

MAGYAR NYELVŰ SZÁMÍTÁSTECHNIKAI TEZAUROSZ I.

Schiff Ervin

Országos Műszaki Könyvtár és Dokumentációs Központ

Rendeltetése szempontjából meghatározva a tezaurusz olyan szaknyelvi utmutató eszköz, amelyet az indexelők vagy a felhasználók akkor vesznek igénybe, amikor a dokumentum nyelvéről az információs rendszer szigorubbán szabályozott nyelvére /dokumentációs nyelv, információs nyelv/ fordítanak, vagy ellentétes irányban, a rendszer nyelvéről vissza a természetes nyelvre.

Szerkezete szempontjából meghatározva a tezaurusz a jelentésük szerint hierarchikusan egymással kapcsolatban lévő kifejezések olyan megállapodásszerű és dinamikus jegyzéke, amely az ismeretek egy körülírt területét fogja át... a természetes nyelvből vett kifejezések szerkezettel rendelkező halmaza, amely dokumentumok, adatgyűjtemények tartalmának leírására alkalmazható...

/UNESCO/SC/WS/500, Paris, 22 December, 1971, Guidelines for the Establishment and Development of Monolingual Thesauri for Information Retrieval. p.5-6./

1973 végére megjelenik a Számítástechnikai Fogalmak Tezaurusza. A kötetet az OMKDK adja ki és korlátozott példányszámban terítés nélkül bocsátja az érdeklődők rendelkezésére. Ez már a második magyar nyelvű műszaki tezaurusz, amelynek előállítására, megjelentetésére, a tudományos, műszaki és gazdasági tájékoztatási munka elméletének és gyakorlatának fejlesztésére az OMF megbizást adott az OMKDK-nak. Szerkesztési munkáinak elvégzését igen nagy mértékben megkönnyítették az első, az általános műszaki tezaurusz összeállításánál szerzett tapasztalatok, kialakított módszerek és eljárások.

⁺ A cikk II. részében MOLNÁR Imre elemzi a kész tezauruszt.

Előzmények

Az OMKDK fejlesztési osztályán 1968-ban kezdtek meg a magyar nyelvű általános tezaurusz létrehozásával kapcsolatos gyakorlati munkát. Abban az időben már sorozatosan jelentek meg tezauruszok különböző idegen nyelveken /elsősorban angol nyelven/. Magyar nyelvű, a hazai /termelési, gazdasági, tudományszervezési stb. strukturális/ feltételeket is figyelembe vevő tezaurusz készítésével kapcsolatos tapasztalatok, módszerek és eljárások azonban nem léteztek, sőt tulajdonképpen a tezauruszokkal kapcsolatos terminológiai és alkalmazási viták sem ültek el. /A fent idézett definíció végleges formában csak 1971 december 22-én jelent meg./

A kérdés előzetes tanulmányozása arra a következtetésre vezetett, hogy

- az OMKDK-nak, amellet, hogy az ETO-val kapcsolatos célkitűzéseit teljesíti, meg kell kezdenie az osztályozási rendszerek másik nagy típusával a tezauruszokkal kapcsolatos munkákat /többek között ez a gépi információkeresés szellemi előkészítésének egyik alapvető lépése is/;

- ez a munka azonban abból a célból, hogy általánosan is alkalmazható tapasztalatokat szolgáltatson, lehetőleg ne kötődjön szűk szakterülethez;

- azért is célszerűbb általános tematikájú tezauruszt előállítani, mert a munka uttörő, iránymutató jellegű és a később kialakítandó szaktezauruszok kidolgozásához az általános tezaurusz szerkesztésével, szókészletével alapot adhat, biztosíthatja a szaktezauruszok elvi kompatibilitását;

- a tezauruszépítéssel kapcsolatos általános problémákat a szakirodalom széles körben tárgyalja, ezért különös figyelmet a magyar nyelv sajátosságai által okozott problémákra kell fordítani;

- a fenti feltételeket /valamint a tapasztalatok hiányát és az OMKDK lehetőségeit/ figyelembe véve, célszerűbbnek látszott ezt az általános műszaki tezauruszt nem teljesen önálló alkotásként létrehozni, hanem valamely, a nemzetközi gyakorlatban már többé-kevésbé elismert tezauruszra támaszkodva; bár teljesen kizárt az a lehetőség, hogy egy idegen nyelvű tezaurusz szakfordításával vele egyenértékűen használható /vagy egyáltalán: használható/ magyar nyelvű tezaurusz jön létre. Így ez a kiszemelt tezaurusz csak mintául, szerkesztés közbeni utmutatóul szolgálhat majd, illetve biztosítja a magyar nyelvű tezaurusz-munkálatok és a nemzetközi tezauruszépítési tevékenység kapcsolatát.

Az akkortájt rendelkezésre álló idegennyelvű tezauruszok közül a kitűzött céloknak a legjobban az 1964-ben kiadott EJC tezaurusz /az Engineers Joint Council kiadásában megjelent Thesaurus of Engineering Terms/felelt meg. Szerkezete és felépítési elve a leginkább elfogadott nézeteket tükrözte, iskolát nyitott; figyelembevételével a nemzetközi gyakorlatban a szaktezauruszok sorát dolgozták ki.

A lényegét nem érintően, inkább csak a felölelt témakörök bővítésével és indexekkel továbbfejlesztett változata, a TEST tezausz 1967-ben jelent meg. Ez utóbbi viszont már mindenben kielégíti a fentebb említett, 1971 decemberi UNESCO ajánlásban rögzített meghatározást és egyéb feltételeket. A magyar nyelvű műszaki tezausz létrehozásakor az EJC tezauszából indultunk ki. Későbbi munkáinkat befolyásolták a TEST /Thesaurus of Engineering and Scientific Terms/ tezausz felépítéséből levonható következtetések is. Nem bővítettük azonban az EJC tezausz általános műszaki szókészletét a TEST tezauszban található biológiai, orvosi stb. kifejezésekkel.

Az 1970-ben megjelentetett "Általános Műszaki Fogalmak Tezausza" szerkesztési munkái olyan tapasztalatokkal jártak, amelyek általánosíthatók. Ezek részben a fogalmakkal magukkal, részben a szókészlettel, illetve a tezausz szerkezetével kapcsolatosak.

A tezauszba felvett /és a szókészlettel tükrözni kívánt/ fogalmak tekintetében megállapítható volt, hogy az alapul vett tezausz fogalomkészlete a hazai feltételeknek nem felel meg /például túl részletezett a többi szakterület bontásához viszonyítva a speciális ásványolajkitermelési fogalomkészlet, űrkutatási -rakétatechnikai fogalomkészlet, míg a vegyipar - gyógyszeripar - műanyagipar csak gyengén reprezentált; de tartalmaz a tezausz óceánkutatási fogalmakat; a közgazdasági fogalmak a kapitalista gazdasági viszonyoknak felelnek meg stb./, így a fogalmak, valamint a fogalmak közötti - a tezauszban feltüntetett - kapcsolatok sem vehetők át mechanikusan az eredetiből. A tudományok rendszerezésének, az ipar ágazati szerkezetének a hazaiétól eltérő volta ismét azt eredményezte, hogy a fogalmak közötti kapcsolatok a hazai gyakorlatnak nem minden esetben feleltek meg. Nemkülönböztetett vitássá tett egyes fogalmi kapcsolatokat a terminológiai tisztázatlanság, vagy vita /például, hogy az öntészet a kohászathoz vagy a gépiparhoz tartozik-e, a "metallurgy" pontos fordítása: kohászat vagy csak a kohászati folyamatokat jelenti stb./.

A terminológiai kérdések átvezettek a szókészlet kérdéséhez. Az egyes nyelvek különböző területeken egymástól eltérő fejlettsége azt eredményezte, hogy az esetek többségében

- egy magyar kifejezés több angol kifejezéssel körülírható fogalmat foglalt magában, vagy

- egy angol kifejezés több magyar kifejezéssel megadható fogalomnak felelt meg.

Ez a tény a szókészlet jelentős átdolgozását kívánta meg /az eredetihez képest szavak elhagyását, más szavak felvételét, a feltüntetett kapcsolatok revidálását, az újonnan felvett szavak kapcsolatainak kijelölését stb./.

A magyarban elkerülhető volt az a - az angol eredetiben alkalmazott - kényszermegoldás, amelynek értelmében a főneveken kívül mellékneveket is felvettek a tezauszba. Bebizonyosodott, hogy a magyarban nem szükséges - sok esetben félreértésre is vezet - a főnevek egyes, illetve többes számú alakjának megkülönböztetett alkalmazása /ami az

angolban magától értetődő a megszámlálhatatlan és a megszámlálható főnevek szigorú megkülönböztetése miatt/.

A szerkesztési munka során igen nagy problémát okozott az összetett szavak helyesírása. Az a nézet alakult ki, hogy az egybe-, illetve különírás szabályait értelemszerűen kell, és nem szabad mechanikusan alkalmazni.

A tezausz szerkezetének folyamatos kimunkálása /az EJC, illetve a TEST tezausz szerkezeti felépítése nyomán/, a kifejezések közötti fogalmi kapcsolatok kijelölése, a különböző szerkezetű indexek létrehozása visszahatott magára a szerkesztői tevékenységre. Minden egyes kapcsolattípus kijelölése, egy újabb szempont szerinti index kialakítása újból és újból a teljes szókészlet átvizsgálásával és helyesbítésével járt. Így a tezausz végső szókészlete és a szavak közötti - a tezauszban feltüntetett - kapcsolatok rendszere a különböző szempontok szerinti iteratív közelítés révén alakult ki. Az elkészített indexek - a szakcsoporthoz és a permutált index - nagymértékben megkönnyítik a tájékozódást is a tezauszban.

Az eddig tárgyalt általános műszaki tezausz létrehozásának célja, a tapasztalatszerzésen kívül, egy olyan alapszókészlet, valamint a szavak segítségével kifejezett fogalmak közötti összefüggéseket tükrözni képes szerkezet kialakítása volt, amelyre alapozva szakmai /ágazati/ tezauszok sora építhető ki. Az első ilyen szaktezausz építését 1970-ben kezdtük meg. A cikk a továbbiakban ezzel a - számítástechnikai tárgykörű - tezausszal foglalkozik.

A számítástechnikai tezausz összeállítása

A számítástechnikai tezausz létrehozására irányuló munka célja és feltételei bizonyos fokig eltértek az előzőekben vázoltaktól.

A cél ebben az esetben egy

- a számítástechnikai tájékoztatás gyakorlatában,
- az információk kezelésére és keresésében
- egyezményesen használható,
- magyar nyelvű
- speciális szakterületi tezausz

kialakítása volt.

Így a szerkesztési munkák során a gyakorlati alkalmazhatóság szempontjait kellett elsődlegesen figyelembe vennünk.

A feltételek több vonatkozásban is mások voltak, mint az általános műszaki teaurusz kialakításának feltételei.

A munkák indításakor olyan idegen nyelvű számítástechnikai teaurusz, amelyre támaszkodnunk lehetett volna, még nem létezett. Bizonyos teauruszok azonban részterületként foglalkoztak a számítástechnikával /mind készülék, mind alkalmazási oldalról/, és idegen nyelvű vagy magyarra fordított értelmező, valamint több nyelvű szótárak, különböző számítástechnikai tájékoztatási szolgáltatások szakcsoportlistái is rendelkezésre állottak. Ugyanakkor a magyar nyelvű számítástechnikai szakirodalom előzetes tanulmányozása során arra a megállapításra jutottunk, hogy a magyar nyelvű egységes terminológia kialakulásáról még korai lenne beszélni. Kisebb a probléma akkor, ha egy-egy kifejezésnek csak a fordítási szóképe vitás, például:

real time = valós idő, valós idejű, reál-idő, vagy azonos idejű;

time-sharing = időosztás, időosztásos, osztott idejű, osztott idő, vagy részüidő /illetve a fenti kifejezések eltérő helyesírású változatai/.

Vitás azonban sokszor a fogalom is, rosszabb esetben ezt a vitás fogalmat többféle magyar kifejezéssel jelölik. Például:

display = az angol eredeti értelemben mindenféle egység, amelyen adatok megjelennek, míg magyarban a katódsugárcsőes készülék, katódsugárcsőes berendezés, katódsugárcsőes megjelenítő, optikai kijelző, sőt a képmű kifejezéseken /amelyek helyett szinonimáként elterjedten a display kifejezést használják/ azt az egységet értik, amelyen az adatok katódsugárcső képernyőjén jelennek meg.

A számítástechnikai gyártó cégek propaganda-fogásai is gyakran a terminológiai zűrzavar fokozása irányába hatnak: az egyik cég "real time", a másik "interaktív" vagy "dialógus" /magyarítva: párbeszéd/ üzemmódról beszél, de e kifejezéseken mindkét cég lényegében ugyanazt érti. A fogalom értelmezői esetleg később árnyalatnyi különbségeket fedeznek fel vagy vélnek felfedezni, és ezzel mindkét kifejezés polgárjogot nyer.

A valódi szinonimák számát tovább növeli az a tény, hogy a számítástechnika lényegében idegen nyelven alakult ki, idegen nyelvű szakirodalom útján ismerkedett meg azzal a hazai szakközönséggel. A kifejezések átültetésével azonban nemcsak a számítástechnikus, hanem, éppen a számítástechnikai eszközök széles körű alkalmazhatósága miatt, más területeken dolgozó szakember is megpróbálkozott. Sok különféle helyi értelmezés, helyi kifejezőkészlet alakult ki, amelyeket azonban az adott szakterületen dolgozók publikációikban alkalmaznak. Így, annak ellenére, hogy a magyar nyelvű terminológiai egységesítés kérdése igen sokakat foglalkoztatott, ezt nem sikerült kialakítani, sőt nagyon sok szinonim kifejezés van használatban.

A feltételek számbavételénél figyelembe kellett venni azonban azt is, hogy már a munka indításakor rendelkezésre állt szókészletével és szerkezetével az általános műszaki teaurusz, valamint az

OMKDK-ban felgülemlett tezaurusz-szerkesztési tapasztalatok is /sőt - az INIS rendszerbe kapcsolódás és az INIS tezaurusz gyakorlati alkalmazás révén - tezaurusz-alkalmazási tapasztalatok is/. Már hosszabb ideje létezett olyan magyar nyelvű tájékoztatói szolgáltatás, amelynek szókészletéből ki lehetett indulni. Ez a Gyorsindex - Számítástechnika és a Gyorsindex - Automatizálás sorozat volt.

E sorozatok egyes füzetei egy adott időközben /két hét, illetve egy hónap alatt/ érkezett különböző, a profilba tartozó dokumentumok - cikkek, kutatási jelentések, szabványok, szabadalmak, vállalati irodalom - tartalmát cimszerűen kifejező mondatokat tartalmaznak. E mondatokat a dokumentum tartalmával való megismerkedés után szakemberek állítják össze, megjelölve a szakkifejezéseket. Az indexeket a jelölések alapján számítógép készíti el úgy, hogy a mondatokat a bennük található szakkifejezéseknek megfelelően betűrendbe rendezi. Így egy-egy füzet felfogható a szakkifejezések szöveggörnyezetes, betűrendbe rendezett jegyzékének.

A SZÁMÍTÁSTECHNIKAI SZÓKÉSZLET ÖSSZEGYÜJTÉSE ÉS AZ ELSŐDLEGES VÁLOGATÁS

A szerkesztési munka már a szakkifejezéseknek a Gyorsindexekből való kigyűjtése során megkezdődött. A szakemberek, akik ebben a munkában részt vettek, olyan utasítást kaptak, hogy a kifejezéseknek az index szövegéből történő kiemelésénél

- azokat névszói formában és egyes szám alanyesetben írják ki /például: számítógép, adatrögzítés/;
 - más szófajok csak összetett kifejezések esetében, azok belsőjében fordulhatnak elő /például: digitális számítógép, mágnesszalagos adatrögzítés/;
 - a számítástechnikai kifejezéseket lehető legspecifikusabb formájukban, míg más szakkifejezéseket általánosabb formájukban rögzítik /például az alábbi mondatból:
- ... MOS /Metall Oxid Semiconductor/ félvezető tömeg tároló fejlesztése az USA-ban ... kigyűjtendő kifejezések
- "MOS félvezető tömeg tároló" és
- "fejlesztés", azaz nem tárolófejlesztés vagy tömeg tárolófejlesztés stb./;
- egyes esetekben a mondatok átfogalmazásával állítsák össze és szerkesszék át a kigyűjtésre kerülő szakkifejezéseket /például az alábbi esetben:
- ... automatikus vegyi elemzés mérési adatainak gépi feldolgozása ...

többek között a

"mérési adat" és az

"adatfeldolgozás"

kifejezések emelendők ki; ez utóbbi példa az átszerkesztésre is: "adatainak feldolgozása", azaz birtokviszony helyett "adatfeldolgozás" a megfelelő formájú kifejezés/.

E munka eredményeképpen több mint 100 ezer kifejezés gyűlt össze, cédulákon és gépi lyukkártyákon rögzítve. A továbbiakban az alkalmazott technika leírásával nem foglalkozom, csak megjegyzem, hogy minden egyes műveletnél a listákkal, cédulákkal teljesen azonos tartalmu lyukkártyasorozatot is előállítottunk a nagy volumenű írásmunka automatizálására.

A cédulákat betűrendbe rendeztük és a gyakoriság feljegyzése mellett egy cédulára vontuk össze a többszörös előfordulásban megjelenő kifejezéseket. E munka során azonban már megkezdtük azt a változtatási és értelmezési tevékenységet is, amelyet aztán egészen a teaurusz végső szövege lezárásának pillanatáig folytattunk. Minden egyes kifejezést megvizsgáltunk ugyanis abból a szempontból, hogy a teauruszban vélelmezhetően szükség lesz-e megtartására:

- nem tulzottan specifikus-e, vagy éppenséggel nem tulzottan általános-e;

- nem helyettesíthető-e más, a teaurusz szerkezetéhez, fogalomkészletéhez jobban illő kifejezéssel;

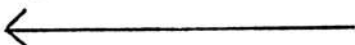
- szakszerűen értelmezett vagy értelmezhető-e: nem helyettesíthető-e pontosabban értelmezett kifejezéssel;

- nem tulzottan kicsi-e az előfordulási gyakorisága és/vagy szakmai szempontból nem jelentéktelen fogalom-e;

- amennyiben a kifejezés deszkriptorként nem kerül be a teauruszba, szükség van-e arra, hogy legalább utaljunk rá /azaz felvegyük az ún. tiltott kifejezések - non-deszkriptorok - közé/.

Ehhez a munkához vitás esetekben különböző kézikönyveket, lexikonokat, értelmező szótárakat, szakcsoportlistákat, idegen nyelvű /nem számítástechnikai/ teauruszokat vettünk igénybe, a döntés ezek figyelembevételével született meg. Ugyancsak nagy segítségünkre volt a munka egyik közbülső lépésében előállított "a tergo" szójegyzék /azaz az olyan típusu felsorolás, amely a kifejezéseket nem a szokványos, előlről hátrafelé haladó, hanem a fordított, hátulról előre haladó betűrendben tartalmazza /1. ábra//. Ebben az esetben ugyanis a kifejezések más metszetükben jönnek össze, mint a szokványos betűrendezésnél, és jól ki lehet használni a szerkesztési munkában azt a tényt, hogy az általánosabb fogalmat általában az összetett kifejezés végén található szó adja meg. Az alábbi kifejezésben például

integrált áramkörös fix-tároló



...
 szervezőautomata
 lyukszalagos szervezőautomata
 helykoordináta
 posta
 csőposta
 mágneslemez-kazetta
 mágnesszalagkazetta
 mágneskazetta
 perspektíva
 műhold pálya
 kártya
 logikai kártya
 lyukkártya
 tizenkét oszlopos lyukkártya
 tartós lyukkártya
 szegély-lyukkártya
 peremlyukkártya
 kétsoros peremlyukkártya
 filmlykkártya
 fénylyukkártya
 duálkártya
 programkártya
 mágneskártya
 fényceruza
 elektronnyaláb
 mágnesdob
 piac
 felvevő piac
 számítógép piac
 világpiac
 markov-lánc
 licenc
 számítógépgyártási licenc
 vállalati licenc
 család
 számítógép család
 algoritmus család
 absztrakt programnyelvcsalád
 beszéd
 átállási sorrend
 trend
 információ-adatfeldolgozási trend
 fejlődési trend
 termelési trend
 kód
 aritmetikai korrigáló kód
 javító kód
 ...

1. ábra. Részlet a számítástechnikai tezaurusz szókészletének szerkesztése során használt "a tergo" szójegyzékből

a nyíl irányában haladva egyre pontosabban írjuk körül azt, hogy milyen tárolóról van szó. Ugyanakkor ez az eljárás az összes azonos végződésű, esetünkben a "tároló" szóval végződő kifejezést egy csoportba hozta össze. Az a tény, hogy a kifejezések között különböző összetételekben előfordult a "memória" szó is, szükségessé tette, hogy döntsünk /ez a döntés a "tároló" javára született meg/, és a - az "a tergo" szótár alapelve következtében szintén egy csoportban található - "memória" végződésű kifejezések korrigálása, adott esetben a memória és a tároló szavak cseréje így könnyen megtörténhetett.

A fenti, hatalmas volumenű munka, amelynek célja a számítástechnikai tezausz szókészletének összeállítása legalább első közelítésben, azonban nem lehetett végbe csak az egyes kifejezéseknek egyedi, illetve egymáshoz viszonyított kapcsolatuk vizsgálatával. Ehhez a kiindulási szókészlet túl nagy, teljességében még a listákban is áttekinthetetlen volt.

A szakcsoportos index összeállítása, az első szakértői munkacsoporti revízió

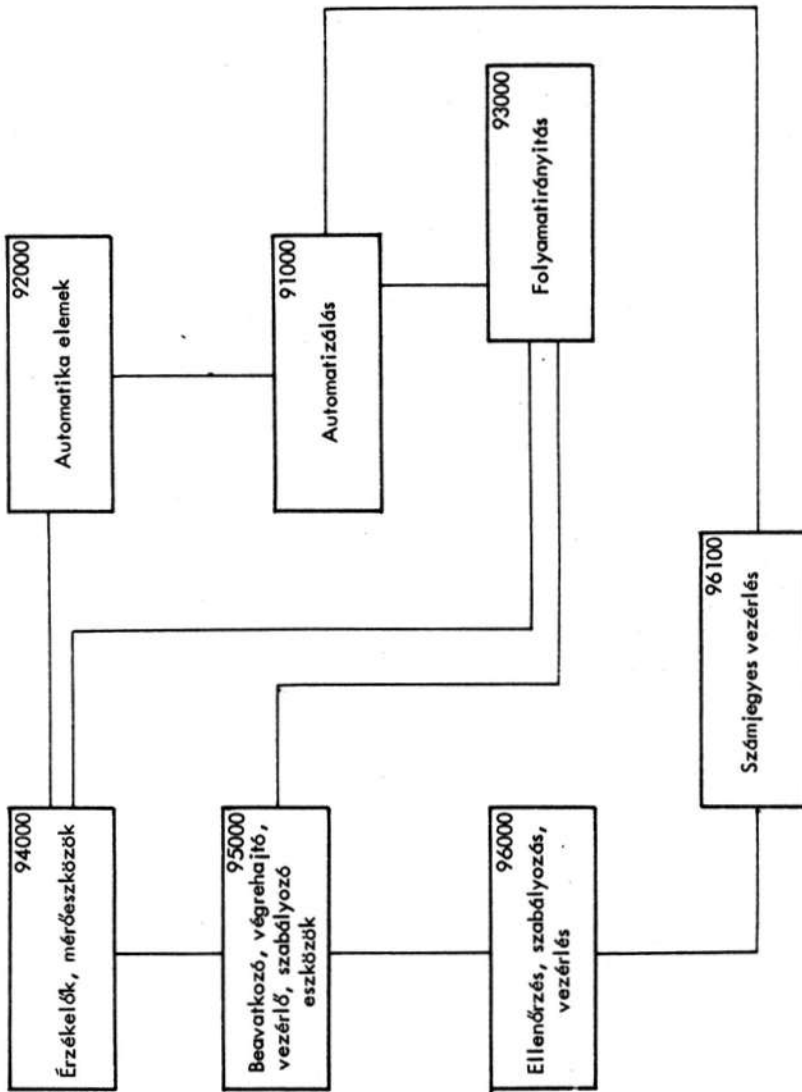
Egy nagy halmaz /például adatok halmaza/ áttekinthetősége általában a halmaz elemeinek csoportosításával /statistikai adatok esetén például a gyakorisági csoportok kialakításával/ könnyíthető meg. A tezausz szerkesztése során ezért az előválogatás után megmaradt kifejezések halmazát részhalmozokra bontottuk, a kifejezéseket egyelőre munkahipotézisként alkalmazott szakcsoportokba osztottuk be. Kilenc nagy szakcsoportot alakítottunk ki /néhány ezek közül: HARDWARE, SOFTWARE, AUTOMATIZÁLÁS, ALKALMAZÁSI TERÜLET stb./.

A kifejezéseket tartalmazó cédulákat a szakcsoportok szerint szétosztva lehetőség nyílt a szakcsoportok között kialakuló aránytalanságok megszüntetésére, valamint a kifejezések további, most már a szakcsoportosítás szempontjából történő válogatására, értelmezésére, módosítására.

Ennek érdekében minden egyes szakcsoporton belül alcsoportokat alakítottunk ki /például a HARDWARE szakcsoport alcsoportjai többek között: áramkör, tárolóegység stb./ és elkészítettük az alcsoportok közötti fogalmi kapcsolatokat bemutató hálózatos ábrákat /2.ábra/. Az egy-egy szakcsoportba tartozó kifejezéseket az alcsoportok szerint csoportosítottuk. Ha az előre kialakított alcsoportok a kifejezések ésszerű besorolását nem tették lehetővé, sor került mind az alcsoportok, mind a hálózatos ábra, esetleg a szakcsoportok módosítására. A kifejezéseknek az ábrákon, illetve a táblázatban elfoglalt helyére kódszám, az ún. szakcsoportjelző szám utal.

Ezzel a művelettel, valamint a kifejezésekre vonatkozó módosításoknak a betűrendes jegyzéken való átvezetésével elkészült az első munkaközi anyag. Ez tartalmazta

- a kb. 2500, a tezauszba deszkriptorként való felvételre javasolt kifejezésből és kb. 500 szinonim /non-deszkriptorként javasolt/ kifejezésből álló betűrendes jegyzéket:



2.ábra. Az AUTOMATIZÁLÁS szakcsoport ábrája a Számítástechnikai Fogalmak Teaurusza kéziratából

- az egyes szakcsoportokba tartozó alcsoportok közötti fogalmi kapcsolatokat bemutató hálózatos ábrákat;

- valamint a deszkriptoroknak a szakcsoportok /ezen belül: alcsoportok/ szerinti felsorolását táblázatos formában /3.ábra/.

E két utóbbi /az ábrák és a táblázatok/ tulajdonképpen a szakcsoportos index munkaközi anyagának tekinthető.

A számítástechnikával hivatásszerűen foglalkozó szakértői munkacsoport* a munkaközi anyagban minden egyes kifejezést megvizsgált abból a szempontból, hogy

- a kifejezés felvétele a tezauszba szükséges-e, és ha igen, az a megfelelő szakcsoportban található-e meg;

- a szinonimia problémákat az előszerkesztés során helyesen oldották-e meg;

- nem szükséges-e további fogalmak beiktatása;

- az alcsoportok közötti fogalmi kapcsolatokat az ábrák helyesen mutatják-e be?

A szakértői munkacsoport véleményezésének figyelembevételével újraszerveztett szókészlet, ábrák és táblázatok alapján állítottuk össze ismét csak munkaközi anyagként a hierarchikus és a permutált indexet.

A hierarchikus index összeállítása

Az elővizsgálatok és a szerkesztési munka során kiderült, hogy annak ellenére, hogy a fogalmakat tudományos követelményeket is kielégítő hierarchiába lehet rendezni, erre nem célszerű törekedni. A javasolt kifejezések ugyanis önmagukban még nem elégségesek egy teljes hierarchikus fa kiépítéséhez. Sok olyan fogalom felvételére lett volna szükség, amelyeknek csak szerkezeti funkciója van a tezauszban /kapcsolatot hoz létre egymástól több szintre elhelyezkedő fogalmak között/, de gyakorlati - indexelési, információkeresési - haszna

* A szakértői munkacsoport tagjai voltak: dr.KÁDÁR Iván,
a közgazdaságtudományok kandidátusa,
dr.EMETY Antal,
információfeldolgozási rendszer
szakértő,
MOLNÁR Imre, műszaki tanácsadó,
PÁL László, okl. villamosmérnök;
az OMKDK részéről:
dr.SCHIFF Ervin, okl.vegyszermérnök,
gazdasági mérnök.

...
Automatizálás

91000	automatika
91000	automatikus rendszer
91000	automatizálás
91000	automatizálás elmélet
91000	automatizált irányítási rendszer
91000	pneumatikus automatizálás
91000	szabályozáselmélet
91000	szerszámgép vezérlés
91000	viSSZacsatolt rendszer

Automatika elemek

92000	automatika-elem
92000	elektrohidraulikus automata-elem
92000	elektromechanikus automata-elem
92000	fluid elem
92000	fluid logikai elem
92000	fluid vezérlőelem
92000	fotoelektromos automata-elem

Folyamatirányítás

93000	áramlás szabályozás
93000	automatizált szállító rendszer
93000	digitális folyamatirányítás
93000	hengernsor vezérlés
93000	reaktor vezérlés
93000	vegyipari folyamatirányítás

Érzékelők, mérőeszközök

94000	adatreggztő mérlegelő berendezés
94000	digitális műszer
94000	érzékelőelem
94000	fluid érzékelő
94000	fluid mérőműszer
94000	letapogató berendezés
94000	mérőadó
94000	mérőátalakító
...	...

3.6bra. Minta a Számítástechnikai Fogalmak Tezaursza kéziratából: részlet az
AUTOMATIZÁLÁS szakcsoport táblázataiból

nincs. Az EJC, a TEST, a Thesaurifacet és az időközben megjelent Thesaurus of Computing Terms /angol nyelvű, számítástechnikai/ tezausz tanulmányozása a probléma következő megoldását ösztönözte:

- az alkalmazott hierarchia olyan laza legyen, hogy az ebben a rohamosan fejlődő ismeretágban felbukkanó új fogalmak, fogalmi kapcsolatok beillesztése ne járjon a teljes hierarchikus fa ismételt átszerkesztésének igényével;

- ezért nem célszerű az összes fogalmat egy egy-pontból kiinduló hierarchikus fába felülről lefelé haladva beépíteni, hanem inkább olyan rendezett szócsoportokat kell kialakítani, amelyeknek felfelé haladó kiépítése adott esetben, átlagban alulról a 3.-4. szint elérésekor abbamarad: a hierarchikus szócsoportok legmagasabb szintű fogalmait hierarchikusan ebben az elképzelésben már nem függnek össze egymással;

- egy adott, hierarchikusan rendezett szócsoporton belül nem célszerű arra törekedni, hogy a kifejezések között csak egyetlen egy típusu - nem-faj, vagy rész-egész - kapcsolat legyen.

A fenti metodikai megfontolást indokolta az is, hogy tezausz - a tárgykör természeténél fogva - tartalmazott igen részletes bontású területeket /a számítástechnikai fogalmakat/, és csak általános kifejezésekkel leírt szakterületet /az alkalmazási területet/.

A hierarchikus index /4. ábra/ előállítását a leírt alapelvek szerint a kifejezéseket tartalmazó cédulák /lyukkártyák/ ismételt átrendezésével történt.

A permutált index előállítása

A permutált index felépítésének alapelve az, hogy minden egyes kifejezés /deskriptor és non-deszkriptor/ minden egyes szakmailag értelmes, jelentéssel bíró szóösszetevője szerint betűrendbe rendezve jelenjen meg benne /5. ábra/.

Az index összeállításánál azt az első ízben az általános műszaki tezausz készítésénél kidolgozott, alkalmazott és kipróbált módszert alkalmaztuk, amely lehetőséget nyújtott a magyar nyelv egyes sajátosságai által okozott problémák megoldására.

Ismeretes, hogy az összetett szavak egybe-, illetve különírásának szabálya a magyarban a kérdés elsődlegesen mechanikus közelítésén túlmenően /amely a szótagszámot és az összetett szót alkotó szavak számát veszi figyelembe/ nagymértékben szubjektív módon kezelhető /attól függően, hogy egy-egy szóösszetételt állandónak, elterjedtnek stb. érzünk-e, vagy sem/. Ez indokolja azt, hogy a permutált indexek előállításánál a mechanikus - a szóközön alapuló - eljárások nem vezetnek sikerre. A tezausz által felölelt szakterület ismeretében lehet csak, szakértői munkával megjelölni az egyes, több szóból álló kifejezésekben azokat a szóösszetevőket, amelyek az adott

...	...
91000	automatizálás
91000	. automatika
96100	. . ipari automatika
96000	. . szabályozástechnika
91000	. automatikus rendszer
91000	. . automatizált irányítási rendszer
96000	. . . automatikus szabályozó rendszer
96000	. . . automatikus vezérlés
96100	 numerikus vezérlés
91000	 pneumatikus automatizálás
96000	. . . önellenőrző rendszer
91000	. . . visszacsatolt rendszer
91000	. automatizálás elmélet
91000	. . szabályozáselmélet
96000	. . . nyílt szabályozás
96000	. . . optimális szabályozás
95000	beavatkozó szerv
95000	. digitális beavatkozó szerv
95000	. elektrohidraulikus végrehajtó szerv
95000	. pneumatikus beavatkozó szerv
95000	. . pneumatikus elektronikus átalakító
26000	bérmunka számítóközpont
42100	. bérelt számítógépidő
24000	. . gépidő eladás
42100	. . használati díj
15400	beviteli egység
15400	. adatbeviteli rajzgép
15400	. input olvasó
15400	. . árcédula leolvasó berendezés
15400	. . bizonylatolvasó
15600	. . . többfunkciós bizonylatolvasó
15400	. . címkeolvasó
15400	. . digitális olvasó
15400	. . filmolvasó
15400	. . jelolvasó
15120	. . . jelfelismerő
15000	. . karakterolvasó
15000	. . . írógép-írás olvasó
15000	. . . kézírás olvasó
15000	. . . mágnesfestékes olvasó
15400	. . kétirányú olvasó
15000	. . kézi bemeneti egység
15000	. . . billentyűzet
15000	 szabványos írógépbillentyűzet
...	...

4.ábra. Minta a Számítástechnikai Fogalmak Teaurusza kéziratából: részlet a hierarchikus indexből

...
 automatizálás +
 automatizálás + levélosztályozás
 *automatizálás + navigáció
 automatizálás + pneumatikus
 automatizálás + rendelésfeldolgozás
 automatizálás + tervezés
 automatizálás elmélet +
 automatizálási szakember képzés +
 automatizált adatviSSzakeresés +
 automatizált fordítás +
 automatizált irányítási rendszer +
 automatizált levélcímzés +
 automatizált szállító rendszer +
 automatizált szedés +
 automatizált szerződésírás +
 autonóm működés +
 autópálya-csomópont tervezés +
 azonnali elérésű tároló +
 azonosító algoritmus +
 bankügylet +
 Barker-féle párhuzamos kód +
 *batch adatfeldolgozás +
 batch végállomás +
 Bayes-féle becslés +
 Bayes-féle döntési szabály +
 bányászat +
 *báziskonfiguráció +
 beavatkozó szerv +
 beavatkozó szerv + digitális
 beavatkozó szerv + pneumatikus
 becslés + adaptív
 becslés + állapot=
 becslés + Bayes-féle
 becslés + gradiens
 becslés + költség=
 becslés + megbízhatósági
 becslés + normaidő
 becslés + rekurzív
 becslés + szekvenciális
 becsléselmélet +
 *beépített tároló +
 beírási ciklusidő +
 belkereskedelem +
 belső kód +
 *belső órajel generátor +
 belső programozású asztali számológép +
 ...

5.ábra. Minta a Számítástechnikai Fogalmak Tezaurusza kéziratából: részlet a permutált indexből

összefüggésben és a tezausz használata szempontjából jelentősek. Így például nyilvánvaló, hogy

a számítástechnika kifejezésben a technika,

az ellenőrzés kifejezésben az őrzés

kiemelésének nincs értelme. De - igen általános jellege miatt - hasonlóan nem érdemes a rendszer szót kiemelni és ezáltal az összes olyan kifejezést, amelyben a rendszer szó előfordul a permutált indexben összehozni /van ugyanis gépi, program-, gazdasági, irányítási, kibernetikai stb. rendszer: ezek egy helyre csoportosítása célszerűtlen/.

Ezért minden egyes kifejezést megvizsgáltunk a szerkesztési munka során, hogy a kifejezés belsőjében előforduló szóösszetevők közül a szakmailag és tezausztechnikailag jelentőseket kiemeljük. A permutáció ezután a kiemelendő szóösszetevők figyelembevételével mechanikusan készült el. A permutált index végső formájában * jellel megjelölve tartalmazza a non-deszkriptorokat is.

A kifejezés eredeti alakjának a végén + jel van, míg az egybeírt, de a permutáláskor szétválasztott kifejezésekben a mesterséges szétválasztásra a permutált alak végén található = jel utal.

A második szakértői revízió

A kifejezések módosított betűrendes jegyzékét, szakcsoportos indexét, valamint a módosított kifejezésekben előállított hierarchikus és permutált indexet adtuk át a szakértői munkacsoportnak másodszori revízióra. Ezt a revíziót most már a kifejezések többszörös összefüggéseinek figyelembevételével, a tezausz teljes szerkezetében elfoglalt helyük és kapcsolataik vizsgálatával lehetett végrehajtani.

A kritikai észrevételek és részletes megjegyzések alapján a szerkesztőség létrehozta a fenti anyagok módosított változatát. Ennek birtokában lehetőség nyílt a tezausz főrészeinek összeállítására.

A tezausz főrésze és végső szövege

A tezausz főrésze /6. ábra/ végső formájában 2603 kifejezést tartalmaz betűrendben. Ezekből 2347 a deszkriptor és 256 a non-deszkriptor.

Minden egyes kifejezés, mint vezérkifejezés után egy un. szóbokor található. Ebben a szóbokorban soroljuk fel a vezérkifejezéssel valamilyen, a tezausz felépítésekor kimutatott kapcsolatban lévő kifejezéseket, a kapcsolat típusának megjelölésével együtt.

- 61000 ...
 - T identifikációs algoritmus
 - *G 61000 algoritmus
 - R 65000 adaptív identifikációs rendszer

- 58000 bankügylet
 - G 58000 pénzüintézet
 - *G 51000 termelési ág - szolgáltatás

- 53000 bányászat
 - *G 51000 termelési ág - szolgáltatás
 - R 54000 geológia

- 72000 Barker-féle párhuzamos kód
 - G 72000 kód
 - *G 72000 kódrendszer

- batch adatfeldolgozás
 - L 28000 csoportos adatfeldolgozás

- 15200 batch végállomás
 - *G 15200 végállomás

- 64000 Bayes-féle becslés
 - G 64000 becslésmélet
 - *G 64000 matematikai statisztika

- 62000 Bayes-féle döntési szabály
 - G 62000 döntésmélet
 - *G 62000 operációkutatás

- báziskonfiguráció
 - L 21200 alapkonfiguráció

- 95000 beavatkozó szerv
 - S 95000 digitális beavatkozó szerv
 - 95000 elektrohidraulikus végrehajtó szerv
 - 95000 pneumatikus beavatkozó szerv
 - ...

A szóbokrok két típusa létezik.

Az egyik típusban a vezérkifejezés egy non-deszkriptor, amelyről a helyette az indexelésben és az információkeresésben használatos deszkriptorra "L" betűvel jelölt utalás mutat.

A másik típusu szóbokor vezérkifejezése deszkriptor. Így ebben a szóbokorban "L" utalás nem lehet. Az egyéb bemutatott kapcsolatok azonban igen sokrétűek.

"T" betű jelzi azokat a szinonim kifejezéseket, amelyek alkalmazása tiltott /non-deszkriptor/, és amelyek helyett a vezérkifejezést kell használni. Az "S" és a "G" betűvel a hierarchikus kapcsolatokat /alacsonyabb és magasabb hierarchia szint/ jelöljük. A "G" a laza hierarchiában kialakított szócsoporthoz legmagasabb szintű kifejezését jelöli.

Az eddig ismertetett kapcsolatok már kidolgozott formában rendelkezésre állottak. Csak a cédulákat /lyukkártyákat/ kellett megfelelően kódolni és a szóbokrok kialakítására, kézzel összeválogatni. A teauruszban feltüntetünk azonban olyan kapcsolatokat is, amelyek nem tartoznak az eddig ismertettek közé, de bemutatásukra a teaurusz használhatósága szempontjából szükség van. Ezt a kapcsolattípust az "R" /rokonsági kapcsolat/ betűvel jelöltük és a teaurusz főrészének összeállításakor dolgoztuk ki, a már említett angol nyelvű számítástechnikai teaurusz segítségével.

E munka során néhányszor még megtörtént, hogy egy-egy kifejezést /vagy kapcsolatát/ módosítanunk kellett. A változtatást természetesen az összes indexen átvezettük.

A végső szöveget a változtatásoknak megfelelően módosítottuk lyukkártya sorozatok felhasználásával, automatikusan irattuk ki.

A számítástechnikai teaurusz
továbbfejlesztése

Általánosan elismert tény, hogy az osztályozási rendszerek, indexelési és információkeresési segédesszközök és így a teauruszok is, ha valóban

gyakorlati célok elérésére használják azokat, megfelelően rugalmasak kell hogy legyenek és tudniuk kell az ismeretek változásával lépést tartani.

A számítástechnikai teaurusz szerkezete erre lehetőséget nyújt. A lehetőséget azonban ki is kell használni. A teauruszokkal foglalkozó és már idézett UNESCO dokumentum is előírja az ismeretek fejlődésének folytonos nyomon követését a teauruszban is.

Ez a feladat azonban akkor hajtható csak végre, ha a számítástechnikai teaurusz gyakorlati alkalmazása során felmerült problémákat a teaurusz továbbfejlesztői megismerik. Kérjük azokat, akik a számítástechnikai teaurusz első kiadását kézbe veszik, vizsgálják

vagy alkalmazzák, hogy segítsék megjegyzéseikkel a továbbfejlesztés elengedhetetlenül szükséges munkáját.

E megjegyzések segítséget jelenthetnek az OMKDK-nak a más szakterületeken végzett tezauruszépítési munkálataiban is.



A CIKKBEN EMLITETT, VALAMINT A TEZAUROSZ ÖSSZEÁLLÍTÁSÁBAN
FELHASZNÁLT KIADVÁNYOK

Általános műszaki fogalmak tezaurusza. I-III. Bp. Orsz. Műszaki Könyvtár és Dok.Közp. 1970. 1113 p.

BÜRGER, E.: Datenverarbeitung. Technik-Wörterbuch. Berlin, Rechner Büromaschinen, 1969. 1463 p.

Dokumentáció az irányítástechnika, elektronika és határterületeik terminológiájához. I-III. Bp. Orsz. Műszaki Könyvtár és Dok Közp. 1971. 887 p.

Guidelines for the establishment and development of monolingual thesauri for information retrieval. Paris, UNESCO, 1971. 30 p.

Gyorsindex - Automatizálás. Bp. Orsz. Műszaki Könyvtár és Dok.Közp. 1970-1972.

Gyorsindex - Számítástechnika. Bp. Orsz. Műszaki Könyvtár és Dok.Közp., SZTI. 1970-1972.

MEEK, C.L.: Glossary of computing terminology. New York, 1972. 372 p.

NCC thesaurus of computing terms. NCC Publications. 4.ed. Manchester, National Computing Centre Ltd.inform.div.1972.

Subject headings for engineering. New York, 1970. 131 p.

Számítástechnikai fogalmak tezaurusza. Bp. Orsz. Műszaki Könyvtár és Dok.Közp. Megjelenik: 1973.

Számítástechnikai Tájékoztató Iroda Tárgyszójegyzéke. Kézirat.

Thesaurfacet. A thesaurus and faceted classification for engineering and related subjects. English Electric Company Ltd. 1969. 491 p.

Thesaurus of engineering and scientific terms. New York, United States Department of Defence. 1967. 689 p.

Thesaurus of Engineering Terms. New York, 1964-65. Engineers Joint Council.

Thesaurus Naturwissenschaft und Technik. ZIID, 1971.

Thesaurus Ökonomie. 1-2.Bd. Berlin, Ökonomische Forschungsinstitut der Staatlichen Plankommiss. 1971.



SCHIFF, E.: Thesaurus on computer techniques in Hungarian

On behalf of the State Committee of Technical Development the Hungarian Central Technical Library and Documentation Centre /OMKDK/ deals with problems of technical thesauri in Hungarian since 1968. As a result of this activity the Department for Development of OMKDK - in collaboration with outside experts and expert groups - compiled and published the "Általános Műszaki Fogalmak Tezaurusza" /Thesaurus of General Technical Terms/ in 1970. The "Számítástechnikai Fogalmak Tezaurusza" /Thesaurus on Computer Techniques/ was completed in 1973. Presently similar works are going on by preparing glossaries and thesauri in other subject fields connected with information processing and retrieval.

Aims and conditions of these thesauri were very different. In the first case

- a general technical thesaurus was made
- foreign examples were applied
- it was a pioneer work without previous experiences in Hungary
- it was very suitable to gain experiences
- as a basic thesaurus it opened ways for further thesauri built on or connected with it.

As a generalizing experience it could be stressed that thesauri reflect the structure of production, economy and sciences of the publishing country, region or organization. Therefore their translation is practically impossible. The differences in grammar, word-formation and concept-building are also increasing difficulties.

Experiences were gained in handling compound words very frequent in Hungarian. Methods for thesaurus-editing were also worked out with the following basic rules:

- gradual construction /identification of the basic vocabulary, step by step forming of various conceptual relationships, elaboration of indexes, building up the main part/;
- revision of the whole word stock during realization of the various steps, modification of terms and their relationships if necessary.

Based on these experiences the editing work of "Számítástechnikai Fogalmak Tezaurusza" /Thesaurus on Computer Techniques/ started in 1970. The aim was the compilation of a special thesaurus in Hungarian to be used in information work on computers, in information processing and retrieval.

Non-Hungarian materials were used only like auxiliary help. The first word-stock was compiled from terms of the Hungarian KWIC index

"Gyorsindex Számítástechnika - Automatizálás" /Quick Index - Computing+Automation/ where they have been used already in practice. Meeting points of the graphic displays represented sub-groups with more terms within the broader categories.

Revision of the word-stock and grouping by experts lead to modifications - was again examined by an expert group. After the critical control of the second expert group the final text of thesaurus was edited including:

- the main part of the thesaurus /with word-clusters illustrating concept relations/;
- subject group indexes /tables including graphical representations of descriptor networks and listing terms of sub-groups/;
- hierarchical index and
- permuted index.

Experiences of the work:

- orderless word-stock compiled could be well handled by forming graphic displays;
- in thesauri made for practical use /so necessarily flexible in structure/ rigid hierarchy is neither practical nor possible;
- permuted indexes are worth to be made only when permutation is not made mechanically /on spaces/.

The second part of the article - to be published later - will analyze the "Számítástechnikai Fogalmak Tezaurusza" /Thesaurus on Computer Techniques/.

*

ИИФ.3.: Тезаурус по вычислительной технике на венгерском языке

Венгерская Центральная Техническая Библиотека и Центр Научно-Технической Информации (ОМКДИК) по поручению Государственного Комитета по Развитию Техники с 1968 года занимается вопросами технических тезаурусов на венгерском языке. В результате этой деятельности Отдел Развития ОМКДИК - с помощью штатных экспертов и рабочих групп - построил и, в 1970 г., издал общий технический тезаурус (Általános Műszaki Fogalmak Tezaurusza). Тезаурус по вычислительной технике (Számítástechnikai fogalmak tezaurusza) создан в 1973 г. В настоящее время и в других областях ведутся работы с целью создания тезаурусов, применяемых в обработке и поиске информации.

Цель и условия создания вышеупомянутых тезаурусов в большой степени различны. В первом случае создан общий технический тезаурус по аналогии имеющихся тогда зарубежных тезаурусов, отчасти с целью приобретения опыта в области построения тезаурусов (поскольку опыты других непосредственно не принимаемые, см. ниже), и, отчасти с целью создания такого основного тезауруса, на базе которого можно построить ряд отраслевых тезаурусов, совместимых с исходным. Этот тезаурус не был предназначен для непосредственного использования в обработке информации.

Из числа приобретенных опытов можно выделить тот факт, что тезаурусы отражают производственные, экономические, научно-организационные и т.п. структуральные представления и подразделение выпускающих стран, областей или организаций и, по этой причине, их переводить невозможно. К этому еще добавляются различия отдельных языков в грамматике, строении слов и понятий.

Также были приобретены опыты в связи с проблемой сложных слов, довольно распространенных в венгерском языке. Выработана методика построения тезауруса, основными принципами которого являются следующие:

- тезаурус должен строиться поэтапно (сбор претендующих на участие слов, построение смысловых связей разного типа в нескольких этапах, создание указателей и, наконец, построение основной части);
- в конце каждого этапа необходимо снова просмотреть весь словарный состав тезауруса и, в случае необходимости, внести поправки.

Опираясь на эти опыты в 1970 г. начались работы построения тезауруса по вычислительной технике. Целью в этом случае являлось создание специального отраслевого тезауруса на венгерском языке, который мог быть результативно использован в практике информации по вычислительной технике, в обработке и поиске информации.

В ходе этих работ иностранные материалы были использованы только лишь в качестве вспомогательных. Первый вариант словарного состава составлен на основании двух Сигнальных указателей типа KWIC: Автоматизация и Вычислительная техника. Для группировки собранных выражений составлены диаграммы (сетевые схемы), указывающие смысловые связи между отдельными подгруппами выражений, находящимися на пересечениях линий схем.

Словарный состав и схемы были просмотрены экспертами и, учитывая высказанные замечания и предложения, в дальнейшем были составлены иерархический и пермутационный указатели. Этот материал снова был просмотрен экспертами, после чего появилась возможность построения основной части тезауруса. Готовый тезаурус состоит из четырех частей:

- основная часть тезауруса (вместе со словарными статьями, указывающими смысловые связи);
- группированный указатель (сетевые схемы и таблицы, содержащие выражения, попавшие в данную подгруппу);
- иерархический указатель;

- пермутационный указатель.

Из числа приобретенных опытов следует выделить следующие:

- собранный, но еще не группированный словарный состав станет удобным в смысле обращения, если построить соответствующие сетевые схемы:

- в таком тезаурусе, который предназначен для использования в практике и, следовательно, который должен быть гибким не целесообразно, и даже часто невозможно построить строгую иерархию;

- пермутационный указатель стоит составить только в том случае, если пермутация происходит не механически (по пробелам).

Вторая часть статьи, которая выходит тоже, будет заниматься анализом тезауруса по вычислительной технике.

*

SCHIFF, E.: Rechentechnischer Thesaurus in ungarischer Sprache, I.

OMKDK /Ungarische Technisch-Wissenschaftliche Zentralbibliothek und Dokumentationszentrum/ befasst sich im Auftrag des Regierungsausschusses für technische Entwicklung seit dem Jahr 1968 mit Fragen der technischen Thesauri. Als Ergebnis dieser Tätigkeit wurde durch die Abteilung für Entwicklung des OMKDK unter Mitwirken von Experten und Experten-teams die Publikation "Thesaurus allgemeiner technische Begriffe" in 1970 verfertigt und veröffentlicht. In 1973 wurde der "Thesaurus der rechentechnischen Begriffe" fertiggestellt. Gegenwärtig sind auch auf sonstigen Fachgebieten ähnliche Arbeiten im Gange, um Wörterverzeichnisse bzw. Thesauri zu erstellen, die in der Informationsverarbeitung und in der Informationsrecherche verwendbar sind.

Die Ziele und die Umstände der Zusammenstellung der erwähnten Thesauri waren sehr verschieden. Im ersten Fall wurde

- ein allgemeiner technischer Thesaurus,
- aufgrund ausländischer Beispiele,
- in ungarischer Sprache als bahnbrechende Arbeit ohne übernehmbare Erfahrungen,
- vor allem eben, um derartige Erfahrungen zu sammeln;
- nicht unmittelbar zur Verwendung in der Informationsverarbeitung;
- als Basisthesaurus, auf welchen aufbauend und an welchen anknüpfend Fachthesauri ausgebaut werden können, zusammengestellt.

Als eine zur Verallgemeinerung geeignete Tatsache ist hervorzuheben, dass Thesauri die strukturellen Vorstellungen bzw. Einteilungen betreffs der Produktion, Wirtschaft, Wissenschaftsorganisation usw. des zusammenstellenden Landes, Gebiets oder der Organisation widerspiegeln und dass sie demzufolge praktisch unmöglich zu übersetzen sind. Hiezu kommen noch die zwischen den einzelnen Sprachen bestehenden Unterschiede in Grammatik, Wortbildung sowie Begriffskonzipierung.

Wir haben in der Handhabung der in der ungarischen Sprache häufigen zusammengesetzten Wörter Erfahrungen sammeln können. Für die Thesauruserarbeitung haben wir eine auf folgenden Grundprinzipien beruhende Methode ausgearbeitet:

- Stufenweiser Aufbau /Schaffung des Basiswortschatzes, Ermittlung der Begriffsverbindungen verschiedener Typen in nacheinanderfolgenden Schritten und nötigenfalls Modifizierungen der Ausdrücke und ihrer Verbindungen/.

Mit der Nutzung dieser Erfahrungen wurde die Redigierung des "Thesaurus rechentechnischer Begriffe" in 1970 begonnen. In diesem Fall war das Ziel

- einen speziellen Fachthesaurus;
 - der in der Praxis der Informationsarbeit auf dem Gebiet der Rechentechnik;
 - bei der Handhabung und Recherche von Informationen
 - vereinbarungsmässig verwendbar ist,
- aufzubauen.

Bei dieser Arbeit wurde ausländisches /nicht in ungarischer Sprache verfasstes/ Material bloss als Hilfsmittel berücksichtigt. Für den Ausgangswortbestand wurden als praktischer Wortgebrauch die herausgestellten Ausdrücke aus den in ungarischer Sprache herausgegebenen KWIC-Index "Schnell-Index Automatisierung" und "Schnell-Index Rechentechnik" zugrunde genommen. Für die Gruppierung der Begriffe haben wir jeweils das die Begriffsverbindungen darstellende Fachgruppendiagramm /Netzdiagramm/ konstruiert. Die Knotenpunkte des Netzdiagramms repräsentieren die mehrere Ausdrücke zusammenfassenden Untergruppen der einzelnen Fachgruppen.

Die Überprüfung des Wortbestands und der Fachgruppenverteilung durch Experten führte zu Modifikationen.

Das vorschlagsgemäss modifizierte Material sowie die unter Berücksichtigung der Modifikationen zusammengestellten hierarchischen und permutierten Indexe wurden vom Experten-team abermals überprüft. Nach der zweiten Expertenkontrolle wurde der endgültige Text des Thesaurus in der folgenden Form fertiggestellt:

- Hauptteil des Thesaurus /mit den auch die Begriffsverbindungen darstellenden Wortbündeln, d.h. Gruppe der zum jeweiligen Des-

kriptor zugeordneten; mit ihm in semantischem Zusammenhang stehenden Deskriptoren:

- Fachgruppenindex /die Netzdiagramme und Tabellen der in die einzelnen Untergruppen gehörenden Ausdrücke/;
- hierarchischer Index und
- permutierter Index.

Von den gewonnenen Erfahrungen sind die folgenden hervorzuheben:

- das gesammelte, ungeordnete Wortmaterial wird durch die Aufstellung von Netzdiagrammen leicht behandelbar;
- bei den für praktische Verwendungen gefertigten /und demzufolge nötigerweise elastisch aufgebauten/ Thesauri ist es nicht möglich und auch nicht zweckmässig, eine strenge Hierarchie anzustreben;
- es ist bloss lohnend, einen permutierten Index zusammenzustellen, falls die Permutierung nicht mechanisch /auf das Spatium bezogen/ erfolgt.

Der später erscheinende 2. Teil der Arbeit befasst sich mit der Analyse des "Thesaurus rechentechnischer Begriffe".

A TEZAUROSZ KIFEJEZÉS A GÖRÖGBŐL SZÁRMAZIK ÉS EREDETILEG RAKTÁRT VAGY KINCSESKAMRÁT JELENTETT. MAI ÉRTELMÉBEN ELŐSZÖR AZ 1736 -OS KIADÁSÚ SHORTER OXFORD ENGLISH DICTIONARY-BEN SZEREPEL, A-HOL MÁR "...AZ ISMERETEK KINCSESHÁZA VAGY TÁRHÁZA, MINT A SZÓTÁR, ENCIKLOPÉDIA, VAGY EHHEZ HASONLÓ..."