

## MAGYAR NYELVŰ SZÁMÍTÁSTECHNIKAI TEZAUROSZ II. A TEZAUROSZ ÖSSZEHAISONLÍTÓ STATIKAI ELEMZÉSE

Molnár Imre

Építésügyi Tájékoztatási Központ

Az alább következő elemzés megjelenésekor a Számítástechnikai Fogalmak Tezaurusza /1/ már megjelent és eljutott az érdeklődő szakemberekhez. A Tudományos és Műszaki Tájékoztatás 1973. decemberi számában részletes ismertetés jelent meg a tezaurusz elkészítésének indítékairól, módszereiről és gyakorlati lefolyásáról /2/. E cikk célja a kész tezaurusz megvizsgálása abból a szempontból, hogy milyen statikai paraméterek jellemzik, s kísérletet teszünk a megállapított paraméterek értékelésére is. Elsősorban annak elemzésére törekszünk, hogy miben és mennyiben van indok a tezaurusz továbbfejlesztésére, illetve mely paraméterek tekintetében látszik megfelelőnek - már a tezaurusz folyamatos felhasználásának kezdete előtt.

Az elemzés során kapott adatokat, jellemzőket összevetjük néhány - korábbi elemzéseink eredményeként rendelkezésünkre álló - egyéb hazai tezaurusz-modell, illetve makrotezaurusz megfelelő adataival /3, 4, 5/, hogy a számértékek plasztikusabb értékelését, közvetlenebb megértését biztosíthassuk. A hasonlítás céljából felhasznált adatok természetesen nem tekinthetők olyan elfogadott és kétségtelenül helyes mennyiségeknek, amelyekhez mérve jogosan minősíteni lehet egy tezauruszt. Az összevetés megegyezői nemcsak a Számítástechnikai Fogalmak Tezauruszát, hanem az idézett modelleket is minősítik. Az a nézetünk azonban, hogy több, egymástól többé-kevésbé független, eltérő körülmények között, eltérő természetű fogalomkörök megmunkálásáról kapott adatok feltűnően megegyező sajátosságai mindenképpen figyelmet érdemelnek, s hisszük, hogy az összehasonlítások számának szaporítása révén egyre jobban általánosítható következtetések vonhatók le a vizsgált és összevetett fogalomcsoportokról. Az elemzések tapasztalatai a magyar nyelv és a magyar szakterminológia szemantikai mélységeihez, sajátosságaihoz is közelebb vezethetnek.

Hasonlóképpen: a referencia-modellek adataitól való szignifikáns eltérés is éppúgy lehet érv a vizsgált tezaurusz mellett, mint az ellen. A pusztán mennyiségi adatok - önmagukban véve - még csak a sajátosságokat mutatják meg, azok értékeléséhez még további, a tezaurusz-építés gyakorlati tapasztalataiból származó megfigyelések és megfontolások szükségesek.

A tezaurusz áttekintő jellegű statikai elemzését 3 területen végezzük el:

1. A tezaurusz fogalom-készletének vizsgálata
2. A fogalmak reláció-hálózatának vizsgálata
3. A szerkezet mélységi tagozódásának vizsgálata

#### 1. A TEZAURUSZ FOGALOM-KÉSZLETÉNEK VIZSGÁLATA

Az elemzés lefolytatásához néhány jelölésre van szükségünk. Ezek az alábbiak:

- a tezaurusz deszkriptorainak száma:  $d$
- a tezaurusz non-deszkriptorainak száma:  $n$
- a tezaurusz összes fogalmainak száma:  $f$

$$f = d + n$$

- a tezaurusz deszkriptorai permutált változatainak száma:  $p_d$
- a tezaurusz non-deszkriptorai permutált változatainak száma:  $p_n$
- a tezaurusz összes permutált fogalom-változatainak száma:  $p$

$$p = p_d + p_n$$

- a tezaurusz összes tételeinek száma:  $t$

$$t = f + p = d + n + p_d + p_n$$

A tezaurusz elemeinek funkcionális megoszlását az 1.sz. táblázat mutatja be.

A táblázat adatai értékelésének elősegítéséül a 2-4.sz. táblázatokban bemutatjuk a fentebb említett hazai kezdeményezések néhány hasonló adatsorát.

Az összefoglaló fogalmi arányok  $f$ ,  $p$  érdekes megegyezéseket mutatnak a számítástechnikai tezaurusz és az összehasonlításhoz bemutatott tezaurusz-modellek között.

A tezaurusz fogalmi kimunkáltságának egyik legfontosabb mutatója a non-deszkriptorok /valóságos és "mesterséges" szinonimák és kváziszinonimák/ és deszkriptorok egymáshoz mért számaránya. Ez a hányados a tezaurusz fogalmi anyagának redundancia-hátterét /R/ adja meg.

1.sz. táblázat

## A TEZAUROSZ FOGALMAINAK FUNKCIONÁLIS MEGOSZLÁSA

tétel	egység	hányad	
d	2347	0,51	
n	256	0,06	
$f = /d+n/$	2603	0,57	
$P_d$	1756	0,38	
$P_n$	233	0,05	
$p = /P_d+P_n/$	1989	0,43	
t	4592	1,00	

2.sz. táblázat

## A KÖZET TEZAUROSZ-MODELL FOGALMAINAK FUNKCIONÁLIS MEGOSZLÁSA

tétel	egység	hányad	
d	210	0,45	
n	80	0,17	
$f=/d+n/$	290	0,62	
$P_d$	108	0,23	
$P_n$	70	0,15	
$p=/P_d+P_n/$	178	0,38	
t	468	1,00	

3.sz. táblázat

A MUNKABÉR TEZAUROSZ-MODELL FOGALMAINAK  
FUNKCIONÁLIS MEGOSZLÁSA

tétel	egység	hányad	
d	39	0,43	
n	9	0,10	
$f=d+n/$	48	0,53	
$P_d$	37	0,41	
$P_n$	5	0,06	
$p=P_d+P_n/$	42	0,47	
t	90	1,00	

4.sz. táblázat

A MUNKAÜGY. MUNKATUDOMÁNY MAKROTEZAUROSZ FOGALMAINAK  
FUNKCIONÁLIS MEGOSZLÁSA

tétel	egység	hányad	
d	244	0,43	
n	51	0,0 <sup>o</sup>	
$f=d+n/$	295	0,52	
$P_d$	219	0,39	
$P_n$	51	0,09	
$p=P_d+P_n/$	270	0,48	
t	565	1,00	

$$R = \frac{n}{d} = \frac{256}{2347} = 0,11$$

Az összehasonlításként bemutatott tezaurusz-modellek R-értékei a következők:

- 0,23 /kőzet tezaurusz-modell/
- 0,21 /munkaügy, munkatudomány makrotezaurusz/
- 0,23 /munkabér tezaurusz-modell/

Az összehasonlítás arra mutat, hogy a tezaurusz R = 0,11 értéke a szükségesnél mindenképpen alacsonyabb redundancia-hátteret takar. A számítástechnika szakterülete modern, dinamikus fejlődésben lévő ismeretág. Fogalom-készletének tulnyomó része angol formában került a magyar szakemberek elé. Számos területen még nem rendelkezünk teljes értékű megfelelőekkel, sok esetben egymás mellett él az angol nyelvű fogalomváltozat és a magyar szakszó /pl. printer - gyorsnyomtató/, számos esetben pedig több magyar megfelelő is forgalomban van egy-egy adott angol szakkifejezésre, mintegy ezzel is jelezve, hogy a magyar számítástechnikai szaknyelv még "in statu nascendi" található. Ezt a helyzetet ismerve, minden különösebb vizsgálat nélkül, a tezaurusz néhány lapjának pusztá átolvasása révén is a kötetbe fel nem vett szinonimáknak, értelmi megfelelőknek tucatjait találhatjuk.

Az R = 0,11 érték igen alacsony volta, illetve a fenti megfontolás egyértelműen utal egy későbbi, dinamikai jellegű elemzés fontosságára. Az ugyanis bizonyosra vehető, hogy a tezaurusz gyakorlati felhasználása egy adott információrendszerben, jelentős non-deszkriptor karakterű anyaggal gazdagítja majd a fogalom-állományt. Jelen megjegyzésünk mindössze annyit kíván rögzíteni, hogy a tezaurusz non-deszkriptor állományának kiteljesítése szükséges és várható. Könnyű belátni, hogy a gyakorlati életben felvetődött igények kielégítését segítő tezaurusz annál hasznosabbá válik, annál több irányból közelíthető meg, minél gazdagabb szinonima- és kvázi-szinonima-környezet veszi körül az egyes deszkriptorokat.

A tezaurusz szerkezetének egy másik lényegi jellemzője a fogalom-szervezési szint /L/ mérőszáma. Ez a paraméter annak vizsgálata révén határozható meg, hogy mennyiben szerepelnek a tezauruszban egytagu, illetve többtagu fogalmak, tehát: milyen helyet foglal el a tezaurusz a tisztán unitermes, illetve a többszörösen összetett fogalmakból építkező, polikoncept típusu tezauruszok között. A fogalom-szervezési szint mérőszáma /L/ a következő képlettel határozható meg:

$$L = \frac{p}{f} = \frac{1989}{2603} = 0,76$$

Az összehasonlítási alapként bemutatott modellek L-értékei a következők:

- 0,61 /kőzet tezaurusz-modell/

0,92 /munkaügy, munkatudomány makrotezaurusz/0,88 /munkabér tezaurusz-modell/

Az  $L = 0,76$  érték belül van azokon a határokon, amelyeket a szemléltetésül megadott referenciák vázolnak, de tartalmi megértéséhez rövid magyarázat szükséges. A tisztán unitermes fogalom-gyűjtemények esetében  $L = 0$  /zéró/. A tisztán kéttagu /unikoncept/ típusu fogalmakat tartalmazó gyűjteményeknél  $L = 1$ . A tisztán polikoncept típusu fogalmakból álló tezauruszok esetében pedig  $L = 2$ . A valóságban azonban ilyen tiszta, szélsőségesen rendezett fogalmi konstrukciókra nem lehet számítani, az adott fogalom-gyűjtemény mindig több alapelv eredő értékével jellemezhető csak. A Számítástechnikai Fogalmak Tezaurusza tehát átmeneti jellegzetességeket mutat a tisztán unitermes és az unikoncept fogalomszervezésű tezauruszok között. Ugy tűnik - több modell vizsgálatának egybehangzó "vallomása" alapján -, hogy a magyar nyelvű szaktezauruszokra, sőt: talán a magyar nyelv szakfogalom-szerkesztési alapelveire is jellemző lehet az  $L = 0,5-1,0$  közötti érték. Ez az észrevétel még további megfigyelést és elemzéseket kíván, de valószínűnek látjuk, hogy objektív tendenciáról van szó.

## 2. A FOGALMAK RELÁCIÓ-HÁLÓZATÁNAK VIZSGÁLATA

A reláció-hálózat vizsgálata során voltaképpen a tezaurusz "belső intelligenciáját", kapcsolási képességét, asszociativitását, tranzitivitását, belső logikáját vesszük szemügyre, ily módon - áttételesen bár - a tezaurusz fogalom-állománya megmunkálásának szellemi tevékenységét minősítjük.

A tezaurusz teljes reláció-hálózata csak kétoldalu, kölcsönös kapcsolatokból tevődik össze. A tezauruszban alkalmazott reláció-típusok a következők:

- hierarchikus relációk / $r_h$ / Jelölés: G - S
- affinitív relációk / $r_a$ / Jelölés: R - R
- preferenciális relációk / $r_p$ / Jelölés: T - L

A tezaurusz a hierarchikus reláció-típust két felmenő szinten is tartalmazza:

- közvetlen generikus relációk /G/  $G = 1204$
- legfelsőbb generikus relációk /\*G/  $*G = 2140$

A tezaurusz relációinak elemzéséhez nekünk a G kapcsolatok mennyiségére van szükségünk, a tezauruszból nyerhető  $\bar{G}$ -érték azonban nem hűséges kifejezője a valóságnak, mivel számos olyan deskriptor van az anyagban, amelynek közvetlen generikus /G/ kapcsolata egyszersmind

legfelsőbb generikus kapcsolata /\*G/ is, hiszen a deskriptorok jelentős csoportjának csak egyetlen felmenő kapcsolata van. A teaurusz ezeket a kapcsolatokat is \*G jelöléssel adja meg. Ezért a reláció-hálózat elemzéséhez a következő adatot kell tekintetbe vennünk:

$$*G = 2140$$

Az \*G kapcsolatjelölések számának - elvileg - egyeznie kell az S /specifikus/ kapcsolatok számával. Ez érthető, hiszen ha X generikus fogalomhoz kapcsolódik X<sub>1</sub> specifikus fogalom, akkor - kétoldalu kapcsolatok esetében - X<sub>1</sub> specifikus fogalomhoz is szükségképpen kapcsolódik egy X generikus fogalom. Ez a teljes mennyiségi azonosság - nyilván egyszerű technikai okokból - nem áll fenn a teauruszban, a hierarchikus kapcsolódás két iránya között, hiszen

$$S = 2103$$

Ily módon

$$r_h = *G + S = 4243$$

Az affinitív relációk /r<sub>a</sub>/ összege

$$r_a = 766,$$

ami 383 fogalompár rokonsági /asszociációs/ kapcsolatát fejezi ki. Az affinitív relációk értelmezését a teaurusz bevezetője világosan és helyesen határozza meg.

A teaurusz preferenciális relációinak hálózatát részint az L /non-deszkriptortól deskriptorhoz vezető/ relációk, részint pedig a T /tiltott fogalom/ relációk adják meg. Ezek számának is egyeznie kellene. A teauruszból kigyíjtott adatok alapján

$$L = 254$$

$$T = 252$$

Ily módon a teaurusz preferenciális relációinak /r<sub>p</sub>/ száma:

$$r_p = L + T = 506$$

Összegezve, a teaurusz összes relációinak száma:

$$r = r_h + r_a + r_p = 4243 + 766 + 506 = 5515$$

Az összes relációk megoszlását az 5.sz. táblázat mutatja be.

## A TEZAURUSZ RELÁCIÓINAK MEGOSZLÁSA

jelölés	egység	hányad
$r_h$	4243	0,77
$r_a$	766	0,14
$r_p$	506	0,09
$r$	5515	1,00

E táblázat adatainak értékeléséhez bemutatjuk a hasonlítási alapként már idézett fogalom-gyűjtemények megfelelő adatait /6.sz. táblázat/.

Az összehasonlítás azt mutatja, hogy a teaurusz markánsan kidolgozott hierarchikus kapcsolat-hálózata mellett az affinitív és -különösen - a preferenciális relációk hálózata még nem eléggé fejlett - a referencia-adatokhoz viszonyítva /márpedig ez utóbbi két kapcsolati kategória az idézett fogalom-gyűjteményekben sem megnyugtatóan kimunkált/.

Egyébként a preferenciális relációkra vonatkozó számszerű adatok alacsony volta is az aránylag kevés számú szinonima feltüntetésével függ össze.

A kapcsolat-hálózat funkcionális megoszlásának bemutatása után lehetővé vált a teaurusz egy újabb jellemzőjének megállapítása, meghatározható a teaurusz egy adott deskriptorának kapcsolódási mérőszáma /k/:

$$k = \frac{r}{d} = \frac{5515}{2347} = 2,35$$

Hasonlítási alapként bemutatjuk a már idézett három forrás megfelelő jellemzőit /7.sz. táblázat/.

Az összehasonlítás egyértelmű eredménye az a megállapítás, hogy a teaurusz fogalom-hálózata még - egészében véve - nem megfelelően részletezett, továbbfejlesztése feltétlenül indokoltnak látszik.

Nem érdektelen annak vizsgálata sem, hogy az egy deskriptorra jutó összes relációk hogyan oszlanak meg az egyes kapcsolat-típusok között. Ezt a 8.sz. táblázat szemlélteti.



## 6.sz. táblázat

## NÉHÁNY TEZAUROSZ-MODELL RELÁCIÓINAK MEGOSZLÁSA\*

Jelölés	Az összes relációk hányada		
	Közet	Munkaügy	Munkabér
$r_h$	0,59	0,71	0,69
$r_a$	0,18	0,13	0,15
$r_p$	0,23	0,16	0,16
$r$	1,00	1,00	1,00

## 7.sz. táblázat

NÉHÁNY TEZAUROSZ-MODELL FAJLAGOS RELÁCIÓ-MENNYISÉGÉNEK /r/d/  
BEMUTATÁSA\*\*

	Közet	Munkaügy	Munkabér
fajlagos reláció- mennyiség /k/	3,36	2,67	2,82

\* A táblázat adatai e tanulmány összehasonlító vizsgálatai számára átdolgozott értékek, mivel az idézett eredeti közleményekben leírt fogalom-gyűjtemények egyirányú /nem kölcsönös/ relációkat is tartalmaznak, s az összehasonlítás érdekében ezeket levontuk a modellek teljes reláció-mennyiségéből, a hányad-értékeket pedig megfelelően módosítottuk.

\*\* Az összehasonlítás kedvéért átszámított adatok.

## 8.sz. táblázat

A TEZAURUSZ EGY DESZKRIPTORA KAPCSOLATAINAK MEGOSZLÁSA  
KAPCSOLAT-TIPUSOK SZERINT

kapcsolat- típus	egység
$r_h$	1,81
$r_a$	0,33
$r_p$	0,21
$r$	2,35

Hasonlításul bemutatjuk a referencia-források megfelelő megoszlási adatait /9.sz. táblázat/.

## 9.sz. táblázat

NÉHÁNY TEZAURUSZ-MODELL EGY DESZKRIPTORÁRA JUTÓ FAJLAGOS  
KAPCSOLAT-MENNYISÉGÉNEK MEGOSZLÁSA KAPCSOLAT-TIPUSOK SZERINT

referencia- modell	$r_h$	$r_a$	$r_p$
Közet	1,97	0,63	0,76
Munkaügy	1,90	0,35	0,42
Munkabér	1,95	0,41	0,46

Az összehasonlítás nemcsak azt mutatja meg, hogy a tezausz lényegesen kevésbé fejlett reláció-hálózattal rendelkezik, mint a bemutatott referenciák, hanem azt is megmutatja, hogy elsősorban a preferenciális relációk száma szorul szaporításra. Eddigi tapasztalataink alapján úgy véljük, hogy a megfelelően kifejlesztett tezauszoknak  $r_a$  és  $r_p$  értéke 0,5 és 1,0 között változhat, az adott szakterület specifikus adottságai szerint, de nemigen maradhat 0,5 alatt. Ez a megjegyzés tapasztalati természetű, bár számos eddigi elemzés támasztja alá. Egyértelmű azonban az is, hogy a vizsgált tezausz még a gyakorlati felhasználás előtt áll, ami kétségkívül gyarapító hatással lesz a preferenciális és affinitív relációk számára.

## 3. A SZERKEZET MÉLYSÉGI TAGOZÓDÁSÁNAK VIZSGÁLATA

A kapcsolat-hálózat áttekintő elemzése után - a tezausz hierarchikus indexének elemzése alapján - vizsgáljuk meg a tezausz-szerkezet mélységi tagozódását. A fogalom-állomány megoszlását az egyes hierarchia-szintek szerint a 10.sz. táblázat, illetve - egy referenciaforrással összevetve - az 1.sz. ábra mutatja be.

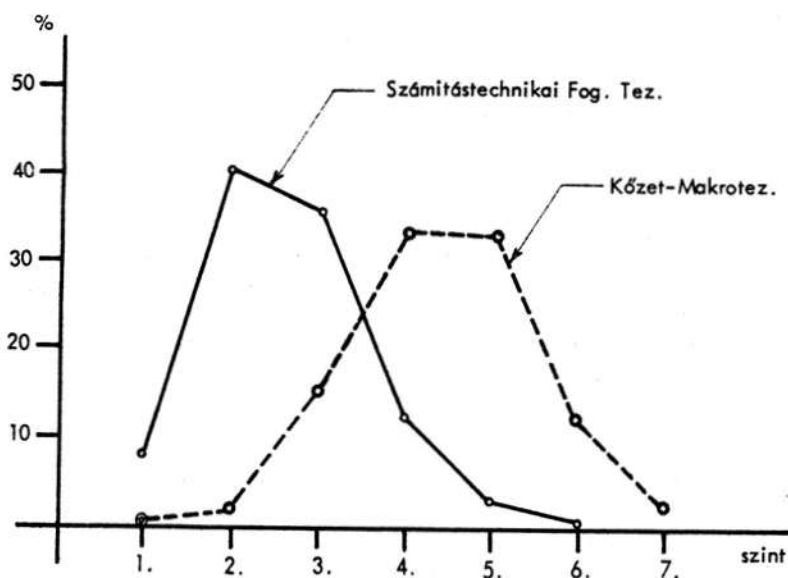
10.sz. táblázatA TEZAUSZ FOGALOM-ÁLLOMÁNYÁNAK MEGOSZLÁSA  
HIERARCHIA-SZINTEK SZERINT

hierarchia-szint	egység	%
1.	197	8,4
2.	944	40,2
3.	838	35,7
4.	293	12,5
5.	63	2,7
6.	12	0,5
1-6.	2347	100,0

Az 1.sz. ábrán pontozott vonallal bemutatott referenciaforrás jellegzetességei jó megegyezést mutatnak a tezauszéival - egy hierarchia-lépcső eltolással. Mint azt a tezausz bevezetője is kifejti, a tezausból - indokoltan - hiányzik az egyértelmű, felső generikus összefonódás /amit a referencia-forrás 1. hierarchia-szintje tartalmaz/. Ebből a szempontból jobbnak kell itélnünk a tezausz strukturáját, mint a felhasznált referenciákét. Egyik legjellemzőbb pozitív tulajdonsága a tezausznak éppen az, hogy viszonylag nagyszámú /197/ fogalom található az 1. /csoportképző deszkriptorokat tartalmazó/ hierarchia-szinten. Ez a jellegzetesség hangsúlyosan mutatja be a laza hierarchikus bontás - igen helyes - szemléletmódját, illetve a tezausz bevezetőjében kifejtett szerkesztési alapelv megvalósulását.

Az ábra tanúsága és a 10.sz. táblázat adatai szerint, a tezausz gerince /a fogalom-állomány 3/4 részénél több/ a 2-3. hierarchia-szinteken található. Az ábra egyébként jól közelíti az un. normál-eloszlási görbét, amit különböző tezauszok esetében egyaránt jellemzőnek találtunk a fogalmi eloszlás grafikus ábrázolásában.

## A TEZAURUSZ FOGALMAINAK SZÓRÓDÁSA HIERARCHIA-SZINTEK SZERINT



Tezauruszról lévén szó, mégpedig olyan tezauruszról, amely arra a célra készült, hogy információfeltáró, -tároló és -kereső rendszerek eszközeivé válják, a fogalmak hierarchikus tagolása kifejezetten a felhasználási szempont kiemelésével - és ezért: olykor a szigorú tudománylogikai megfontolások háttérbe szorításával - történt. Ezzel a szemléletmóddal azonban minden olyan információs szakembernek egyet kellene értenie, aki az információkezelés mindennapos gyakorlatában él. Ez a szemléletmód olyan "önkényes" megoldásokat kíván meg, amelyek lényegesen nehezebb munkát okoznak, mint a legszigorúbb tudománylogikai osztályozási elvek alkalmazása. A tezaurusz a fogalmi rend világa, de nem a merevségé, hanem a dinamikus, a változásra min-

dig kész és alkalmas rendé. Találón fejezi ki ezt a szemléletmódot Mephistoteles egy mondata GOETHE Faust-jában /6/:

"Grau, teurer Freund, ist alle Theorie,  
und grün des Lebens goldner Baum."

/"Szürke, öcsém, minden teoria,  
s az élet aranyfája zöld."/

A fentiek során a teaurusz statikai elemzésének néhány fontos irányát követtük s megállapítottuk a kialakított fogalom-állomány és reláció-hálózat néhány szerkezeti és mennyiségi jellemzőjét. Világosan látjuk azonban, hogy a teaurusz gyakorlati használhatóságát, a fogalom-állomány célszerű méretét és szükséges - horizontális és vertikális - tagolását majd csak a gyakorlati kísérletek adataira alapozott dinamikai elemzés eredményei határozhatják meg, s indíthatják el a használati szempont kiemelésével megcélzott optimum felé.

\*

Az elemzéshez felhasznált és idézett szakirodalom

- /1/ Számítástechnikai fogalmak teaurusza. Szerk.: Schiff Ervin. Bp. 1973. OMKDK, 396 p.
- /2/ SCHIFF Ervin: Magyar nyelvű számítástechnikai teaurusz. I.-TMT, 20.k. 11-12.sz. 1973.nov.-dec. p.785-809.
- /3/ MOLNÁR Imre: Ujszerű teaurusz-szerkezet kialakítása. = TMT, 19.k. 3.sz. 1972.márc. p.149-169.
- /4/ MOLNÁR Imre: Az információszaketeaurusz fogalma, rendszere és felépítése, különös tekintettel az építészeti ágazati szaketeaurusz kialakítására. Bp. 1973. ÉTK. 392 p.
- /5/ MOLNÁR Imre: Az információkezelés fogalomrendszere. Bp. 1973. Közgazd. és Jogi Kiadó. 218 p.
- /6/ GOETHE, J.W.: Faust. Gesamtausgabe. Leipzig, é.n. Inselverl. 573 p. Teil 1. 190 p. /magyarra Sárközi György fordítása/

\*

MOLNÁR, I.: Thesaurus on computer techniques in Hungarian. Part II. Comparative static analysis of the thesaurus

The "Számítástechnikai Fogalmak Tezaurusza" /Thesaurus on Computer Techniques/ is examined, a statical analysis of descriptor inter-relationships and of structural depth is given in 3 parts:

1. In the course of the word-stock analysis the followings are illustrated: quantitative indicators of the functional distribution; value of the redundancy factor /R/ referring to synonyme-ampleness; index of the level of word compilation /L/ referring to characteristics of descriptor-complexities. These characteristics compared with the concerning data of some thesaurus models prove that the Thesaurus consists of relatively few non-descriptor surroundings, however, it is a collection with suitable structural level.

2. Descriptor inter-relationships are analyzed. In the meantime basic types of inter-relationships /hierarchical, affinitive and preferential/ and their quantitative division is stated. Specific relation value /k/ per descriptor expressing the stage of inter-relationships development as well as the structure of the various connections are also examined.

Parameters registered are also compared with reference data.

3. As to the structural depth it is found that the Thesaurus consists of 6 hierarchical levels, this structure proves the application of the weak hierarchical principle. The dispersion of terms within these levels is in congruence with the normal dispersion diagram. The statical analysis is to be continued later by a dynamic one adopting characteristics from the use of the Thesaurus.

\*

МОЛНАР, И.: Тезаурус по вычислительной технике на венгерском языке. Часть II. Сравнительный статический анализ тезауруса

В статье в трех частях дается статический анализ словарного запаса, сети связей между понятиями и также глубинной структуры тезауруса по вычислительной технике.

1. В ходе анализа словарного запаса вычисляются следующие величины:

- количественные показатели функционального распределения понятий;
- величина фактора редунданции (т.е. отношение числа недескрипторов к числу дескрипторов), характеризующая "богатство" тезауруса в синонимах;

- величина, характеризующая уровень сложности понятий (т.е. величина, равная отношению числа всех пермутированных вариаций к числу всех понятий).

Сравнивая полученные данные с данными других тезаурусов утверждается, что данный тезаурус обладает сравнительно небольшой окрестностью недескрипторов, но имеет достаточно высокий уровень сложности понятий.

2. В ходе анализа сети связей вычисляются:

- количественное распределение основных типов связей (т.е. иерархических, аффинитивных и предпоченциальных);
- величина, характеризующая развитость сети связей и равная отношению числа всех связей к числу всех дескрипторов.

Полученные данные сравниваются с имеющимися.

3. В ходе анализа глубинной структуры выясняется, что тезаурус характеризуется 6 уровнями иерархии, а иерархические связи установлены по принципу так называемой "слабой иерархии". Распределение понятий между отдельными уровнями в достаточной мере совпадает с одной стороны с функцией нормального распределения, с другой стороны с распределительными особенностями других моделей. После статического анализа должен следовать и динамический анализ, в котором должны быть учтены и показатели тезауруса, полученные в ходе его использования.

\*

MOLNÁR, I.: Rechentechnischer Thesaurus in ungarischer Sprache  
II. Vergleichende statische Analyse des Thesaurus

Es wird der "Thesaurus rechentechnischer Begriffe" einer statischen Analyse seines Begriffsbestandes, des Relationsnetzes der Begriffe und seiner Tiefenstruktur, in die folgenden drei Teile gegliedert, unterzogen:

1. Im Laufe der Analyse des Begriffsbestandes werden die quantitativen Kennziffern der funktionellen Verteilung, der Wert des auf den Synonymenreichtum hinweisenden Redundanzfaktors  $/R/$  /d.h. Quoti-

ent aus den im Thesaurus enthaltenen Nichtdeskriptoren und Deskriptoren/, sowie die Messziffer des auf die Komplexität der Deskriptoren hinweisenden Begriffsorganisationsniveaus /L/ /d.h. Verhältnis der permutierten und der nicht-permutierten Ausdrücke/ festgestellt. Diese Kenndaten werden den einschlägigen Angaben einiger Thesaurusmodelle gegenübergestellt und aufgrund der erhaltenen Werte wird der Thesaurus als eine Sammlung mit einer geringen Nichtdeskriptorenumgebung klassifiziert, die jedoch ein entsprechendes Begriffsorganisationsniveau besitzt.

2. Bei der Untersuchung des Relationsnetzes wird die quantitative Verteilung der grundlegenden /hierarchischen, affinitiven und präferentiellen/ Relationstypen nachgewiesen und es wird der auf den Entwicklungsstand des Relationsnetzes hinweisende, für einen Deskriptor berechnete spezifische Relationswert /k/, bzw. seine Gliederung je nach den Relationen verschiedener Typen festgestellt.

Die ermittelten Parameter wurden mit den Referenzwerten ebenfalls verglichen.

3. Die Untersuchung der tiefenmässigen Gliederung der Struktur zeigte, dass der Thesaurus durch sechs Hierarchieniveaus charakterisiert wird und diese Gliederung beweist, dass sich das Prinzip der schwachen Hierarchie durchsetzt. Die Streuung der Begriffe je nach den einzelnen Niveaus steht in guter Übereinstimmung mit den Streuungscharakteristiken des vorgeführten Referenzmodells. Der statischen Analyse soll später unter Anwendung der Gebrauchskennwerte des Thesaurus eine dynamische Analyse folgen.

