

ÚJSZERŰ TEZAUROSZSZERKEZET KIALAKÍTÁSA

Molnár Imre

Építésügyi Tájékoztatási Központ

A tájékoztatási rendszerek folyamatos tevékenységében felhasználható tezauszok kialakítása sok követelmény egyeztetését kívánja meg a tezauszszervezet megválasztásánál. A legfőbb és legfontosabb követelmény - természetesen - az információigények minél jobb kielégítése. E cikk célja az, hogy a tezauszépítési munkálathoz használható és hatékony segédeszközt adjon az indexelést végző, az információt kérő, a kéréseket szabványosító és az információkereső szakemberek kezébe.

A betürendes
szerkezet

A gyakorlati információszolgáltatás céljára épített tezauszok leghatékonyabb, legtöbb célra használható formája a betürendbe rendezett változat /1/.

Ezért fontos feladat annak biztosítása, hogy az érdekeltek a lehetséges legtöbb információhoz juthassanak e tezausz-forma révén. Fő törekvésünk tehát olyan betürendes, egyedi azonosságjelzetekkel rendelkező deszkriptorokat, viszonylag nagyszámú non-deszkriptort és az említettek összes célszerű permutált változatait egyesítő szerkezet kialakítása, amely tükrözi a benne található fogalmak hierarchikus, nem-hierarchikus és permutációs kapcsolatait is.

Ez a forma megfelelő információt adhat nemcsak az indexelő szakember, de az információkereső specialista és a "fogyasztó", a tájékoztatást igénylő számára is. A tezausz tájékoztatni képes a benne megtalálható összes fogalom összes célszerű /deszkriptor, non-deszkriptor és permutált/ változata révén.

A hierarchikus szerkezet

A fogalmak hierarchikus rendezése szükséges és elengedhetetlen feltétel az említett betürendes szerkezet kialakításához. Ez a munkafolyamat teszi lehetővé a fogalmak közötti különböző kapcsolatok megállapítását.

A hierarchikus szerkezet kialakítása sok hazzonnal jár. A rutinszerűen működő tájékoztatási rendszeren belül az indexelő szakember gyakran bukkan olyan fogalomra, amely még nem található meg teauruszában. Az ilyen esetek gyakran vezetnek a teaurusz fogalomállományának bővítéséhez. A bővítés folyamata azt kívánja meg, hogy a felmerült új fogalom a maga sajátos szakmai, ismeretági környezetében kerüljön megítélés alá. Csak ily módon lehetséges eldönteni, hogy az új fogalom alkalmas-e a rendszerbe való beiktatásra. Tekintettel a fenti megállapításokra, a rendszer második célkitűzése a fogalmak hierarchikusan rendezett gyűjteményének kialakítása.

A grafikus szerkezet

A teaurusz fogalomállományának jobb áttekinthetősége, továbbá a könnyű továbbfejleszthetőség biztosítása és az érintkező területek kölcsönös kapcsolatainak kimunkálása érdekében

elengedhetetlen a fogalom-csoportok grafikus ábrázolása irányított gráfok segítségével /2, 3/.

Ez a teauruszépítés munkafolyamatának harmadik célkitűzése.

A fő célkitűzés megvalósítása lehetővé teszi a teaurusz üzemszerű működését - az adott fogalmi határok között - a második és harmadik célkitűzés megvalósítása létrehozta a teaurusz karbantartásának és folyamatos fejlesztésének feltételeit. Ez utóbbi célok elérésének interdiszciplináris és tudományszervezési jelentősége is van.

Az alábbiak során a fő célkitűzés realizálásának egyes lépéseit vesszük sorra, a második és harmadik célkitűzést pedig a fő célkitűzés megvalósításának menetében, annak részeként és előfeltételeiként tárgyaljuk.

A teauruszépítés alapozó lépései

A három főbb lépés a következő:

- a belső szerkezet meghatározása;
- a szerkezeti modell kidolgozása;
- a szerkezeti modell elemzése.

A tezaurusz belső szerkezetének meghatározása

A tezaurusz részint fogalmakat, részint pedig azok közötti kapcsolatokat foglal magába. A tezaurusz fogalmi elemei a következők:

deszkriptorok;
non-deszkriptorok;
permutált fogalom-változatok.

A tezaurusz fogalmainak kapcsolatai a következők /4/:

1. generikus-specifikus kapcsolatok /G-S/

Ez a kapcsolat-típus az átfogóbb, tágabb értelmű, illetve a részletezőbb, szűkebb értelmű deszkriptorok közötti kapcsolatot fejezi ki.

2. affinitív kapcsolatok /R-R/

Ez a típus mindazon deszkriptorok közötti kapcsolatot jelöli, amelyeket nem köt össze generikus-specifikus kapcsolat, de rutinszerű együttes visszakeresésükre, használatukra reálisan számítani lehet.

3. preferenciális kapcsolatok /T-L/

Ez a kapcsolat-típus deszkriptorok és non-deszkriptorok összeköttetését jelöli, megmutatja egy deszkriptor non-deszkriptor-környezetét, illetve megjelöli, hogy egy adott non-deszkriptor melyik deszkriptorhoz tartozik.

4. permutációs kapcsolatok /P/

A kapcsolatoknak ez a típusa is deszkriptor és non-deszkriptor kapcsolatát jelöli.

A generikus-specifikus, affinitív és preferenciális kapcsolat szimmetrikus jellegű.

A fent megfogalmazott fő célkitűzés értelmében, a betűrendes szerkezet tartalmazza a deszkriptorok és non-deszkriptorok összes célszerűnek ítélt permutált változatát is. Hatékony megoldásnak látszik a non-deszkriptorok permutált formáit is a deszkriptorokhoz kapcsolni /és nem a vonatkozó non-deszkriptorokhoz, hiszen ez utóbbiak ugyis a deszkriptorokhoz vezetnek vissza/. Ez a megoldás teszi igazán értékesé a betűrendes szerkezet permutált fogalomváltozatainak gazdagságát. A permutált formák és a deszkriptorok közötti kapcsolat értelmezése új kapcsolat-típus bevezetését eredményezi, amelynek specifikus jellegét a non-deszkriptorok permutált formáinak a deszkriptorokhoz való kapcsolása emeli ki. Ez az új típus - funkcionális szempontból

nézve - a preferenciális kapcsolat egy specifikus változata /jele: P/. A rendszer a permutált fogalomváltozatokat deszkriptorok kvázi-szinonimáiként kezeli. A permutációs kapcsolat kialakítása - a folyamatos használat követelményeinek szellemében - aszimmetrikusan történik: a permutált fogalomváltozat a deszkriptorra utal és nincs reciproka ennek az utalásnak. A kapcsolatok e sajátos típusára alább még visszatérünk.

A fogalmak jelzeteleése

A tezaurusz fogalmi vázát a deszkriptorok rendezett halmaza teremti meg. Minden egyes deszkriptornak egyedi, azonossági jelzete van. A többi fogalmi elem nem rendelkezik egyedi jelzettel, hanem a deszkriptorok jelzeteinek segítségével kötődik azokhoz. Az egyedi azonossági jelzet minden egyes deszkriptor verbális megnevezésének mindig jelenlévő kiegészítő eleme, de a jelzetek önmagukban is hatékonyan alkalmazhatók, elsősorban számítógépi tárolás esetén, mivel minden szöveges értelmezés nélkül is egyértelműen utalnak az adott deszkriptorra. A célszerűen megválasztott egyedi deszkriptor-jelzet közvetlen információt adhat az adott fogalom hierarchikus szintjére vonatkozólag, sőt: a hierarchikusan felépített jelzetrendszer a tezaurusz fogalmainak teljes vertikális szerkezetét is hordozhatja.

Leghatékonyabbnak az alfanumerikus jelzetrendszer használata látszik. A jelzetrendszer kapacitásának megállapításakor a várható maximális deszkriptor-mennyiséget kell meghatározni. Tapasztalataink és vizsgálataink alapján úgy véljük, hogy bármelyik ismeretág fogalmi anyaga feltárható egy kb. 10 000 - 15 000 deszkriptort tartalmazó tezauruszal. A jelzetrendszer rugalmasságát az ugrószámozás módszerének alkalmazásával biztosíthatjuk, s ezért tiszteres kapacitást veszünk tervbe. Jelzetrendszerünk maximális kapacitását 200 000-ben állapítjuk meg, ez mindenképpen elégségesnek látszik. Minden egyes jelzet egy verzális betű és négy számjegy kapcsolatából áll /pl. D 5812/. A negyedik számjegyet tartalékként kezeljük, egyelőre csak hármat használunk fel.

Az ábécé 6 nagybetűjét - G, L, P, R, S, T - már lefoglaltuk a különböző típusú kapcsolatok jelzésére. A még fennmaradó 20 betű meghatározza a deszkriptorok lehetséges főosztályainak számát, a három számjegyes jelzetrészek pedig minden egyes főosztályon belül maximum 999 jelzet kiosztását teszik lehetővé. Ily módon lehetővé válik $20 \times 1000 = 20\ 000$ deszkriptor ellátása egyedi alfanumerikus azonossági jelzettel. Ez a kapacitás elégségesnek látszik egy készülő tezaurusz fogalmainak jelzeteleéséhez, sőt a tezaurusz első néhány évi működéséhez is.

A deszkriptorok kódolása

A tezaurusz dinamikus, nyílt rendszer, és ezt az alapelvet kell tükröznie a jelzetrendszernek is. Az ugrószámozás módszere eleget tesz e követelménynek. A jelzeteleés előkészítéseként, a deszkriptorokat csoportonként hierarchikusan kell rendezni. Az egy-

egy deszkriptor-osztályban kiosztható maximális jelzet-szám /999/ és az adott osztályban jelzetelésre váró deszkriptorok számának hányadosa meghatározza az ugrószámozás számközeit:

$$u = \frac{j}{d}$$

ahol u: az ugrószámköz nagysága; j: a kiosztható maximális jelzetek száma; d: az adott csoport deszkriptorainak száma. A kapott érték természetesen csak alap-érték és mindenkor módosítani kell az adott tudományág természete, valamint az adott deszkriptor-osztály kimunkáltsági foka szerint. A módosítás egyes esetekben éppúgy jelenthet számközsűkítést, mint -bővitést.

A jelzetek kiosztásakor hasznosnak látszik megkülönböztetett jelentőséget tulajdonítani a zéró számjegynek, oly módon, hogy a több zérót tartalmazó jelzeteket /pl. 008, 100, 050/ mindig magasabb hierarchia-szintű /generikus/ fogalmak, az egy zérót tartalmazó jelzeteket közepes, a zérót nem tartalmazó jelzeteket pedig alacsony hierarchia-szintű /specifikus/ fogalmak kapják.

A zérótól eltérő, más numerikus értéket is lehet rendezőelvül választani, de ez az elv megfelelőnek látszik a szükséges proporcionálás megvalósításához. Az adott háromjegyű számban található zérók száma és helyzete hasznos információt ad az egyes deszkriptorok hierarchikus szintjére vonatkozólag. Pl.

600 munkabér

610 időbér

618 hetibér

A negyedik számjegy esetleges felhasználása természetesen megnöveli a kitüntetett számok mennyiségét is.

A non-deszkriptorok kódolása

A non-deszkriptoroknak nincs egyedi azonossági jelzetük, mindig arra a deszkriptorra utalnak, amelyhez tartoznak. Ezért a non-deszkriptorok utalása mindig a deszkriptorok verbális megnevezéséhez, illetve egyedi alfanumerikus kódjához

mutat vissza. A deszkriptorok és non-deszkriptorok permutált változatai sem rendelkeznek egyedi azonossági jelzettel; az utalás mindig a vonatkozó deszkriptorra irányul.

A tezaurusz írásmódja

A tezaurusz jobb informativitásának biztosítása érdekében, a könnyebb kezelhetőség céljából tanácsos az eltérő funkcióju fogalomtipusokat tipográfiailag is jelölni. Olyan szedésre

van szükség, amely egyértelműen és feltűnően megmutatja, hogy egy adott fogalom deszkriptor, non-deszkriptor vagy permutált fogalomváltozat.

<p>A deskriptorok írás módja</p>

A deskriptort /és csak a deskriptort!/ mindig alfanumerikus azonossági jelzet vezet be, amely a hasáb balszélétől indul. Az egyedi azonossági jelzetet szóköz követi, majd a deskriptor verbális megnevezése következik. Minden non-deszkriptor jellegű tétel a deskriptor verbális megnevezésével egy vonalban kezdődik, ily módon a deskriptor azonossági jelzete "kiugratja" a szedéstükörből a deskriptorokat.

A deskriptor jelzetét és megnevezését tartalmazó sort a különféle kapcsolatok felsorolása követi. Minden egyes kapcsolódó fogalom új sorban, a deskriptor verbális megnevezése alatt, ötbetűs bekezdéssel indul. A kapcsolat jellegére utaló rövidítés vezet be az egyes kapcsolódó fogalmakat, ezt a verbális megnevezés, majd a vonatkozó azonossági jelzet követi. A különböző kapcsolat-típusok felsorolási rendje kötelező:

1. T /tiltott fogalom/
2. G /generikus fogalom/
3. S /specifikus fogalom/
4. R /vonatkozó fogalom/

A T rövidítés /tiltott fogalom/ az egyetlen non-deszkriptorra utaló kapcsolat jelölése, ezért a T után következő verbális non-deszkriptor-megnevezés után nem áll azonossági jelzet, hiszen a non-deszkriptor nem kap ilyen jelzetet.

Az elmondottakat az alábbi példa szemlélteti:

C 650 bérpótlék

T munkabérpótlék
G munkabér C 600
S családi pótlék C 652
S korpótlék C 654
S nyelvpótlék C 656
R tulóradij C 627

A példából látható, hogy a szedésben három bekezdés egyértelmű megkülönböztetését kell elérni:

1. azonossági jelzet;
2. verbális megnevezés;
3. kapcsolatok és kapcsolódó fogalmak felsorolása.

Az 1. bekezdés csak deskriptorok számára szolgál, mivel csak azoknak van azonossági jelzetük.

A non-deszkriptorok
írasmódja

A non-deszkriptorokat nem vezeti be egyedi azonossági jelzet, az ilyen jellegű tételek írása a deszkriptorok verbális megnevezése alatt, azzal egy vonalban kezdődik. A non-deszkriptor megnevezése alatti sor 5. betűhelyén kezdődik

az utalás, amely mindig egy deszkriptorra irányul. Az utalást L /lásd/ rövidítés vezeti be, ezt követi a vonatkozó deszkriptor verbális megnevezése, majd a deszkriptor jelzete. Pl.:

munkabérpótlék

L bérpótlék C 650

Az L betű jelenléte egyértelműen mutatja a tétel non-deszkriptor-jellegét, illetve a kapcsolat preferenciális mivoltát. A non-deszkriptor jellegét egyébként a tételt bevezető azonossági jelzet hiánya is világosan jelzi.

A permutált változatok
írasmódja

A fogalmak /deszkriptorok és non-deszkriptorok/ permutált változatait sem vezeti be egyedi azonossági jelzet. A permutálás műveletének eredményeképpen, egy adott fogalomnak eltérő vezérszavakkal kezdődő formáit

kell besorolni a betűrendbe. Az egyes permutált változatok vezérszavát mindig vessző követi, aztán következik a fogalom eredeti formájának közlése, amelynek végéről hiányzik azonban a vezérszóra kiemelt tag. A kapcsolat megjelölése a következő sorban, annak 5. betűhelyén indul, és mindig P /permutált változat/ rövidítés vezeti be. A rövidítést a vonatkozó deszkriptor verbális megnevezése, majd egyedi azonossági jelzete követi. Pl.:

pótlék, bér=

P bérpótlék C 650

pótlék, kor=

P korpótlék C 654

pótlék, munkabér=

P bérpótlék C 650

pótlék, nyelv=

P nyelvpótlék C 656

Szükséges és fontos a sajátos tördelés, mivel ez a tagolás közvetlen kiegészítő információt nyújt a tezaurusban keresőknek, és általában véve - kitűnő vizualitású tezaurusszerkezetet eredményez.

A teaurusz modelljének felépítése

A fentiekben kifejtett elveket és javasolt módszereket gyakorlatban is megvizsgáltuk néhány teauruszmodell felépítése és a modellek elemzése során. Az alábbiakban egy modell felépítésének menetét mutatjuk be,

majd megvizsgáljuk a bemutatott modell néhány alapvető paraméterét. A modell egy kisméretű deskriptor-osztályt ölel fel; fogalmainak kódolt betűrendes jegyzéke az 1. táblázaton, grafikus ábrázolása pedig az 1. ábrán szemléltethető.

1. táblázat

A MUNKABÉR fogalom-osztály deskriptorainak
kódolt betűrendes jegyzéke

C 635	akkordbér	C 628	különmunkadij
C 640	alapbér	C 605	levonás
C 602	átlagbér	C 600	munkabér
C 650	bérpótlék	C 616	napibér
C 652	családi pótlék	C 656	nyelvpótlék
C 632	darabbér	C 675	nyereségrészesedés
C 655	éjszakai pótlék	C 612	órabér
C 643	előleg /részfizetés/	C 660	prémium
C 687	étkezési hozzájárulás	C 657	rakodási pótlék
C 689	fizetési előleg	C 696	rendkívüli segély
C 693	gyermekgondozási segély	C 670	részesedés
C 621	havibér	C 690	segély
C 618	hetibér	C 638	szakmánybér
C 653	idénypótlék	C 624	személyi fizetés
C 610	időbér	C 692	szülési segély
C 663	jutalom	C 630	teljesítménybér
C 614	kiemelt órabér	C 626	tulmunkadij
C 681	kiküldetési díj	C 627	tulóradij
C 654	korpótlék	C 684	utazási hozzájárulás
C 658	különélési pótlék		

A modell összeállításának főbb lépései a következők voltak:

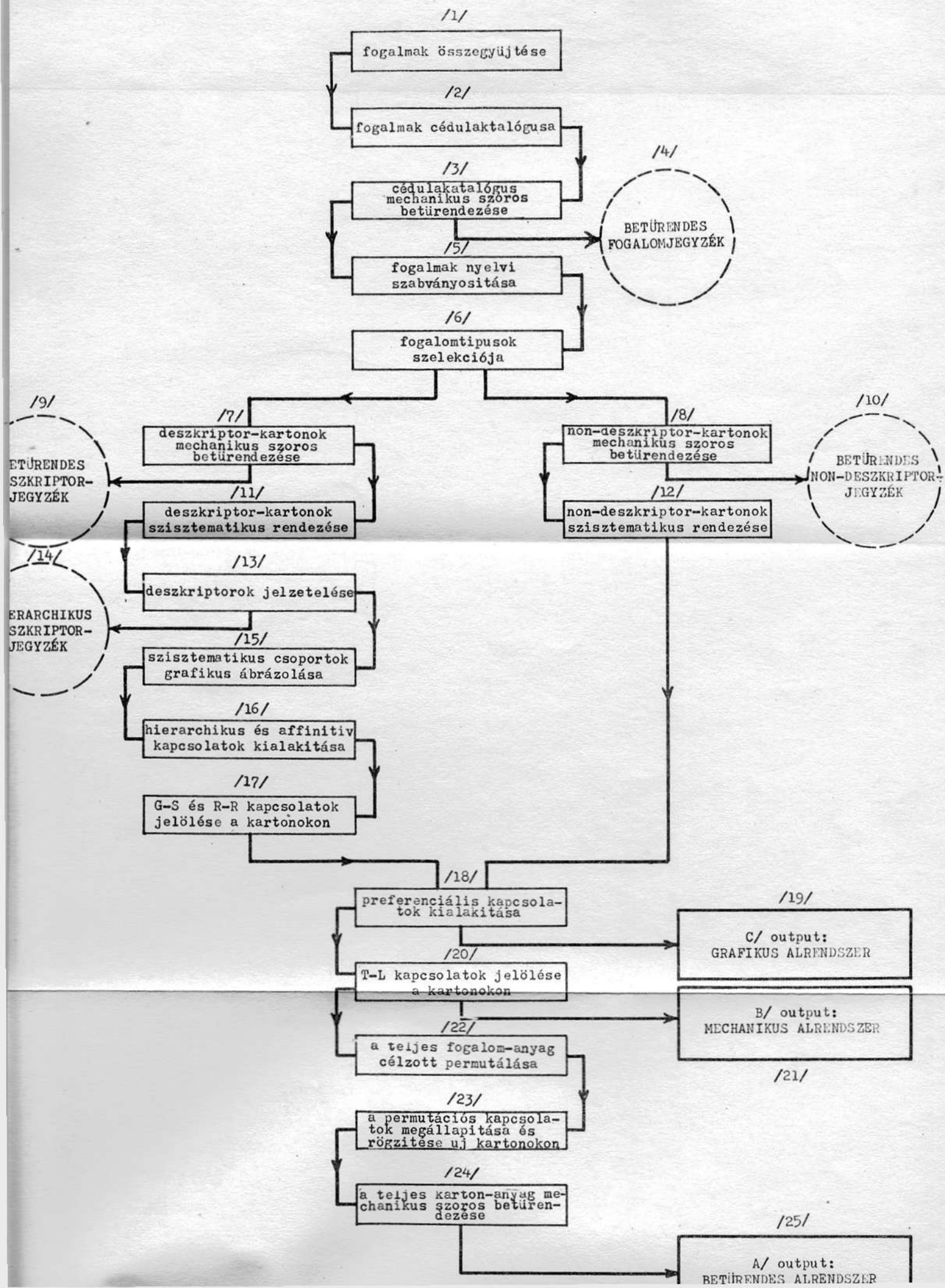
- a tárgyszóanyag összegyűjtése;
- cédulakatalógus készítése a tárgyszavakból;
- a tárgyszavak grammatikai szabványosítása; az írásmód egységesítése;
- a deszkriptorok és non-deszkriptorok különválasztása;
- a deszkriptorok szisztematikus csoportosítása;
- a deszkriptorok alfanumerikus jelzetelése; a deszkriptorok kiírása növekvő azonossági jelzetek sorrendjében /2. táblázat/;
- a deszkriptorok grafikus ábrázolása /1. ábra/; a hierarchikus /G-S/ és affinitív /R-R/ kapcsolatok megállapítása, rávezetése a térképre, illetve a cédulákra;
- a non-deszkriptorok /szinonimák és kvázi-szinonimák/ csatolása a deszkriptoranyaghoz; alfanumerikus utalásokat tartalmazó új katalóguslapok kiállítása a non-deszkriptorok számára; az új és régi katalóguslapok egybeosztása; a preferenciális /T-L/ kapcsolatok megállapítása; e kapcsolatok feltüntetése a térképen, illetve a cédulákon;
- a grafikus térkép végleges megrajzolása;
- a szisztematikus tezaurusz-modell kiírása;
- a modell teljes fogalomállandóknak célzott permutálása; új katalóguslapok kiállítása a permutált fogalomváltozatok számára; a permutációs kapcsolat /P/ megállapítása és rávezetése az új katalóguslapokra; az új katalóguslapok beosztása a cédulakatalógusba;
- a teljes cédulakatalógus mechanikus szoros betűrendezése;
- a betűrendes tezaurusz-modell kiírása.

A modell felépítésének áttekintő algoritmusát a 2. ábra szemlélteti.

A tezaurusz-modell
elemzése

A modell elemzését két irányban végeztük el:

- a fogalmak elemzése;
- a kapcsolatok elemzése.



A MUNKABÉR fogalom-osztály deszkriptorainak
kódolt szisztematikus jegyzéke

C 600 munkabér

C 602 átlagbér

C 605 levonás

C 610 időbér

C 612 órabér

C 614 kiemelt órabér

C 616 napibér

C 618 hetibér

C 621 havibér

C 624 személyi fizetés

C 626 túlmunkadíj

C 627 túlóradíj

C 628 különmunkadíj

C 630 teljesítménybér

C 632 darabbér

C 635 akkordbér

C 638 szakmánybér

C 640 alapbér

C 643 előleg /részfizetés/

C 650 bérpótlék

C 652 családi pótlék

C 653 idénypótlék

C 654 korpótlék

C 655 éjszakai pótlék

C 656 nyelvpótlék

C 657 rakodási pótlék

C 658 különélési pótlék

C 660 prémium

C 663 jutalom

C 670 részesedés

C 675 nyereségrészesedés

C 681 kiküldetési díj

C 684 utazási hozzájárulás

C 687 étkezési hozzájárulás

C 689 fizetési előleg

C 690 segély

C 692 szülési segély

C 693 gyermekgondozási segély

C 696 rendkívüli segély

A fogalmak elemzése

Jelölések:

a modell deskriptorai: d

a modell non-deszkriptorai: n

a modell fogalmai: f ; $f = d + n$ a modell permutált deskriptor-változatai: p_d a modell permutált non-deszkriptor változatai: p_n a modell permutált fogalom-változatai: p ; $p = p_d + p_n$ a modell összes tétele : t ; $t = f + p$

Az egyes tétel-csoportok mennyiségi jellemző adatait a 3. táblázat mutatja.

3. táblázatA modell számszerű adatai

tétel	egység	hányad	tétel	egység	hányad
d	39	0,43	f	48	0,53
n	9	0,10			
p_d	37	0,41	p	42	0,47
p_n	5	0,06			
t	90	1,00	t	90	1,00

A 3. táblázat adatai lehetővé teszik a modell néhány jellegzetességének megállapítását.

$$\frac{p}{f} = \frac{42}{48} = 0,88$$

Ez az érték megmutatja a modell egy deskriptorára jutó permutált változatok mennyiségét, tehát közvetlenül jelzi a modell fogalomszervezési szintjét. Ez érthetővé válik, ha tekintetbe vesszük,

hogy egy permutált változatokat nem tartalmazó halmaz $\frac{p}{f}$ értéke = 0. Egy csak uniterm jellegű fogalmakat tartalmazó rendszer fogalomszervezési szintjének mérőszáma ily módon: 0.

Csak egyszeresen összetett szavakat tartalmazó rendszer esetében a permutált fogalomváltozatok száma megegyezik a fogalmak számával, ily módon

$$p = f \quad \text{és ebből:} \quad \frac{p}{f} = 1$$

Ez az érték jellemzi az unikoncept jellegű fogalmakat tartalmazó rendszer fogalomszervezési szintjét.

A csupa háromtagu fogalommal dolgozó rendszer jellemzője a következő:

$$p = 2f, \quad \text{illetve} \quad \frac{p}{f} = 2$$

A fentiekből következik, hogy a modell fogalomszervezési szintjének 0,88 értéke egyértelműen unikoncept jellegű fogalomszervezésre utal. Természetes és köztudott, hogy egy adott teaurusz fogalmai mindig tartalmaznak szórást a különböző típusu fogalomszervezési szintek között; a jelzőszám mindig az átlagot mutatja, ily módon joggal nevezhető a teauruszok egyik alapvető paraméterének.

A cranfieldi vizsgálatok /5/ az unikoncept-típusu fogalmakat tartalmazó rendszereket átlagos, vagy annál valamivel jobb hatékonyságúaknak minősítik, így a modell fogalomszervezési szintjét megfelelőnek értékelhetjük.

A modell fogalmainak fogalomszervezési szintek szerinti megoszlását a 4. táblázat mutatja be.

4. táblázat

A modell fogalom-típusainak megoszlása

tipus	egység	hányad	példa
uniterm	11	0,23	bér
unikoncept	32	0,67	órabér
polikoncept	5	0,10	kiemelt órabér
f	48	1,00	

Az n/d hányados a deszkriptorok és non-deszkriptorok egymáshoz mért mennyiségi viszonyát fejezi ki, s a fogalomgyűjtemény szemantikai kimunkáltságára utal. Ismeretes, hogy az egyes ismeretágak eltérő árnyaltságu non-deszkriptor környezetel veszik körül a deszkriptorokat. Mégis, a már létező és alkalmazott tezauruszok példája és saját tapasztalatunk is azt mutatja, hogy az n/d hányados értéke általában nem kisebb mint 0,5.

Modellünk redundancia-faktora:

$$r = \frac{n}{d} = 0,23$$

A fenti megállapítás értelmében: a modell non-deszkriptor környezete a szükségesnél szegényebb. Ez az érték még akkor is elégtelen, ha tekintetbe vesszük, hogy ezt az értéket nagymértékben befolyásolja a fogalombontás legalsó szintjének meghatározása, ami mindenkor a reális szükségletektől elinduló, de végeredményben - adott tezaurusz esetén - önkényes eljárás.

A kapcsolatok elemzése

A modell részint kétirányú /szimmetrikus/, részint pedig egyirányú /aszimmetrikus/ kapcsolatokat tartalmaz.

Szimmetrikus kapcsolatok: r_2

hierarchikus /G-S/ kapcsolat: r_h

affinitív /R-R/ kapcsolat: r_a

preferenciális /T-L/ kapcsolat: r_p

$$r_2 = r_h + r_a + r_p$$

Aszimmetrikus kapcsolat: r_1

permutációs /P/ kapcsolat.

A modellben fellelhető kapcsolatok összessége: r

$$r = r_2 + r_1$$

A kapcsolatok mennyiségi megoszlását a modellben az 5. táblázat mutatja.

5. táblázat

A kapcsolatok megoszlása a modellben

tipus	egység	hányad	tipus	egység	hányad
r_h	76	0,50			
r_a	16	0,11	r_2	110	0,73
r_p	18	0,12			
r_l	42	0,27	r_1	42	0,27
r	152	1,00	r	152	1,00

Ezekkel az adatokkal meg lehet határozni a modell egy adott deskriptorának átlagos kapcsolat-mérszámát $k/$:

$$k = \frac{r}{d} = \frac{152}{39} = 3,9$$

Ez az érték egy deskriptor összes kapcsolatainak átlagos mennyiségét adja meg. E mennyiség egyes kapcsolat-típusonkénti megoszlását a 6. táblázat szemlélteti.

6. táblázat

Egy deskriptor kapcsolatainak megoszlása

tipus	egység	tipus	egység
$\frac{r_h}{d}$	1,95		
$\frac{r_a}{d}$	0,41	$\frac{r_2}{d}$	2,82
$\frac{r_p}{d}$	0,46		
$\frac{r_l}{d}$	1,08	$\frac{r_1}{d}$	1,08
$\frac{r}{d}$	3,90	$\frac{r}{d}$	3,90

A kapcsolat irányát /a szimmetriát, illetve aszimmetriát/ is tekintetbe véve, a mérőszámok a következőképpen alakulnak /7. táblázat/:

7. táblázat

A kapcsolatok megoszlása irány szerint

tipus	d	d	összesen
r_h	0,975	0,975	1,95
r_a	0,205	0,205	0,41
r_p	0,230	0,230	0,46
r_l	1,080	-	1,08
r	2,490	1,410	3,90

Látható, hogy a deskriptorok felé irányított kapcsolatok mennyisége lényegesen meghaladja a deskriptorokból induló utalások mennyiségét, s az is, hogy ezt a permutációs kapcsolat aszimmetrikus jellege magyarázza.

A fogalmaknak és azok között feltüntetett kapcsolatoknak megvizsgálása után megfigyeltük a modell vertikális szerkezetét, a fogalombontás részletességét is. A modellben található deskriptorok vertikális tagozódásának számszerű adatait a 8. táblázat szemlélteti.

8. táblázat

A modell vertikális szerkezetének tagozódása

hierarchia-szint	egység	%
1.	1	2,5
2.	15	38,5
3.	22	56,4
4.	1	2,5
1-4.	39	99,9

Az a tény, hogy a deskriptoroknak több mint fele /56,4%/ a harmadik hierarchia-szintet képviseli, eldönti a modell vertikális tagozódásának jellegét: a modell 3-hierarchia-szintes.

Következtetések

A fentiek egy kisméretű modell elemzésének első eredményei. Az adatok néhány megfontolást ajánlatossá tesznek:

a deskriptorokhoz kapcsolódó non-deszkriptorok számát meg kell növelni;

szaporítani kell a feltüntetett affinitív kapcsolatok számát. Tanácsosnak látszik több deskriptor-osztály modelljének összehasonlító vizsgálata, hiszen a jelenlegi modellből teljességgel hiányoznak a deskriptor-főosztályok közötti és kismértékben vannak feltüntetve az osztályközi affinitív kapcsolatok;

a fogalmak fogalomszervezési szintje megfelelő;

a modell vertikális tagozódása a több szakember által is kívánatosnak ítélt gyenge hierarchiát tükrözi /6/.

A bemutatott modellelemzés egyszersmind elemzés-modell is kíván lenni. További módszerekre, az elemző tevékenység elmélyültebb és kiterjedtebb végzésére van szükség a teauruszszerkezet alapvető jellegzetességeinek pontosabb megállapításához. Feltétlenül szükségesnek tartjuk néhány jelentősebb teaurusz hasonló paramétereinek megvizsgálását, mert úgy véljük, hogy csak összehasonlító vizsgálatok vezethetnek el általánosítható megállapításokhoz.

A fenti - és több más jellegű és méretű - modell mélyebb vizsgálata folyik.



I R O D A L O M

- /1/ MOLNÁR, I.: Remarks on the general principles of thesauri building. = Proceedings of the international conference on general principles of thesauri building. Warsaw, 23-27 March, 1970. p.45-50.
- /2/ ROLLING, L.: The role of graphic display of content relationships in indexing and retrieval vocabularies. Elsinore, 1964. 29 p. /FID/CR International Study Conference on Classification Research./
- /3/ KUNSZT Gy.: Tudománylogikai modellelmélet. Bp. ÉTI, 1970. 89 p.
- /4/ UNESCO guidelines for the establishment and development of monolingual scientific and technical thesauri for information retrieval. Paris, UNESCO, 1970. 18 p.
- /5/ CLEVERDON, C. - MILLS, J. - KEEN, M.: Cranfield research project. Factors determining the performance of indexing systems. Vol.1-2. Cranfield, National Science Foundation, 1966. 2 db.
- /6/ TOMAN, J.: Problems of thesauri. = Proceedings of the international conference on general principles of thesauri building. Warsaw, 23-27 March, 1970. p.99-104.

mm^mmm

MOLNAR, I.: Designing a new-type thesaurus-structure

The paper aims at proposing a method to construct a structure combining the alphabetically arranged descriptors, supplied with individual identification marks, with a relatively large number of non-descriptors and all feasible permuted variations of the former so that the structure may adequately mirror the hierarchical /generic - specific/, non-hierarchical /affinitive - preferential/, as well as permutative relations of all concepts /terms/ covered.

The author points out that this structure may be properly built up only after the hierarchical arrangement of the terms /hierarchical sub-system/ by a graphical representation of the term-groups /graphical sub-system/.

The author looks upon the whole of the thesaurus as a system, and considers both the alphabetical main form and the hierarchical resp. the graphic form necessary to compile a thesaurus suitable for serving several information purposes.

The author also tackles the practical problems of constructing the thesaurus structure with a special view to the notations of terms, which may be handled by computer, too.

He also establishes a new relation-type, which he terms as permutative relation, and describes the characteristic features and role of this relation, as well as its merging with the main - alphabetical - part. The relation-type is deemed significant by the author both as regards the practical use of the thesaurus and its analysis.

All this is illustrated by means of an elaborated model containing a minor group of terms taken from labour management /it only contains 39 descriptors/. The author also shows the algorithm of thesaurus construction and an elaborate graphic map of the model.

By analyzing the constructed model, he tends to determine certain essential thesaurus parameters /e.g. redundancy factor, level of term organization/.

Using larger models, the author continues his investigations expecting to reach generalizable results.

!!!

МОЛНАР, И.: О создании новой структуры тезауруса

Автор статьи ставит себе целью показать такую структуру тезауруса, которая объединяет в себе дескрипторы, имеющие однозначные шифры, расположенные в алфавитном порядке, сравнительно большое число не - дескрипторов и все перемутированные варианты их, считаемые полезными, и отражает также иерархические (генерические - специфические), не - иерархические (аффинитивные и преференциальные) и пермутационные отношения между понятиями тезауруса.

Показывается, что только создав иерархическое собрание понятий (иерархическую подструктуру) и графически изобразив группы понятий можно успешно создать эту структуру.

Автор считает тезаурус системой, и как алфавитную главную форму, так иерархическую или графическую форму необходимыми для создания эффективного тезауруса, соответствующего многообразным целям информации.

В дальнейшем рассматриваются практические вопросы создания структуры тезауруса, и обращает особое внимание на зашифрование понятий при помощи ЭВМ.

Выявляется новый тип отношения, названный автором пермутационным отношением. Представляются признаки, роль этого отношения и его место в алфавитной главной части. Автор считает новый тип отношения значительным как в области использования тезауруса, так и в области его анализа.

Вышеупомянутые идеи представляются и на примере модели обработки небольшой группы понятий (содержащей 39 дескрипторов) по теме труда. Показывается алгоритм создания тезауруса и графическая карта выявленной модели.

После анализа модели автор стремится к установлению некоторых основных параметров тезауруса (фактор избыточности, уровень организации понятий).

Исследования продолжаются на моделях большего размера, автор делает дальнейшие анализы и ожидает обобщаемых результатов.

" "

MOLNÁR, I.: Aufstellung einer neuartigen Thesaurus-Struktur

Das Ziel der Abhandlung ist die Darstellung einer solchen Struktur, die Deskriptoren mit individuellen Identitätszeichen und in alphabetischer Ordnung, ferner zahlreiche Nondesriptoren, sowie sämtliche zweckmässigen permutierten Varianten derselben in sich vereint und auch die hierarchischen /genetisch-spezifischen/, nicht-hierarchischen /affinitiven und präferenziellen/ sowie Permutationsrelationen der darin enthaltenen Begriffe widerspiegelt.

Es wird gezeigt, dass diese Struktur nur nach Ausbildung der hierarchisch geordneten Begriffssammlung /hierarchisches Subsystem/ und mittels einer graphischen Darstellung /graphisches Subsystem/ erfolgreich aufgebaut werden kann.

Der Autor betrachtet den Thesaurus als ein System und hält sowohl die alphabetische Hauptform, als auch die hierarchische, bzw. graphische Form in der Bearbeitung eines wirksamen, für zahlreiche Informationszwecke anwendbaren Thesaurus für unerlässlich.

Der Artikel bespricht im weiteren die praktischen Probleme der Aufstellung der Thesaurus-Struktur, mit besonderer Rücksicht auf die auch durch EDV behandelbare Bezeichnung der Begriffe.

Es wird ein neuer Relationstyp aufgestellt und Permutationsrelation genannt. Er stellt die Kennzeichen, Rolle und den Einbau der Relation in die Struktur des alphabetischen Hauptteils dar. Der Autor hält den neuen Relationstyp sowohl auf dem Gebiet des praktischen Gebrauchs, als auch für die Analyse des Thesaurus für höchst bedeutend.

Die Kenntnisse und Feststellungen werden anhand eines Modells durch die Demonstration der Bearbeitung einer kleinen - 39 Deskriptoren zählenden - Begriffsgruppe auf dem Gebiet des Arbeitswesens illustriert. Der Artikel beschreibt den Algorithmus des Thesaurusaufbaus und die graphische Karte des Modells.

Mittels der Analyse des aufgebauten Modells ist der Autor bestrebt einige grundlegende Thesaurus-Parameter festzustellen /Redundanzfaktor, Begriffsorganisationsniveau/.

Die Untersuchungen werden an grösseren Modellen weitergeführt, der Autor nimmt weitere Analysen vor und erwartet von diesen verallgemeinbare Ergebnisse.

::':::

*Hogyan lehet folyamatosan, gyorsan tájékozódni
az ipargazdaság és üzemszervezés korszerű kérdéseiről?*

*A rohamos fejlődéssel csak megfelelő információk birtokában
lehet lépést tartani.*

Est a célt szolgálja a

GYORSINDEX-IPARGAZDASÁG

c. információs kiadvány, amely havonta mintegy 60 válogatott külföldi ipargazdasági és üzemszervezési folyóirat cikkeinek bővített címfordítását (tartalomjegyzékét) és a címekben előforduló tárgyszavak permutált indexét közli.

*Ez a jól áttekinthető gyorstájékoztató kiadvány
jelentős időmegtakarítást
eredményez a szakirodalom tanulmányozásánál, mert módot nyújt
a publikált cikkek, tanulmányok beszerzését megelőző
válogatásra.*

A „Gyorsindex-Ipargazdaság” 12 számának évi előfizetési díja: 600,— Ft.

Megrendelhető az

ORSZÁGOS MŰSZAKI KÖNYVTÁR
ES DOKUMENTÁCIÓS KÖZPONT
Kereskedelmi Igazgatóságánál
Budapest, VIII., Reviczky u. 6.