentumhoz, csupán közbenső lépés a kérdés-kifejezéstől a releváns dokumentumig vezető hosszú, rögzös uton. A Subject category index (Tárgykategóriák mutatója) - a TEST többnyire a COSATI-kategóriákat használja - talán csak a témakörök ilyenfajta elrendezéséhez szokott használóknak jó, mivel a kategóriák önkényesen szelektáltak és egyedül szűk specializált használói gárda érdeklődését reprezentálják.

FELHASZNÁLÁS

A folyamatábrát módszeres tézauruszhasználati segédeszköznek szántuk - emberi munkaerők számára. Fel lehet-e használni számítógépes indexelésre is? A válasz sajnos egyértelműen nem, legalábbis belátható időn belül. Ha egy számítógép képes lenne IT-k kijelölésére ilyen viszonylag egyszerű folyamatára segítségével, akkor beköszöntené az indexelés új korszaka. Közismert tény, hogy sajnos nem állunk ilyen jól. S bár az is igaz, hogy a folyamatára legtöbb döntését és műveletét végrehajthatja egy megfelelően programozott számítógép (amelyben a tézauruszt géppel olvasható formában tároltuk), mégis vannak döntő lépések, ahol emberi ítélőképességre van szükség. Az első ilyen döntés nyilvánvalóan a dokumentum kiválasztása a gyűjtemény számára, de mivel ez nem tartozik szorosan az indexelési folyamathoz, figyelmen kívül hagyható. Abban egyébként biztosak vagyunk, hogy a választás egy dokumentum felvétele vagy kiselejtezése között mindig is értelmes emberi lény munkája marad.

Értékitéleten alapuló kritikus döntést kell hozni ebben a kérdésben: "Van-e a jegyzékben egyáltalán IT-ként megfelelő kifejezés?" Amint láttuk, a tézauruszban vannak NT-k és RT-k, amelyek nincsenek szó szerint említve a dokumentumban vagy kivonatában, s mégis kiválóan alkalmas IT-k. Sem az elnagyolt KWIC-indexek, sem a sokkal finomabb tömörítő vagy csoportosító módszerek nem képesek még távolról sem megközelíteni egy jó indexelő problémamegoldásait. A tézauruszban nem szereplő kifejezések rutinszerű kikeresése (szinonima-szótárakból vagy magából a dokumentumból, esetleg előzőleg indexelt dokumentumok asszociatív IT-iből), az előfordulási gyakoriság statisztikai rögzítése - mindez túl kevés vagy túl sok (tehát nem releváns) kifejezés-jelöltet eredményezne. A helyzet ahhoz hasonlít, amikor a jegyzékben nem szereplő kifejezések lehetséges BT-i után kell kutatni egy hierarchikus jegyzékben. Nem minden kifejezés tartozik csupán egy hierarchia alá, de a legjobb esetet véve is, a kifejezést durván helyettesítő BT eldöntése is értékitéleteken nyugszik - uti figura docet.

A harmadik pont, ahol emberi ítélőképességre van szükség, mikor arról kell dönteni, hogy egy eddig jegyzékben nem szereplő kifejezést IT-ként kell-e használni vagy sem. Itt még elképzelhető olyan számítógépprogram, hogy a jegyzékben nem szereplő kifejezéseket a számítógép kiküszöböli mint IT-eket, de emellett tárolja, és ha előfordulásuk eléri az előre megállapított küszöbértéket, akkor érvényes IT-kként felveszi őket.

Egyelőre úgy látszik, hogy a tézauruszok szerkesztését támogathatják számítógépprogramok, felhasználásuk azonban nagymértékben emberi lények döntéseitől függ. Másfelől azonban nem csak magas képesítésű szakemberek használhatják sikeresen a tézauruszt. 1971-ben felmérték, hogy a gyűjtemény profilját jól ismerő, jól strukturált tézauruszokkal dolgozó indexelők 80 százalékban el tudták végezni munkájukat szakértők bevonása nélkül. Ujfajta szervezéssel sokat át lehet hárítani az ún. technikai személyzetre az indexelési munkafolyamatokból.