

ORSZÁGOS SZÉCHÉNYI KÖNYVTÁR
KÖNYVTÁRTUDOMÁNYI ÉS MÓDSZERTANI KÖZPONT

KOHÓ- ÉS GÉPIPARI TUDOMÁNYOS MŰSZAKI TÁJÉKOZTATÓ INTÉZET

HORVÁTH TIBOR – VARGA DÉNES
INFORMÁCIÓS TEZSAURUSZOK

OSZK
Országos Széchényi Könyvtár

Népművelési Propaganda Iroda
Budapest 1977

ORSZÁGOS SZÉCHÉNYI KÖNYVTÁR
KÖNYVTÁRTUDOMÁNYI ÉS MÓDSZERTANI KÖZPONT

KOHÓ- ÉS GÉPIPARI TUDOMÁNYOS MŰSZAKI TÁJÉKOZTATÓ INTÉZET

HORVÁTH TIBOR – VARGA DÉNES
INFORMÁCIÓS TEZAUROSZOK

OSZK
Országos Széchényi Könyvtár

Népművelési Propaganda Iroda
Budapest 1977

Lektorálták: Horváth Magda
Patek Ferenc
Vajda Erik



ISBN 963 201 049 3 – KMK
ISBN 963 562 283 X – NPI

Készült a Népművelési Propaganda Iroda nyomdájában
Felelős kiadó: Nemes Iván
Felelős vezető: Vymeták Ferenc
Budapest 6534–77

**„Azért dolgoztam ennyit ezen a rajzon,
hogy figyelmeztessem barátaimat egy
veszélyre, amelyet nem ismernek, noha
régóta fogva ott leselkedik a sarkukban,
akárcsak az enyémben.”**

(Antoine de Saint-Exupéry)

OSZK

Országos Széchényi Könyvtár

TARTALOM

Előszó	6
1. A tezaurusz fogalma, célja, feladatai	8
1.1 A tezaurusz fogalma	8
1.2 A tezaurusz célja	12
1.21 Osztályozás	12
1.22 Tárolás	14
1.23 A keresőtémák leírása	15
1.24 Egyéb funkciók	15
1.3 A tezauruszok általános szerkezete	16
1.31 A tezaurusz főrésze	17
1.32 A tezauruszok hierarchikus része	20
1.33 A tezauruszok grafikus része	22
1.34 A tezauruszok kiegészítő részei	23
2. Szógyűjtési módszerek	25
2.1 Általában	25
2.2 Szóstatistikai módszer	26
2.3 Tapasztalati módszer	28
2.4 Meglevő osztályozási rendszerek és szókészletek felhasználása	30
2.5 Elméleti módszer	33
2.6 Módszerek kombinációja	35
3. Relációk	37
3.1 Általában	37
3.2 Relációelméleti alapfogalmak	38
3.3 Tezauruszrelációk	44
3.31 Nem-faj viszony	44
3.32 Egész-rész viszony	46
3.33 Rokon kapcsolatok	48
3.34 Szinonimák és kváziszinonimák. Homonimia	52
3.35 Külső tapasztalatok	54
3.4 A deskriptor szóalakja	54
3.5 Fazetták (vegyületek)	56
4. A tezaurszkészítés menete	57
4.1 A célok meghatározása és a tervezés szakasza	58
4.2 Az induló szóanyag összegyűjtése	59
4.3 A szóanyag elemzése és a legfontosabb deskriptorok kiválasztása	60
4.4 A legfontosabb deskriptorok relációinak gráfszerű kifejtése	61
4.5 A deskriptorcikkek kialakítása	63
4.6 A tezaurusz szerkesztése	65
4.7 A tezaurusz hierarchikus részének szerkesztése	68
4.8 Ellenőrzés, modellkísérletek	69
Bibliográfia	73

ELOSZO

Ennek a könyvnek, amelyet az olvasó kezében tart, a kézírata a jelenlegi formában már 1975-ben készen állt. Két intézet közös erőfeszítése és együttműködése alapján született: a Kohó- és Gépipari Tudományos Tájékoztató Intézet gyakorlati tezaszuszépítő feladatok előtt állt, az OSZK Könyvtartudományi és Módszertani Központja pedig az elméleti problémák tisztázását tűzte ki feladatául. Szerencsés találkozás volt, hogy a két intézet feladataikat szorosan együttműködve oldhatták meg, mindkét fél gazdagodásával.

Az olvasó észreveheti, hogy a könyv mondandója teljes összhangban áll a magyar nyelvű tezaszuszok szabványával, nemcsak terminológiájában és jelöléseiben, hanem koncepciójában is. Ez annak köszönhető, hogy a magyar nyelvű tezaszuszszabvány ugyan-ezen munkálatokból nőtt ki, a szabvány kidolgozó személyükben is ugyanazok a szakemberek voltak, akik a két intézet együttműködésében a tezaszuszok fejlesztési-kutatási feladataiban vettek részt.

Külön említésre méltó az a hatékony támogatás, amelyet az MTA informatikai és matematikai-nyelvészeti bizottsága nyújtott a tezaszusz kutatás számára. Ez a téma a bizottság érdeklődésének középpontjában állt, a problémákról ankétsorozatot is rendezett.

A könyv koncepciójának néhány sarkalatos pontjára itt hívjuk fel az olvasó figyelmét.

A szerzők mindenképpen igyekeztek elkerülni a tezaszusz kérdésség leegyszerűsítését — sajnos, ilyen megalkuvásokkal találkozni lehet a hazai szakirodalomban is. Igénytelen tezaszusz nem biztosítja hatékonyabban a szakirodalom feltárását a hagyományos eljárásoknál, tehát nincs is szükség rá. Másfelől a túlságosan bonyolult, a feleslegesen szaporított relációk sokaságával felépített tezaszuszok nehézkessé tehetik a visszakereső rendszerek működését. Az optimális tezaszusz e két szélsőség veszélyeit kerüli el, és a jelen könyv ezeket a megoldásokat igyekezett felmutatni.

Elméleti alapvetésre törekedtek a szerzők, de mindenütt érezni lehet a gyakorlat kontrollját is: lényeges pontokon azon az alapon foglaltak állást, hogy az adott megoldásnak milyen kihatásai vannak a visszakereső rendszer működésére, a szakirodalmi szolgáltatásokra.

Lényeges vonása a koncepciónak, hogy kerüli a formai megoldásokat. A kapcsolatokat — számos munka vagy akár szabvány álláspontjával szemben — az objektumok között értelmezi, illetve a fogalmak síkján, nem pedig nyelvi kifejezéseik között. A defi-

níciókat szakmai problémákból vezeti le és fejti ki, de nem bocsátkozik a meghatározások területén meddő vitákba.

Végül a könyvet szakmai mérséklet jellemzi, több tekintetben is. Egyrészt a szerzők nem állítják, hogy az indexelő-osztályozó feladatok egyetlen üdvöztető megoldása a tezauszokon nyugszik. Csak abban foglalnak állást, hogy ha a tezauszok mellett születik döntés, akkor annak milyen követelményei vannak. Másrészt korántsem kívántak ideális megoldást, mindenhol, mindenre alkalmas elméletet kidolgozni. Ellenkezőleg, hangsúlyozzák, hogy az általános elvek konkrét funkciók alapján telnek meg élettel, s a jó tezausz mindig adott feladat számára készül.

Budapest, 1977. március

A kiadók

OSZK

Országos Széchényi Könyvtár

„Sohasem definíciók alapján ismerünk meg egy dolgot. Mielőtt eljutnánk a valódi megértéshez, tapasztalatokra kell szert tennünk az alapfogalmakkal kapcsolatosan, míg dolgozunk velük és vizsgáljuk őket.”

(Lánczos Kornél)

1. A TEZAUROSZ FOGALMA, CÉLJA, FELADATAI

1.1 A TEZAUROSZ FOGALMA

Századunk első éveitől kezdve a dokumentációs tevékenységet – a szót első ízben Paul OTLET használta 1905-ben a nemzetközi közgazdasági konferencián – a szétsugárzó szolgáltatások egyedurialma jellemezte. A cél az újdonságértésítés volt: rendszeresen és folyamatosan tájékoztatni a felhasználókat az új szakirodalomról, annak tartalmáról – lett légyen a szakirodalom az adott esetben folyóirat-közlemény, szabadalom, könyv vagy bármi más. Az új információk birtokába a szakirodalom rendszeres figyelemmel kísérésevel lehetett eljutni.

A dokumentációs tevékenység abban nyújtott ehhez segítséget, hogy felhívta a figyelmet a különböző helyeken szétszórta – esetleg különböző nyelveken – megjelenő irodalomra, elősegítve ezzel a tájékozódást a kutatási témákban, sőt bizonyos fokig értékelést is adott a szakirodalomról.

A tezauszproblematika akkor vált a szakirodalmi tájékoztatás egyik fő kérdésévé, amikor az információk mennyiségi növekedése nyomán a szétsugárzó szolgáltatások primátusát megtépázta a visszakereső tájékoztatás előtérbe kerülése. Valóban, a szétsugárzó tájékoztatás nehezen birkózott meg két feladattal. Az elsőt úgy lehet megfogalmazni, hogy a folyamatos tájékoztatás kicsi valószínűséggel kecsegtet arra vonatkozóan, hogy az igény egy konkrét információval kapcsolatban ugyanakkor merül fel, amikor azt a folyamatos tájékoztatásban éppen közléstesszik. A másik probléma az volt, hogy az akkori szétsugárzó tájékoztatások többé-kevésbé a tudományok aktuális tartalmából indultak ki, leggyakrabban tehát tudományorientáltak voltak, szemben a ma mind gyakrabban hangoztatott és hangsúlyozott felhasználóorientáltsággal.

Mindkét probléma a szétsugárzó tájékoztatás idő szerinti rendezettségéből fakad. Ha bizonyos késedelemmel is, de a dokumentumok megjelenési ideje szerint jelennek meg a róluk szóló tájékoztatások. (A regisztratív tájékoztatások természetesen hamarabb, az értékelő tájékoztatások nagyobb késéssel.) A gondolatok megéréséhez azonban idő kell, a gyakorlati alkalmazásokról nem is szólva. A szétsugárzó tájékoztatás mellett – főleg

túlságosan bőséges választék esetén — régen feledésbe merülhetnek olyan dokumentumok, amelyek tematikusan egy később beérő gondolat kifejlődését kiszítették elő. Az *időláncra való felfűzés mellett tehát a tematikai hasonlóság szempontja került előtérbe*. Természetesen nemcsak a közvetlen hatások, közvetlen előzmények előbányászásáról van szó — ez tisztességes szerzők esetén a hivatkozásokból viszonylag könnyen elvégezhető és nem is mindig fontos —, hanem a látszólag véletlen tematikai egyezésekről, *ugyanannak a problémakörnek más irányból való megközelítéséről*, módszertani analógiákról stb. A tudomány és gyakorlat kacsaringói pedig, minél közelebről, minél részletesebben nézzük, annál inkább fittyet hánynak a kronologikus sorrendnek. Nem segít tehát a problémán az időrenden *belül* való tematikai bontás sem. Minél részletesebb a bontás, annál ritkábban kerül egy-egy kis részskatulyába valami dokumentum, ha pedig nem részletes a bontás, akkor a tematikai sokféleségben éppúgy elveszhet a kutató, mint a dokumentumok özönében.

Ezt a dilemmát a szétszaggató szolgáltatások valóban nem tudják megoldani, egyszerűen azért, mert nem erre valók. Ilyen értelemben hiányos volt a dokumentációs tájékoztatás — teljesen ráhagyta a felhasználóra a finomabb tematikus válogatást. A maga feladatát viszont kitűnően ellátta továbbra is, azt, amit viszont a visszakereső tájékoztatás vagy sokkal drágábban, vagy sokkal nagyobb késéssel, vagy egyáltalán nem tudna elvégezni.

A dokumentációs tevékenység feladatának megítélésében tehát lényeges szemléletváltozás állt be. Ez a változás azonban korántsem azt jelenti, hogy ma a visszakereső tájékoztatás lép egyoldalúan a folyamatos helyére, vagy hogy a felhasználói orientáció teljesen felváltja a diszciplinaorientáltságot. Az történt, hogy a vagylagos megoldások helyébe a korszerű, komplex rendszerszemlélet lépett. Ez a szemlélet *ugyanazon szolgáltatási rendszer különböző oldalainak fogja fel a különböző szolgáltatási típusokat*, hangsúlyozva, hogy mindegyikre szükség van, ha azok különböző feladatok, funkciók betöltésére valók. A mai követelményeknek nem képes megfelelni a szakirodalmi tájékoztatás a régi, többé-kevésbé egységes szolgáltatási típussal, hanem csak a szolgáltatások egész rendszerével, ahol a rendszer egymásra épülő, egymást kiegészítő, egymást kölcsönösen feltételező részrendszerekből áll. A korszerű szolgáltatási rendszer alapja a *rendszeres, folyamatos irodalomfigyelés, osztályozás és indexelés* — egyelőre szinte kizárólag emberi kapacitással, megfelelő szervezéssel biztosítva a viszonylagos teljességet és a fedések minimumra való korlátozását.*

Ez az irodalomfigyelés hazai viszonylatban elsősorban a szakma reprezentáns folyóiratcikkeinek, dokumentumainak bibliográfiai adataira, a ma már egyre inkább a

*A későbbiekben itt is egyre nagyobb szerepe lesz az ember-gép kapcsolatra épülő dokumentációs előkészítő tevékenységnek, ahol az előzetes szelektálást és csoportosítást már a számítógép végzi, mindig fejlettebb módszerekkel, de soha nem hibátlanul. Az automatikus indexelés jórészt még a jövő feladata, ha nem is olyan nagyon távoli jövőé. A ma készülő teauruszoknak a számítástechnika rohamos növekedése miatt ezzel a lehetőséggel is számolniuk kell. Az ember közreműködésével folyó ún. interaktív keresés pedig már a jelen problémája, akárcsak a több nyelvű bemenettel rendelkező kereső rendszerek. A technikai lehetőségek mindehhez adva vannak, a gazdaságosság is egyre kevésbé probléma, a fő akadály a jelenleg működő rendszerek teauruszainak fejletlensége és a régen túlhaladott technikai lehetőségekhez való túlzott ragaszkodás egyes dokumentációs szakemberek részéről.

dokumentumokkal együtt megadott kulcsszavakra és témaközlő összefoglalásokra, valamint a referáló folyóiratok adataira kell hogy támaszkodjék — kiegészítve természetesen a hazai szakirodalom hasonló természetű feldolgozásaival. (Egyes szakterületeken kitűnő nyersanyagot szolgáltatnak a nemzetközi szakirodalomról a rendszeres mágnesszalagos szolgáltatások.)

Erre az alapra mint dokumentumbázisra épülhet a korszerű gépesített tájékoztatás, a különböző rendeltetésű és szintű visszakereső rendszerek éppúgy, mint a különböző mélységű és átfutási idejű szétsugárzó szolgáltatások. Ezek között is a két alaptípus: a tudományos eredményeket rendszeresen közlő diszciplinára orientált és a konkrét felhasználói igényeket figyelembe vevő másik, témafigyelő jellegű szolgáltatási típus.

Amennyiben a korszerű komplex gépesítés válik alapjává a szolgáltatási rendszernek, azonnal felmerül egy újabb kérdéscsoport: nevezetesen a szakirodalmi adatok tárolásának kérdései. A technikai problémák ennek a kérdéskomplexumnak csak egyik oldalát jelentik — ezt megelőzően mindenekelőtt arra kell választ adni, hogy milyen szempontok alapján történjen a tárolás. Melyek azok a paraméterek, szempontok, jellemző jegyek, amelyeket figyelembe kell venni, vagy éppen a szakirodalomból kell kihámozni ahhoz, hogy ezek alapján az irodalom a kívánalmaknak megfelelően visszakereshető legyen. Természetesen ezek a szempontok sem túl bonyolultak, sem túlságosan nagyszámúak nem lehetnek, de nem lehetnek önellentmondóak, következetlenek sem, hiszen nyilvánvaló, hogy ezek a szempontok egymással is kapcsolatban állnak. És bármennyire igyekszünk is általános követelményrendszer kidolgozására, tudatában kell lennünk annak is, hogy a kidolgozandó struktúra függ az adott ismeretág (tudományág) jellegétől és a felhasználói igényektől is.

A teauruszok azt a feladatot tűzik ki célul, hogy szabályozzák, rendszerezék a fenti szempontokat, a rendszer belső logikája és a gazdaságosság mérlegelésével limitálják a szempontok számát, kiemelik a legfontosabbakat és megállapítsák ezek összefüggéseit. Mindezek alapján megfelelő eszközt, nyelvet kell definiálniuk, melyek lehetővé teszik a szakirodalmi közlemények — vagy általában az információk —, valamint az információs igények leírását, a kettő egybevetését, s ennek alapján az információs igények kellő pontosságú és mélységű kielégítését. Ennek megvalósítása a természetes nyelvek szabályainál szigorúbb és egyöntetűbb, könnyebben formalizálható és könnyebben kezelhető rendszerek kialakítását kívánja meg.

Ahol lehet, a teauruszok a szaknyelvek kifejezésanyagára épülnek, így a természetes nyelvvel is megtartják a kapcsolatukat, *elemi egységnek azonban nem a szavakat, hanem a fogalmakat, illetőleg az ennek megfelelő nyelvi kifejezésformákat tekintik*, függetlenül attól, hogy az egyszerű vagy összetett szó vagy pedig kifejezés. A lényeg a fogalom és a kifejezésforma *kölcsönösen egyértelmű* megfeleltetésén van. A nem egyértelmű nyelvi kifejezések csak mint *utalók*, a teaurusz szóanyagának perifériáján szerepelhetnek, de sem a dokumentumok tartalmának, sem az igényeknek a megfogalmazásában nem szerepelhetnek, és így természetesen a kereséssel kapcsolatos további műveletekben sem vehetnek részt. Utalóként való felvételüket éppen az indokolja, hogy a kifejezés többértelműségére időben felhívják a felhasználó (indexelő vagy információt igénylő) ember figyelmét, s döntésre bírja, hogy a felkínált — most már egyértelmű — lehetőségek

közül válasszon. Az utalások között tehát *vagylagosak* is szerepelhetnek. Másutt — éppen a rendszer aktív részének tehermentesítése és redukálása érdekében — az összetettebb fogalmak *felbontására* ad utasítást az utalás (a — lásd b és c és . . .), ahol $b, c, \dots a$ -nál egyszerűbb, elemibb fogalmakat jelöl, amelyek *együttes* alkalmazása — az adott rendszer keretein belül — egyértelmű a alkalmazásával. Ez a módszer természetesen csak addig a pontig alkalmazható, míg a visszakeresés egyértelműségét komolyan nem veszélyezteti, amíg tehát az a tulajdonsággal jellemzett dokumentumok halmaza kifejezhető a b, c, \dots tulajdonságokkal jellemezhető dokumentumhalmazok metszeteként:

$$D(a) = D(b) \cap D(c) \cap \dots$$

Így tehát a tezauruszban szereplő *aktív* fogalmak mindegyike egy-egy szempont, jellemző jegy, amelyek segítségével a dokumentumok leírhatók a kívánt — tárolásra és visszakeresésre egyaránt alkalmas — formában. A természetes nyelvtől ezek a kifejezések tehát elsősorban abban különböznek, hogy szabályozottak, mondhatnók szabványosítottak maguk a kifejezések, valamint az aktív és passzív kifejezésanyag közötti határvonal is.

Ha egy-egy kifejezés, fogalom ilyen formában szabályozásra és meghatározásra kerül a tezaurusz számára, akkor már nem is fogalomnak vagy kifejezésnek hívjuk, hanem *deskriptornak*, jelezve azt, hogy már az információk, a dokumentumok, vagy általában nyilvántartásban szereplő objektumok szabatos leírására való kifejezésekről van szó.

De ez még nem minden. A kifejezésekbe a szakember kimondatlanul is, implicit módon *beleérti* a kifejezéssel jelölt fogalom szerteágazó, sokrétű kapcsolatát más fogalmakkal. A gép (vagy a más módon realizált gépies kereső rendszer) azonban nem értheti bele azt, amit kifejezetten nem közölnek vele. Nem elég tehát csak a kifejezésanyagot szabályozni — szabályozni és implicit módon tárolni kell az aktív kifejezőkészlet elemei között fennálló kapcsolatokat, *relációkat* is, olyan részletességgel, amely a rendszerrel szemben támasztott követelményeknek megfelel.

Összefoglalva az elmondottakat: az információs tezauruszokat úgy definiálhatjuk, mint a szakmai ismeretek egy-egy ágához tartozó fogalmak szabályozott kifejezéseinek — deskriptorainak — és az adott információs rendszer keretein belül nélkülözhető szakmai kifejezéseknek — a nem deskriptoroknak — egy olyan bővíthető, de az ismeretág adott fejlettségi fokán tematikailag teljesnek tekinthető gyűjteményét, amely explicit módon tartalmazza e fogalmak legfontosabb összefüggéstípusait is, az egyes deskriptorokhoz rendelt relációmutatók, illetőleg a nondeskriptorokhoz rendelt utaló kifejezések formájában.

A magyar nyelvű információkereső tezauruszok szabványának (MSz 3418) a fentivel azonos szellemű meghatározása szerint a *tezaurusz* „természetes nyelven kifejezett fogalmak olyan szabályozott, szükség szerint változtatható szótára, amely feltünteti a fogalmak közötti legfontosabb összefüggéseket. Fő rendeltetése az információk feldolgozása és keresése.”

1.2 A TEZAUROSZOK CÉLJA

Az információs tezauszoknak sokirányú feladatkört kell ellátnia. E feladatok közül hármat szükséges kiemelni.

1.21 OSZTÁLYOZÁS

A tezausz első feladata biztosítani a dokumentumok osztályozását és indexelését. Klasszikus értelemben a dokumentumok osztályozásán osztályba sorolást értenek, azaz az azonos vagy majdnem azonos tartalmú dokumentumok besorolását valamely fogalomosztályba — egy olyan általánosabb fogalom körébe, amely elég híven kifejezi a dokumentumok tartalmát. Az osztályok általában előre adottak és szükségszerűen átfogóbbak, mint a dokumentumok konkrét tartalma. Ha nem így lenne, minden dokumentum önmagában képviselne egy osztályt, s az egész folyamat értelmetlen lenne. Másfelől a dokumentumok tartalma rendszerint nem jellemezhető egyetlen fogalommal, és így az esetek többségében minden dokumentum több osztályba sorolható.

A mai követelményeknek az osztályozási tevékenység tartalma a fenti — kissé egyszerűsített — megfogalmazásban korántsem felel meg. Az osztályozásnak ma már hármas feladatot kell ellátnia:

- a) osztályba sorolást a fenti értelemben,
- b) a dokumentumok egyedi információinak megfelelő formai előírások szerint való kifejezését,
- c) a fenti ismérvek alapján annak bizonyítását, hogy a dokumentumok esetenként kialakított halmazai elkülöníthetők legyenek különböző szempontok vagy szempontcsoportok szerint, vagy más műveleteket lehessen elvégezni, továbbá lehetővé kell tennie változó szempontok szerint rendezett halmazok kialakítását.

A második követelmény bizonyos értelemben ellentmond az elsőnek. Míg az osztályba sorolás a dokumentumra mint egészre, tartalmi egészre vonatkozik és egy tárgyabb jelentésű, magasabb szintű fogalom segítségével mehet végbe, addig az egyedi információk kifejezése éppen ellenkezőleg, a dokumentumok mikrostruktúrájának kifejezésére vonatkozik. Példával érzékeltetve: ha egy monográfia a gráfelmélettel foglalkozik, akkor besorolható a „gráfok” vagy a „gráfelmélet” osztályba. Ezzel eleget tettünk az első követelményeknek, de elkendőztünk egész sor konkrét információt, pl. hogy ad-e halmazelméleti bevezetést vagy sem, foglalkozik-e a gráfok alkalmazásával a hálós tervezésben stb. Az utóbbi szempontokra már csak egy mélyebb osztályozás ad választ, erre azonban nem lehet előre elkészített, dokumentumra szabott osztályozó rendszert adni. Ennek a *hajlékonyságnak*, flexibilitásnak a követelményét fejezi ki éppen a második pont.

Az osztályozás harmadik funkciója a visszakereshetőség szempontjának érvényesítését jelenti már az osztályozás és indexelés fázisában. Az osztályba sorolás — azzal, hogy átfogóbb, generikusabb fogalmak segítségével alakít ki csoportokat a dokumentumok

között, önmagában hordja a visszakereshetőség kritériumát, legalábbis az adott osztályozás logikája szerint. Az adott esetben csak arra kell ügyelnünk, hogy a dokumentumot abba az osztályba soroljuk be, ahol várhatóan keresni fogják. Bevált módszer a hagyományos könyvtári osztályozásban is, hogy a dokumentumot inkább egy, esetleg két hierarchiaszinttel is magasabbra sorolják be, eleve generikusabb fogalommal, mint amilyen a dokumentum tartalma. Ez a módszer itt is alkalmazható, mivel a visszakeresés teljességét a zaj lényeges növekedése nélkül emeli.

A visszakereshetőség biztosítása az egyedi jellemzők kifejezése alapján már bonyolultabb feladatot jelent. Világosan elkülönített ekvivalenciaosztályok helyett itt részben egymást fedő fogalmakkal, ún. toleranciaosztályokkal kell dolgoznunk. A halmazok megfelelő részhalmazainak kiválasztását redundáns indexeléssel lehet csak biztosítani. Az osztályozótól azonban nem várható el, hogy esetenként izzadjon ki a lehetséges keresési szempontokat — megfelelő választékokat és választási lehetőséget kell számára biztosítani a rokon fogalmak közül, hogy az indexképet kellőképpen kiszélesíthesse a legfontosabb és legáltalább kifejezések kiválasztásával. A keresőkép kiszélesítése automatikusan is végbemehet a rendszerben (bizonyos feltételtől függően), akkor azonban válogatás nélkül minden irányban megtörténik a kiszélesítés, a zaj tehát eleve nagyobb lesz.

A kereső rendszer lehetőségei nagymértékben függenek a tezausz szerkezeti kimunkáltságától. Minél jobb a tezausz, annál tökéletesebb kereső rendszert lehet rá építeni. Ezzel a lehetőséggel azonban élni is kell — tudni kell a tezauszt jól használni már az indexelés során. Soha nem szabad elfelejteni, hogy minden egyes ísmérv egy-egy dokumentumhalmazt határoz meg, s ennek sem túlságosan szűknek, sem túlságosan bőnek nem szabad lennie.

Ahol pontos fogalmak vannak, ott nincs különös probléma. A négyzet, a téglalap, a rombusz, a trapéz, a deltoid például egyértelműen besorolható a magasabb osztályt képviselő „négyszögek” alá, vagy még átfogóbban a „síkidomok” alá. A kereső rendszernek természetesen olyan kérdésekre is fel kell készülnie, hogy mely síkidomok szimmetrikusak, melyek jellemezhetők a derékszöggel, az oldalak egyenlőségével stb. A kereső rendszerek azonban általában nem ilyen világosan elkülöníthető kapcsolatokkal dolgoznak — ha így volna, akkor nem is volna helye egyedi mérlegelésnek, az egész probléma csak a tezausz készítőinek jelentene gondot.

Az osztályozás fenti feladatai magukban foglalják azt is, hogy a tezauszokban szereplő deskriptorok segítségével a kiválasztott dokumentumok a deskriptorok alapján rendezhetők is. A rendezés történhet bármely olyan szempont szerint, amellyel előzőleg számolt a tezausz, tehát lehet mechanikus, pl. a deskriptorok szoros betűrendjét követve, de nyugodhat a deskriptorok közötti tartalmi kapcsolatokon, akár előre megadott szempontok szerint, akár a keresőkép fokozatos automatikus készítése alapján.

Az osztályozási feladatkör ilyen szélesebb értelmezésének a klasszikus osztályozási rendszerek nem tudnak megfelelni. A tezauszok a több szempontú osztályozás számítógépes realizálásával minőségi ugrást okoztak az osztályozásméletben is, nemcsak a gyakorlati tájékoztatásban. A tartalékok pedig az elért eredményeknél is nagyobbak, éppen az osztályozás és indexelés automatizálása területén.

1.22 TÁROLÁS

A dokumentumok csak abban az esetben kereshetők vissza, ha biztosítva van a megfeleltetés az osztályozás ismérvei és a tárolt dokumentumok között.

A nyilvántartó rendszer, amely a visszakeresés eszköze, ezen a ponton rendszerint elválik a dokumentumtól. (Kivétel néhány mikrofilmes visszakereső rendszer, amelyeknél a dokumentumok teljes szövege és az ismérveket tartalmazó ún. kódmező együttesen kerül fotózásra, ennek következtében a dokumentumszöveg együtt marad a deskriptorokkal.) A dokumentumtárat és a visszakereső apparátust valamilyen számrendszer, hivatkozási jelzetek, raktári szám stb. köti össze. A tárolás technikai problémáival itt nem kívánunk foglalkozni, funkcionálisan mégis szükségesnek tartjuk a feladat tisztázását.

A tároló rendszerrel szemben kettős követelmény támasztható: egyrészt biztosítania kell a dokumentumokra vonatkozó információk bevitelét és állandó bővülését, másrészt magát a visszakeresést. Ehhez azonban meg kell tudnia állapítani, hogy egy dokumentumnak melyek az ismérvei, valamint választ kell adnia arra a kérdésre, hogy egy adott ismért mely dokumentumra vonatkozik.

A feladat nehézkesen (vagy túlságosan drágán) oldható meg egyetlen file-lal, mivel a megközelítés módja pontosan fordított a két esetben. Az első esetben úgy gazdaságos, ha együtt szerepel mindaz, ami egy dokumentumra vonatkozik: azonosítók, bibliográfiai adatok, deskriptorok együtt képviselnek egy tárolandó egységet (rekordot). Így hosszú leírások keletkeznek, amelyek kezelése nehézkes. Egyetlen szempont szerinti visszakeresés így vagy valamennyi rekord vizsgálatát szükségessé teszi, vagy eleve feltételezi, hogy a rekordok rendezve vannak éppen az adott szempont szerint. A második esetben külön-külön az egyes visszakeresési szempontok szerint kell a dokumentumot leíró rekordoknak elérhetőnek lenniük. Ehhez azonban semmi szükség nincs az ún. járulékos információkra — ezek csak lassítanak és megnehezítik a keresést.

Durván a következő modellel szemléltethetjük a rendszert:

Maga a visszakeresési rendszer egy mátrixszal reprezentálható, amelynek sorai a tárolandó dokumentumokat képviselik, oszlopai az ismérveket. A mátrix elemei 0 és 1 jegyek lesznek, aszerint, hogy a sor dokumentumára vonatkozik-e egy oszlop ismérve vagy sem. (A mátrix a_{ij} eleme tehát 0, ha az i -edik dokumentumhoz nem rendeljük hozzá a j -edik deskriptort, és 1, ha hozzárendeljük.)

Maga a keresés az ismérveket jelentő *oszlopoknak* megfelelő Boole-vektorokon végrehajtott logikai műveletek sorozataként fogható fel. Egy dokumentum kiválasztásra kerül, ha a megfelelő logikai műveletek elvégzése után az adott sorszámú dokumentumot jelölő pozícióban 1 van, nem kerül kiválasztásra, ha 0 van. A valóságban a kiválasztás mechanizmusa bonyolultabb — de a lényeg az, hogy maga a kiválasztás jól elvégezhető az ismérveknek megfelelő ún. invertált file segítségével, majd az így kiválasztott dokumentumoknak megfelelő teljes információ kiíratható a dokumentumleírások direkt file-jából.

1.23 A KERESŐTÉMÁK LEÍRÁSA

A teaurusz harmadik jelentős feladata, hogy segítségével leírhatók legyenek a vizszakeresési témák. A keresőprofilok leírása ugyanolyan szabályok alapján végezhető, mint a dokumentumok osztályozása.

Elég ritka kivétel, hogy egy dokumentum képe (vektora) minden ismérv tekintetében éppen egy keresőképeknek (keresővektornak) feleljen meg. Láttuk, hogy éppen ezért a visszakeresés szempontjait már a dokumentumok osztályozása során érvényesítenünk kell. Azt is említettük, hogy az osztályozásnak redundánsnak kell lennie a keresőtémákra vonatkoztatva. Ez úgy valósítható meg, hogy a dokumentumokat általában lényegesen több deskriptorral osztályozzuk, mint amennyiből egy keresőkérdés általában áll. A keresőtémák a működő rendszerek tapasztalatai szerint átlag 4–6 deskriptorból állnak, míg a dokumentumokat – a rendszerek sajátosságaitól függően – 6–10, sőt néha 25–30 deskriptorral írják le.

A keresőkép kiszélesítését ugyancsak érintettük már az osztályozás tárgyalásakor. Itt csak azt akarjuk hangsúlyozni, hogy a teauruszban tárolt relációrendszer teljes egészében az automatikus gépi kereső rendszer rendelkezésére áll. Ennek segítségével sokkal fejlettebb kereső rendszerek valósíthatók meg, mint amit a fent leírt leegyszerűsített modell lehetővé tenne. A deskriptorok között kiépített kapcsolatrendszeren keresztül lépésről lépésre továbbhaladhat a keresés. Így nemcsak a keresőtémát leíró 4–6 deskriptor kerül vizsgálatra és összehasonlításra a tárukban, hanem azok is, amelyek ezekkel valamilyen kapcsolatban vannak: a magasabb rendű fogalmak (főlérendeltjei), vagy éppen az alacsonyabbak, részei vagy a nagyobb egységek, amelyeknek egyik alkotóeleme a deskriptor által megnevezett objektum stb.

A visszakereső rendszereknek módot kell nyújtani arra, hogy a keresőtémát akár külső beavatkozással, akár automatikusan javítani lehessen. A témaleírás módosítása két szempont értékelése alapján történhet: egyrészt annak alapján, hogy milyen típusú dokumentumok kerültek, illetve nem kerültek elő, másrészt annak figyelembevételével, hogy az előkerültek között melyek a legrelevánsabbak. Az is megvalósítható, hogy a vizszakeresés második menetében ne az eredeti keresőkép vegyen részt, hanem a legrelevánsabb dokumentum(ok) dokumentumvektorából kialakított keresőkép.

1.24 EGYÉB FUNKCIÓK

A fenti három feladaton kívül a teauruszok más szerepet is betöltenek, nem is csak szorosan a közvetlen értelemben vett informatikai igény oldaláról. A teauruszok a szaktudományok élő nyelvének szakkifejezéseire épülnek, ezeket és a köztük fennálló legfontosabb relációkat rögzítik, ezzel a szakmai nyelv és terminológia egységesítésében is nagy szerepet játszanak. A terminológia kérdéseit nem nyelvművelő szinten – a kifejezések oldaláról – vizsgáljuk, hanem elsősorban tartalmi oldalról. A szakkifejezések relációi mögött ugyanis – amelyekből a teaurusz struktúrája kialakul – tartalmi, logikai és nyelvi elemzés áll.

A szakkifejezések, majd utóbb a deskriptorok *gyakoriságának* vizsgálata az egyik legfontosabb eszköz a teaurusz kialakítására és továbbfejlesztésére. Ez nemcsak az információkereső rendszerek tervezése számára nyújt fontos támpontot, hanem a kutatási témák megoszlására és a kutatások intenzitására is rámutat, vagyis beállítható a tudáspolitikai és tudományszervezés, kutatásirányítás szolgálatába is. Hasonlatos ez a szociológia ún. content analysis (tartalomelemzés) módszeréhez, lényegében éppen azt valósítja meg. Ha a gyakorisági mutatókból idősorok állnak rendelkezésre, akkor a kutatások, a műszaki fejlesztés trendjeinek elemzését is lehetővé teszik egzakt módszerekkel.

A teauruszok, mivel az osztályozást szabályozott fogalmakkal végzik el, látszólag ellentétesek az automatikus osztályozással. Ez azonban csak a felhasználás módjára igaz, nem pedig a benne foglalt információkra. A teauruszok állhatnak az automatikus osztályozás vagy indexelés szolgálatában is, ha például ún. *pozitív szótárként* használjuk fel őket. Az automatikus osztályozás során vagy a szavak gyakorisága alapján emelhetjük ki az ún. kulcsszavakat, amelyekkel a dokumentumokat osztályozzuk, vagy ún. *negatív szótárban* tüntetjük fel az *irreleváns szavakat*. A szövegben előforduló összes többi szót kulcsszónak tekintjük, amivel a dokumentum indexelhető. Technikai okokból rendszerint a dokumentumcímek elemzésére korlátozódnak ebben az esetben. Ha azonban pozitív szótárra alapozzuk a kulcsszavak kiemelését, azaz az elemzett szöveg minden szavát egybevetjük a *pozitív szótárral*, és a kulcsszavakat ennek az összehasonlításnak eredményeként emeljük, akkor lehetővé válik nagyobb terjedelmű szövegek, pl. referátumok elemzése is osztályozási célokra. A teaurusz pozitív szótárként minden további nélkül felhasználható. Az így nyert indexelés igen tömör és tartalmas lehet. Negatív szótárként – egy nyelvi negatív szótárral együtt – olyan értelemben használjuk fel a teauruszt, hogy az abba fel nem vett, de gyakran előforduló szakkifejezéseket gyűjtjük össze mechanikusan részint a teaurusz továbbfejlesztése, részint a tudományos érdeklődés fejlődési tendenciáinak vizsgálata érdekében.

1.3 A TEZAURUSZOK ÁLTALÁNOS SZERKEZETE

Az előző fejezetekben felvázolt sokféle funkciót a teauruszok nem tudnák ellátni, ha a deskriptorok egyetlen mechanikus sorrendbe volnának állítva, nem nyújtanák az információkat többféle tálalásban. Korszerű információs teauruszok a deskriptorokat sokféle metszetben, sokféle összefüggésben fejezik ki, a különböző célú és szempontú igénybevételnek, szerepkörnek megfelelően. Általában a teauruszoknak az alábbi különböző felépítésű részét szokták elkészíteni:

a) Az egyedi deskriptorokra orientált részét – a teaurusz főrése –, amelyben minden deskriptor teljes kapcsolatrendszere kidolgozásra kerül. Itt minden deskriptor mintegy „szervező központ”, ahonnan elindulva valamely deskriptortól el lehet jutni a vele közvetlen vagy távolabbi kapcsolatban lévő kifejezésekhez, illetve a vele meghatározott típusú kapcsolatban lévő deskriptorokhoz. Ezt szokás a teauruszok „betűrendes

részének” nevezni. (A lényeg természetesen nem a deskriptorok sörrendjén van — az ábécé szerinti rendezés kizárólag a deskriptor fellelését biztosítja.)

b) Jelentéskörök szerint csoportosított részét, amelyben a deskriptorok átfogó fogalmak alá kerülnek, de ezen fogalmak alatt a deskriptorok betűrendben követik egymást. A deskriptoroknak így kapott átfogó csoportjai lehetnek mérsékeltén strukturáltak is a betűrenden belül abban az esetben, ha a deskriptorokat a velük közvetlen kapcsolatban álló deskriptorokkal együtt iktatjuk be. Ennek a résznek az a célja, hogy együtt legyenek a tartalmilag összetartozó deskriptorok.

c) A több szintű logikai (genus-species) kapcsolatrendszer teljes mélységében főfogalmak szerint kifejező részt. A tezausz ennek az olyan része, mely nem deskriptorközpontú, hanem rendszerközpontú. Célja a deskriptorok rendszerbeli helyének kijelölése a legfontosabb relációk módszeresen és teljességben kiépített hierarchiájában. Hasonló szerkezet készíthető bármely olyan relációrendszerre, amely rendezési relációt valósít meg (rész-egész, irányított asszociatív kapcsolatok stb.).

d) Végül az elsővel rokon szerkezetű, szintén deskriptorközpontú rész, a deskriptorok relációrendszerének szemléletes ábrázolása. Gyakran itt csak a legfontosabb deskriptorokat és relációkat szokták feltüntetni egy gráfokkal reprezentált rendszerben. A szemléletességen kívül célja lehet a *kapcsolatok erősségének* kifejezése is. Ezt szokás a tezausz grafikus részének nevezni.

A tezausz anyagának „tálalása” természetesen más hasznos részekkel is bővíthető: permutációs mutatókkal, idegen nyelvű kifejezések szótárszerű mutatóival, a használat során szerzett gyakorisági mutatókkal stb.

Részletesebben kell szólnunk a tezausz főrészéről, a hierarchikus és grafikus részről.

1.31 A TEZAUSZOK FŐRÉSE

A dokumentumok osztályozása és a keresőtémák leírása a tezausz főrészének segítségével megy végbe. Egységei: a *deskriptorcikkek* és nem-deszkriptorcikkek. A deskriptorcikk a *vezéredeskriptorból* áll és a vele közvetlen kapcsolatban álló deskriptorból, a kapcsolat jellegének jelölésével együtt. A deskriptorcikknek részei a nem-deszkriptorok (vagy tiltott deskriptorok) is, amelyek helyett a vezéredeskriptort kell használni.

A nem-deszkriptorcikk lényegében utalás. A könnyű megtalálhatóság érdekében a deskriptorokat és a nem-deszkriptorokat betűrendben rendezzük.

A deskriptorcikk részletesebben a következő részekből áll:

- a vezéredeskriptorból
- a deskriptorra vonatkozó, az osztályozó munkában közvetlenül nem használható információkból, mint a deskriptor kódjele, a szakterület jelzete, esetleg egyéb információkból, mint ETO-szám stb.;
- azokból a kifejezésekből, amelyek helyett is az adott deskriptor áll, tehát a szinonimákból;

– a deskriptorral közvetlen kapcsolatban álló deskriptorokból, a viszony jellegének feltüntetésével együtt;

– egyéb adatokból (tezauruszonként változhat), mint pl. a deskriptor fazettájának jeléből, esetleges idegen nyelvi ekvivalensekből stb.

A vezérdeskriptorral meghatározott relációban álló kifejezések maguk is szükségszerűen deskriptorok, tehát megjelennek olyan deskriptorcikkekben is, amelyekben éppen ők a vezérdeskriptorok, önnálló címszóként a saját kapcsolatrendszerükbe ágyazva. Ha pl. a vezérdeskriptor a „bibliográfia” és ennek egyik kapcsolódó deskriptora a „nemzeti bibliográfia”, akkor ez az utóbbi maga is megjelenik egy deskriptorcikk élén, mint „nemzeti bibliográfia”, amelynek viszont kapcsolódó deskriptora a „bibliográfia” lesz, de megjelenhet más, pl. az „általános bibliográfia”, „szakbibliográfia” . . . deskriptorok cikkeiben is. (Mivel mind az általános, mind a szakbibliográfiák lehetnek nemzetiek.)

Ha minden deskriptornál fel akarnánk tüntetni valamennyi kifejezést, deskriptort, amely vele kapcsolatban áll, akkor a deskriptorcikkeknek nem lenne vége. Ezért minden deskriptorcikkben csak a legközelebbi kapcsolódó deskriptorokat vesszük fel. Többre nincs is szükség, mert a kapcsolatrendszer teljes kiépítettségű, láncszerűen tovább keresve a legközelebbi kapcsolódó deskriptorok alatt az eredeti deskriptor távolabbi kapcsolatait is megtalálhatjuk. (Ha A deskriptor közvetlenül kapcsolatban áll B-vel, B deskriptor C-vel, C deskriptor D-vel, akkor A kapcsolatainál csak B-t tüntetjük fel. Ha A távolabbi kapcsolatai is érdekesek, akkor A-ból B-hez jutunk, és B-ből olvassuk le, hogy A-nak közvetve C-vel is kapcsolata van, és C alatt jutunk D-hez.) A legközelebbi kapcsolatos deskriptorokat viszont valamennyi tartalmazza, külön feltüntetve a kapcsolat típusát. A kapcsolatok feltüntetése mindig kétirányú, tehát külön relációként szerepel a tezauruszban valamennyi nem szimmetrikus reláció inverze. (A szimmetrikus relációnál is szerepel a visszautalás, de az inverz reláció típusa itt ugyanaz, mint az eredeti.)

A tezaurusz főrészeinek a *nem-deszkriptorok* esetében is tartalmaznia kell a szükséges utalásokat. Az utalóknak a tezauruszhasználatban fontos szerep jut. Akármilyen nyelvi alakban, a fogalmak bármely szóképében merül is fel egy téma, a tezauruszban utalók segítségével el kell tudnunk jutni azokhoz a deskriptorokhoz, amelyek az adott téma képviselőjét ellátnak. Az utalóknak több fajtája alakult ki és szerepkörük is eltérő lehet.

a) Utalni kell a szinonimákról a szinonimákkal egyenértékű jelentésű deskriptorra, hiszen a tezaurusz az azonos jelentésű szavak egyikét emeli csak a deskriptor rangjára. A szinonima saját betűrendi helyén áll, innen „lásd” jelentésű utaló vezet magához a deskriptorhoz. Pl. „vizuális lyukkártya – lásd fénylyukkártya”.

b) Ha a deskriptor összetett kifejezés és osztályozási szempontból valamelyik fordított szórendű alakja is értelmes, elfogadható alakváltozat, akkor ettől a változottól szintén „lásd” jelentésű utaló vezet a deskriptorhoz. (Az ilyen alakváltozatokat a könyvtárosi szóhasználatban inverz alakváltozatoknak nevezik, ami azonban nem tévesztendő össze a relációk inverzével.)

Tulajdonképpen az inverz alakváltozatok a szinonima egyik esetének foghatók fel. Legyen példa a „nagyfeszültségű erőátviteli berendezés”.

Akkor a hozzátartozó utalók:

erőátviteli berendezés, nagyfeszültségű

lásd

nagyfeszültségű erőátviteli berendezés,

illetve — megfontolandó — adható másik utalás is:

berendezés, nagyfeszültségű erőátviteli

lásd

nagyfeszültségű erőátviteli berendezés.

A szükséges alakváltozatok kiválasztása a szakember feladata.

c) Utalót kell készíteni a különböző helyesírású szóalakokról, ha a szakmai szóhasználatban több változat is elterjedt. Példa: computer *lásd* komputer, vagy éppen: számológép *lásd* számítógép, attól függően, hogy melyik alakváltozat válik deszkriptorrá. Ez a példa alkalmas arra is, hogy az eddig tárgyalt esetekre utalási változatokat adjon. Tegyük fel, hogy a deszkriptor „számítógép” alakú lesz. Ekkor a következő szinonimákról, illetve alakváltozatokról szükséges utalót készíteni:

elektronikus számítógép

lásd

számítógép;

komputer

lásd

számítógép;

computer

lásd

számítógép.

d) Más típusú utalást jelentenek azok, amelyek tartalmilag nem azonos deszkriptorra utalnak. A deszkriptorok számos esetben több fogalom közös, gyűjtő deszkriptoraiként jelennek meg, mivel az adott rendszerben nincs szükség vagy lehetőség további részletezésre. A deszkriptor jelentéstartalmát részleteiben kifejező fogalmakról ilyen esetben utalni kell az összefoglaló deszkriptorra. Ha pl. nincs szándékunkban a különböző szintű jogszabályok között különbséget tenni, akkor a törvény, törvényerejű rendelet, rendelet, utasítás, határozat stb.-ről utalást szükséges beépíteni. Az utalás alakja szintén „lásd” lesz, hiszen a felsoroltak helyett áll maga a deszkriptor:

törvény

lásd

jogszabály;

rendelet

lásd

jogszabály;

...

Ugyanez a helyzet áll elő minden hierarchikus kapcsolatrendszer legalacsonyabb (legbontottabb) szintjén. A deszkriptorokat felbontjuk alárendelt deszkriptorokra, azokat ismét alárendeltekre és így tovább. Valahol a felbontásnak abba kell maradnia, jöllehet a valóságban vannak még a legspecifikusabb deszkriptoroknál is specifikusabb

fogalmak. Ezekről *lásd* típusú utalás vezet a hozzájuk legközelebb eső genushoz, amely már deskriptorként szerepel a tezauszban. Az egész-rész reláció hasonló eseteket produkál.

A tezauszoknak itt ismertetett része — a relációk és utalások rendszerével — képezi a tezausz lényegét: a fogalmak bonyolult hálózata feldarabolva egyedi fogalmakra, úgy, hogy sértetlenül megmaradjon a fogalom közvetlen környezete, s belőle szükség szerint a távolabbi környezet is reprodukálható legyen.

1.32 A TEZAURUSZOK HIERARCHIKUS RÉSZE

A tezauszok hierarchikus része összefüggően mutatja meg a fogalmi hálózatot (vagy annak legfontosabb kapcsolatait). Nem fogalomcentrikus, hanem rendszercentrikus. Legtöbbször az osztályozás különböző szintjeinek egymáshoz kapcsolódását szokták vele kifejezni, de elvileg semmi akadálya nincs annak, hogy más legyen a rendező szempont (például a részekre bontás, vagyis az egész-rész reláció, vagy a fogalmak egymásra épülése, a fogalmi hierarchia).

Kezelhetőségét tekintve nagy előnyei vannak a fa alakban felírható hierarchikus rendszereknek (amilyen például egy könyv fejezetekre, ezen belül alfejezetekre, még tovább paragrafusokra, pontokra stb. való bontása). Ez annyit jelent, hogy az adott rendszerben minden fogalomnak csak egyetlen felettese lehet. Gyakorlati szempontból ez kétségtelenül hasznos megszorítás, a fogalmak valóságos összefüggései azonban nem mindig engedelmessé válnak neki — a valóság nem monohierarchikus, hanem polihierarchikus rendszereket produkál. A megfelelő elemek megismétlésével (vagy megfelelő keresztutalásokkal) ugyan ez is felbontható monohierarchikus struktúrává, ez azonban sok felesleges ismétléssel, ugyanannak a fogalomnak több helyen való felbukkanásával stb. jár.

Tudomásul kell venni tehát, hogy a tezausz hierarchikus része *korlátozott lehetőségeket* tartogat, nem olyan hajlékony eszköz, mint a fogalomcentrikus feldolgozás. Rendszerint arra szokták használni, hogy a *legfontosabb összefüggéseket* fejezzék ki vele, pl. egy-egy csomópontból kiindulva teljes mélységben mutassák a hierarchiarendszert.

A hierarchikus résznek többféle feladata lehet.

Mindenekelőtt gyakorta kerülhet sőr a tárolt dokumentumok bizonyos szempont szerinti következetes kilistázására. A jegyzék szakrendjét a hierarchia csúcsán lévő fogalmak adják meg. Ezen belül olyan mélységű bontást alkalmazhatunk, amilyent a tezausz hierarchikus része megenged — egyszerűen be kell járnunk a megfelelő fát. A hierarchikus rész alapján tehát az anyag következetesen — a hierarchiának megfelelően — kifejezhető.

Ha az osztályozó vagy a keresést irányító szakember tippeket akar kapni az index vagy a keresőkép összeállításához, akkor a megfelelő úton felülről lefelé haladva a fogalmi hierarchiában egyre pontosabban megközelítheti a leírandó témát jellemző fogalmakat. Ha már a kívánt deskriptort megtalálta, a tezausz betűrendes részéből felderítheti a további kapcsolatokat és viszonylag könnyen elvégezheti az osztályozást.

Végül szükség lehet egyszerűen a tezauszban tükröződő szakterület deskriptor-

állományának áttekintésére. Ahogy egy kézikönyv tartalomjegyzéke feltárja nagy vonalakban a vizsgált terület összefüggéseit, úgy a hierarchikus rész önmagában is áttekintést ad egy-egy nagyobb témáról. A betűrendes rész erre nem alkalmas, mivel áttekintése rengeteg lapozgatással járna. Természeténél fogva az első rész — a teaurusz főrésze — a fogalmak mikrovilágát tárja fel, mechanikus (ábécé szerinti) elrendezése pedig a természetes kapcsolatok feltöredezését vonja maga után. Ebből a szempontból is hasznos tehát valahol a természetes összefüggéseknek megfelelően szerepeltetni a deskriptorokat.

Felvetődhet a kérdés: létezik-e egyáltalán olyan hierarchia, amely a dolgok természetes rendjén nyugszik? Monohierarchia semmi esetre sincs ilyen. Ki kell választani tehát a lehetséges hierarchiák közül valamelyiket — a legmegfelelőbbet, amely az ember fogalmi gondolkodásának legjobban megfelel. Elménk sajátossága, hogy sokkal könnyebben tudunk tájékozódni egy olyan rendszerben, amely egyre speciálisabb eseteit különbözteti meg a fogalmaknak. Egy több szintű osztályozási rendszer van tehát az ember fejében, amely mindig készen áll arra, hogy további megkülönböztetésekkel tovább lehessen finomítani. Arról azonban végképp le kell mondanunk, hogy azonos felosztási alapon lehessen egy ilyen rendszert kiépíteni, mivel a kiválasztott szempont előbb-utóbb alkalmazhatatlanná válik, a rendszer pedig annál mesterkéltébbé, minél jobban erőltetjük a fogalmak belegyörmöszölését egy előzetesen kigondolt, logikusnak tűnő rendszerbe. Ami megfelel a dolgok logikájának az egyik helyen, az teljesen csődöt mondhat a másik helyen.

Ha jól belegondolunk, a fogalmaknak ez a renitens viselkedése voltaképpen nem is olyan meglepő. Az osztályozás mindig valamilyen újabb szempont kiemelését, érvényesítését jelenti — s milyen alapon várjuk el a valóságtól, hogy ugyanaz a szempont akárhányszor egymás után alkalmazható legyen? Az ember fogalom felbontható *férfira* és *nőre*, ezzel azonban ez a szempont ki is merült, többször nem alkalmazható. Ez azonban nem zárja ki a további osztályozást, pl. életkor szerint. Éppen az a kivétel, ha valamilyen szempont többször egymás után alkalmazható (pl. valaminek a *helye* több lépésben pontosítható — de előbb-utóbb ez a szempont is kimerül).

Adjuk meg tehát magunkat a dolgok — adott esetben az osztályozó emberi értelem — logikájának, és fogadjuk el azt a természetes hierarchiát, ami az emberi értelem és az adott ismeretág fejlődése során többé-kevésbé szükségszerűen kialakult. Fogadjuk el, hogy *felosztás alapja általában változik a hierarchia különböző szintjein és változnak a hierarchia ágai szerint is*. Legfeljebb retusálhatjuk néhol a struktúra saját logikáját pl. egy-egy összefoglaló fogalom mesterséges beiktatásával, alapjában azonban nem változtathatjuk meg.

A teaurusz hierarchikus szerkezetében elég a deskriptorokat feltüntetni, sem a szinonimákra, sem a nyelvi alakváltozatokra nincs szükség — az csak zavarná az áttekinthetőséget. Nemcsak felesleges, de az esetek többségében egyenesen káros volna feltüntetni a kapcsolódó fogalmakat, mivel ezek nem mutatnak olyan világos rendező tendenciát, mint a fentiekben vázolt fogalmi specializálódás.

A teaurusz említett két része között tehát olyan kapcsolat van, hogy a hierarchikus részben jelentkező összefüggések szükségszerűen megjelennek a teaurusz első részében is, de ami a teaurusz betűrendes részében kapcsolatként beépült, abból csak *egyfajta*

reláció és abból is csak egyfajta felosztású kapcsolat kerül a hierarchikus részbe. Az már a teaurusz készítőinek felelőssége, hogy ami bekerül, az a legfontosabb legyen.

Mint futólag említettük, annak sincs kizáró akadálya, hogy *egy-egy teaurusz többféle hierarchikus részt tartalmazzon*. A hierarchia felépíthető másfajta logikai alapon, vagy más relációk, pl. az egész-rész kapcsolatok alapján is. Ezzel a lehetőséggel azonban a gyakorlatban ritkán szoktak élni. A teaurusz betűrendes részéből a hierarchiából hiányzó kapcsolatok a szükségleteknek megfelelő szintig esetenként úgys kihámozhatók.

A hierarchikus rész az áttekinthetőség kedvéért olyan tipográfiai formában szokott megjelenni, ahol a deskriptorok elhelyezkedését a hierarchiában számkódok fejezik ki: a számjegyek helyi értéke a hierarchiaszintet, alaki értéke az azonos szinten elfoglalt sorrendi helyet fejezi ki.

A hierarchikus részben szereplő valamennyi fogalom deskriptor, s – ugyanúgy, mint a relációknak – természetesen szerepelnie kell a teaurusz első részében is. Fordítva a követelmény nem áll. A teaurusz tartalmi körébe szorosan bele nem illő, de a betűrendes részbe mégis felvett deskriptorok között természetes, hogy lehetnek olyanok – egy műszaki teaurusz esetében például bizonyos gazdasági, szervezési, jogi stb. deskriptorok, – amelyek nem jelennek meg a teaurusz hierarchikus részében. Erről a problémáról később még szólunk.

1.33 A TEZAURUSZOK GRAFIKUS RÉSZE

Szokás a teauruszoknak egy további, ún. grafikus részét is kiépíteni, esetleg a hierarchikus rész helyett. A grafikus részre még inkább igaz az, hogy nem tartalmazza feltétlenül a teaurusz valamennyi deskriptorát.

A grafikus rész igyekszik egyesíteni a teaurusz két előzőleg tárgyalt, különböző szerkezetű és funkciójú megoldását. A hierarchikus résszel ellentétben több különböző reláció szemléltetésére alkalmas. Korlátot az így kapott ábra áttekinthetősége szab. Az ábra középpontja egy kiemelt deskriptor, inkább fogalomközpontúnak tekinthetjük, tehát a közvetlen kapcsolatokon kívül a közvetett kapcsolatokat is feltünteti.

A grafikus ábrázolt teaurusz szintén csak deskriptorokat tartalmaz, szinonimákat soha.

A grafikus teaurusz ábrázolhatja:

- a relációk tulajdonságait: irányát, közvetlen vagy közvetett jellegét, szimmetrikus vagy aszimmetrikus voltát;
- a reláció logikai tartalmát: hogy fogalmi bontásról, részekre bontásról, asszociatív vagy más relációról van-e szó;
- a reláció pragmatikus használhatóságát: a kapcsolat erősségét, intenzitását, előfordulási vagy felhasználási gyakoriságát. (A gyakoriság itt nem a deskriptorokra, hanem az adott relációra vonatkozik!)

Konkrét megoldásokban természetesen nem kötelező, hogy a fenti három lehetőség mindegyike szerepeljen.

Az ábrázolás céljára felhasználható eszközök a következők:

- a síkbeli távolság, az, hogy az ábrán közelebb vagy távolabb helyezkednek-e el a deskriptorok (az azonos szemantikai távolságra lévő fogalmakat néha körökkel vagy keretekkel fogják egybe);

- különböző típusú élek, összekötő vonalak: folytonos, szaggatott stb. vagy különböző vastagságú élek;

- az élek mellé beírható súlyozási értékek.

Matematikailag értelmezve a grafikus tezaurszt, egy ilyen grafikus ábra egy ún. színezett (többféle relációt feltüntető) gráf. A gráf lehet „súlyozott” vagy súlyozás nélküli, tartalmazhat irányított és irányítás nélküli éleket. (E gráfban a csúcsok a deskriptoroknak, az élek a kapcsolatoknak felelnek meg. A súlyozást az élekhez írt pozitív számokkal fejezhetjük ki, ha nemcsak a kapcsolat megléte, hanem erőssége is lényeges.

A gráfok felírhatók a szokásos módon mátrix alakban is. E mátrixoknak akkor van igen nagy jelentősége, ha a tezaurszt különböző szempontokból elemezni akarjuk, pl. különböző kapcsolatok statisztikai eloszlását, gyakoriságuk egybevetését, korrelációját stb. vizsgáljuk.

Az ábrák megrajzolásánál vigyázni kell, hogy ne legyen túlságosan bonyolult, szövevényes, áttekinthetetlen. A grafikus ábrázolást értelmétől fosztja meg, ha a teljesség illúziója kedvéért éppen az áttekinthetőségről mondunk le.

1.34 A TEZAURUSZOK KIEGÉSZÍTŐ RÉSZEI

Ha egy tezaurszt elkészül, érdemes kiegészíteni egyéb tartozékokkal is, amelyek használatát megkönnyítik vagy sokoldalúbbá teszik. Vannak például olyan kiegészítő információk, amelyek nélkül hatékony visszakereső rendszer aligha valósítható meg, illetve amelyek nélkül a tezaurszt csak sokkal korlátozottabb körben használható.

Léteznek mindenekelőtt ún. formai deskriptorok, amelyek nem kívánnak hasonló elemzést és a kapcsolatrendszerre vonatkozó olyan kiépítettséget, mint fentebb láttuk – gyakran ennek nem is lehetne hasznát venni. Ilyen formai deskriptoroknak tekinthetők a földrajzi neveket, az időt, a dokumentumok formáját stb. kifejező deskriptorok.

A felsoroltak közül a földrajzi deskriptorok érdemelnek a legtöbb figyelmet. Ezek között bizonyos kapcsolatrendszer kidolgozására szükség van. Az országnevek vagy városnevek mellett politikai, gazdasági, katonai tömörülések, övezetek, nagyobb tájegységek, térségek, különböző egyezményekhez tartozó országok stb. fordulnak elő, s így köztük bizonyos kapcsolatrendszer is szükséges.

Az időt – időpont és időtartam kifejezését – elég általánosan szabályozni, nem szükséges a fogalmakat felsorolni.

A dokumentumok formáját kifejező, egyébként nem túlságosan sok deskriptort viszont előre rögzíteni kell. Ehhez a típushoz tartoznak (a kifejezetten dokumentumformátumra mutató jegyek mellett) azok a deskriptorok, amelyek a dokumentumok informatív értékére utalnak. Pl. hogy tartalmaz-e a dokumentum idegen nyelvű összefoglalót, statisztikai táblázatot stb.

A tezausz használátát ugyancsak kidolgozott szabályzatnak kell előírnia. A szabályzatnak ki kell térnie mind a dokumentumosztályozás, mind a keresőtema leírásának kérdéseire, elő kell írnia az osztályozás mélységét, az osztályozási műveleteket, szabályozni kell e műveletek hatását a visszakeresés, a formai deskriptorok használatát, az osztályozás formai kérdéseit stb. is.

Szintén szabályzatban kell gondoskodni a tezausz rendszeres fejlesztéséről. Milyen esetekben és hogyan lehet új deskriptorokat a rendszerbe beiktatni, milyen esetekben lehet a rendszerben már szereplő deskriptorokat megváltoztatni, pl. összevonni, szétbontani, milyen esetekben lehet a relációkon változtatni – ezek mind-mind olyan kérdések, amelyekre külön előírások kellenek. Egy tezausz soha nincs készen. Ha nem fejlődik, képtelen követni a tudományok fejlődését, s előbb-utóbb alkalmatlanná válik a szakterület tezauszának feladatkörét ellátni.

Végül a tezausz különböző részei és szabályzatai mellett szükség van a dokumentumok leírásának szabályaira is. A bibliográfiai leírás része egy-egy dokumentumrekordnak, amely tárolásra kerül. Közvetve így a tezausz használatát is befolyásolja, hogy a bibliográfiai leírás milyen adatokat tartalmaz, mely adatok alapján biztosítja a visszakeresést.

A gyakorlatban elkészült tezauszok legalább két részből állnak. Mindenképpen kötelező a főrész elkészítése, ehhez a jelentéskörök szerint csoportosított vagy hierarchikus rész csatlakozik. Harmadiknak a grafikus táblázatok alkotják a tezauszt. Ezeket még számos mutató követheti. A jelentéskörök szerint csoportosított és hierarchikus részt ugyanazon tezauszhoz elkészíteni fölösleges. Megengedhető az is, hogy hierarchikus rész ne készüljön a tezausz egészéhez, csak bizonyos fejezetekhez. Az ugyanazon deskriptorokat tartalmazó több hierarchia kiépítése azonban nemcsak megengedett, hanem néha kívánatos is.

*„Mephisto: Mi fő – kapaszkodjál a szóba
S megnyílik záros kapu módra
A Bizonyosság Temploma.*

Diák: A szónak maggal is kell bírnia.

*Mephisto: Igaz, de ezért magad gyötörni káros,
Mert hol a fogalom hiányos
Idején lelhetsz kellő szót oda.
Lehet szavakon lovagolni
Szóból egész rendszert koholni,
Hitet építeni rá előre
S van úgy, hogy egy betűt sem csenhetsz
el belőle.”
(Goethe)*

2. SZÓGYŰJTÉSI MÓDSZEREK

2.1 ÁLTALÁBAN

Osztályozási rendszerek, teauruszok, általában a visszakeresés céljait szolgáló osztályozási kifejezések induló szóanyagának összegyűjtésére az alábbi lehetséges módszerek alakultak ki:

- szakszövegek szóstatistikai elemzésének módszere;
- tapasztalati úton való szógyűjtés;
- meglévő indexek, osztályozási rendszerek, teauruszok adaptálásának módszere;
- elméleti módszer (ún. deduktív vagy filozófiai módszer);
- a fentiek valamilyen kombinációja.

A módszer megválasztása befolyásolja a kapott szóanyagot. Különböző eljárások eredményeként csak részben kapunk azonos szóanyagot, az induló kifejezések különbözősége pedig bizonyos mértékig a későbbi munkát és eredményét is befolyásolja. Ha különböző módszerekkel ugyanazokhoz a kifejezésekhez lehetne eljutni, akkor a módszer kérdése közömbös lenne és az egyik vagy másik módszer melletti döntést gazdaságossági és időtényezők szabnák meg. A különbözőséget természetesen nem kell eltúlozni. A más-más módszerrel nyert induló szóanyag eltérő volta azt jelenti, hogy egyes dokumentumok osztályozására másféle kifejezéseket vagyunk kénytelenek használni, de bármely módszert is választjuk, attól még a dokumentumok nem válnak osztályozhatatlannak. Ami az egyik eljárás esetén két-három kifejezéssel (deskriptorral) oldható meg, a másik esetben egyetlen kifejezéssel realizálható. Az egyik módszer inkább specifikusabb, a másik inkább

generikusabb kifejezéseket eredményez, egyik módszer inkább hajlik az összetett kifejezések felé, a másik inkább egységszavakból álló rendszer felépítésére ösztönöz stb. A későbbi elemzések kétségtelenül a különbségek kiegyenlítését eredményezik, teljes mértékben azonban sohasem lehet a kapott induló kifejezések meghatározó voltát megkerülni.

A módszer megválasztása nemcsak elméleti megfontolástól függ, hanem attól is, hogy mi áll rendelkezésre. Minden szakterületnek vannak osztályozási hagyományai, eszközei. Aligha képzelhető el tehát, hogy egy teaurusz elkészítését – beleértve a szógyűjtés módszerének megválasztását is – teljesen a zérusszintről kellene indítani.

A módszert azonban leginkább az osztályozási cél befolyásolja. Előzetesen dönteni kell, hogy milyen dokumentumtípusokat kívánunk osztályozni és milyen mélységben. Milyen lesz a rendszer outputja: bibliográfiák, katalógusok kötetformában való publikálása, egyedi visszakereső témákhoz végzett irodalomkutatás, szelektív információszűrés vagy ezek bármilyen kombinációja.

Nyilvánvaló, hogy másként kell monográfiákat osztályozni, mint folyóirat-közleményeket, másként a szabadalmi leírásokat, megint másként prospektusokat vagy más vállalati irodalmat. Pl. a Magyarország közlekedési statisztikáinak adatait tartalmazó kiadvány osztályozható úgy, hogy Magyarország, közlekedés, ilyen és ilyen időhatárok között, statisztika. Ugyanezt a művet lehet bontottabban osztályozni, ti. hogy megtalálhatók benne a közúti közlekedés, vasúti (sínpályás), vízi, légi közlekedés, teher- és személyszállítás adatai. Lehet ennél még részletesebben is analizálni a kiadványt, egészen az egyes táblázatok rovatainak részletességéig. Ebben az esetben választ kell kapnunk pl. arra, hogy a termelőszövetkezeti tulajdonban lévő Békés megyei nagy teherbírású gépkocsik árutonnakilométer adatai hogyan alakultak az utolsó öt évben.

A példa megvilágítja, hogy az *előzetes döntés* az osztályozást illetően már a választott szógyűjtési módszert is meghatározza. Elméleti úton gyűjtött szóanyag nem válhat alkalmassá mélyebb osztályozásra, a statisztikai módszer pedig nem nyújt megfelelő fogalmakat generalizáló osztályozásra, amelyre pl. a nyomtatott katalógusoknak lehet szüksége.

Egy átlagos esetet célszerű szem előtt tartani tehát, dokumentumtípusok tekintetében elsősorban a folyóiratcikkeket, az osztályozás mélységét illetően azt az esetet, amikor a cikk egészét, vagy annak egy-két résztemára bontásával kapott kisebb részeit osztályozzuk. Ez a szint felel meg körülbelül az átlagos követelményeknek. Sem a túlrészletezés (megéri-e?), sem a túlzottan általános osztályozás (mire jó?) nem merül fel ezzel az átlagos igényszinttel.

2.2 SZÓSTATISZTIKAI MÓDSZER

Szóstatisztikai elemzésre kiválasztott szövegek lehetnek:

- dokumentumok teljes szövegei;
- a kivonatok, referátumok, monográfiák tartalomjegyzékei;
- dokumentumcímek.

Az eljárás véghezvihető manuálisan, számítógépes segítséggel és teljesen automatikusan.

A módszer realizálható előkészítéssel — az elemzendő szövegek bizonyos előzetes bejelölésével és átalakításával — vagy végezhető teljesen mechanikusan.

Az automatikus szóstatisztikai elemzés magyar nyelv esetén ez idő szerint nem realizálható, mert a természetes szövegekben a szavak toldalékolt alakban fordulnak elő és nem állnak rendelkezésre azok a számítógépes programok, amelyek a szavak tőalakra hozását elvégzik. A vonatkozó nyelvészeti kutatások igen előrehaladott szakaszban vannak.

Számítógépes elemzésre tehát a szövegek előzetesen előkészítendőek. Kompromisszumos megoldás, ha a számítógéppel nem végeztetünk gyakorisági számlálást, ennek folytán bízni lehet abban, hogy egy szó ragozott, jelekkel ellátott, illetve képzett alakjai és összetételei a betűrendes szólistában egymás mellé kerülnek vagy egymástól nem nagy távolságra. A prefixumos alakok (a szavak elejére kerülő nyelvi funktorok, pl. igekötők) viszont távolra kerülnek a tőalaktól. Az így kapott szólistát tehát később, a listázás után lehet egységesíteni és a gyakoriságot megállapítani.

Ha előzetesen előkészítjük a szöveget, akkor a statisztikai elemzés korlátozódhat a magas redundanciával megállapított releváns kifejezésekre.

A számítógépes listázó programnak ki kell terjednie a szavak, szópárok, esetleg szóhármak kiírására is. Ennek következtében erősen terjedős jegyzéket kapunk, amit később még szelektálni kell.

Manuális elemzésnél a szöveg-előkészítés kiterjed:

- a nemkívánatos végződések levágására;
- a szóösszetételek esetleges felbontására;
- az összetett kifejezések megjelölésére;
- az összetett kifejezésekbe beékelődött irreleváns szavak törlésére.

A szóstatisztikai elemzés során minden szóra, illetve kifejezésre meg kell állapítani gyakoriságát és a gyakoriságot ki kell fejezni arányosan, vagyis százalékban is. A releváns kifejezések intervalluma tapasztalati úton határozható meg. Általában ez az intervallum a 0,5–1,7%-os gyakorisági tartományba esik. Kézenfekvő, hogy sem a leggyakoribb szavak (névelők, köztörzsek stb.), sem a legritkábbak nem eredményeznek hasznosítható kifejezéseket.

Ha az előkészített szövegben előzetesen kizárjuk az irreleváns kifejezéseket, akkor mind a gyakoriság abszolút számának, mind százalékszámának jelentése megváltozik. A releváns kifejezések százalékban kifejezett intervalluma ebben az esetben lényegesen kibővül. Megnö vizont az átlagos gyakoriságtól való eltérések fontossága, ugyanis ettől az átlagtól való erős pozitív vagy negatív eltérés kevésbé használható szavakat fog jelezni.

A kapott szóanyag egyaránt tartalmazni fogja a specifikus és átfogó fogalmakat. A specifikussági fokozat átlaga a teljes szövegek elemzése esetén lesz a legszélsőségesebben részletező, címek esetén sokkal kevésbé. De még mindig jóval részletezőbb lesz, mint más szógyűjtési eljárások kifejezései.

A tapasztalatok szerint referátumok elemzése eredményezi a leghasználhatóbb szóanyagot. A referátum — a címekkel szemben — elég részletezően írja le a dokumentum-tartalmat, vizont elég tömör és összefoglaló a teljes szöveghez képest. A referáló személy

általában a közhasználatú terminológiát használja, szemben a szerzők esetleges egyénies stílusával.

A szógyűjtés elégségességének ellenőrzése két úton mehet végbe. Vagy további szövegek hasonló módszerű elemzését végezzük, s megnézzük, kapunk-e új kifejezéseket, vagy úgy, hogy más módszerű szógyűjtést végzünk, és azt vizsgáljuk, hogy az így kapott fogalmak kifejezhetők-e a meglévőkkel vagy sem.

Az összegyűjtött és gyakorisági számmal ellátott, a fenti módon szelektált szóanyagot a gyűjtés befejezése után rávezetjük az elemzőcédulákra, szavanként, illetve fogalmanként egy-egy cédulára. A további elemzések már ennek segítségével mennek végbe.

2.3 TAPASZTALATI MÓDSZER

A tapasztalati szógyűjtést ajánlatos képzett, gyakorlott osztályozó szakemberekre bízni. Lényegében úgy megy végbe, hogy jól kiválasztott dokumentummintát adunk az osztályozó szakember kezébe, aki megkísérli osztályozni a dokumentumot a maga elképzelései szerint, osztályozó rendszer nélkül. Minden egyes dokumentum esetén maga tesz javaslatot azokra a kifejezésekre, amelyeket alkalmasnak ítél a dokumentum leírására, illetve osztályozására. Természetesen igénybe vehet segédeszközöket, terminológiai szótárakat, kézikönyveket vagy mást, de a tárgyszóalkotás egyéni és egyedi javaslat alapján jön létre.

Az osztályozás eredményeként létrejött kifejezéseket össze kell gyűjteni, és gyakoriságukkal együtt – későbbi elemzés céljára – adatlapra írni.

A siker egyik titka a jól választott dokumentumminta. A tárgyköröket jól kell lefednie és eléggé reprezentatívnak kell lennie. Ezt leginkább véletlen mintavétellel érhetjük el. Komplex szakterület esetén 800–1000 dokumentum, egyszerűbb, kevésbé összetett szakterület esetén 400–600 dokumentum elegendő lehet, ha a minta megfelelő volt. Osztályozási célra megfelelnek a dokumentumok referátumai is, sőt – tömörségük-nél fogva – tapasztalati osztályozással még alkalmasabbak is a szógyűjtésre, mint a teljes dokumentumszövegek. Referátumok könnyen nyerhetők referáló folyóiratokból, egyáltalán nem szükséges külön erre a célra referátumokat készíteni.

A szógyűjtés során már bizonyos szervezettséget vihetünk a módszerbe azzal, ha minden kifejezést az osztályozó azonnal kiír egy cédulára, s e cédulákat betűrendben tartja. Ahogy halad az osztályozásban, megnézheti saját döntéseit, már meglévő kifejezéseit. Így a szavak, kifejezések megválasztása következetesebb lesz. A tárgyszavak céduláira a dokumentumok (referátumok) azonosító számát is rá kell vezetnie, hogy ellenőrizhető legyen később is a döntések következetessége. A tételek száma egyúttal mutatja a gyakoriságot is.

A kigyűjtött kifejezések száma arányosan nő a feldolgozott dokumentumok számával, később egyre kevesebb lesz az új tárgyszó, végül egyre több dokumentumot (referátumot) kell feldolgozni akár egy-egy új kifejezésért. Amikor a dokumentumok már az

összegyűjtött tárgyszavakkal is osztályozhatók, a gyűjtést abba lehet hagyni, következhet a tárgyszavak elemzésének szakasza.

Ha az osztályozást egyidejűleg több szakember végzi, ügyelni kell a gyűjtés egyöntetűségére. A következtetlenség a szóanyag specifikusságában jelentkezik majd elsősorban. Ezért vagy ugyanazt a dokumentumanyagot kell kiadni több szakembernek, vagy a szétosztásnak kisebb egységeként véletlen kiválasztással kell végbemennie. Ha a szétosztást szakterületi bontásban végezzük, a gyűjtött szóanyag szakterületenként nagyon egyenletes lehet.

A módszer két veszélyt rejt magában. Először is minden osztályozó szakember valamilyen meglevő rendszerben sajátította el az osztályozó munkát, s a tárgyszóalkotásban ettől nehezen tud szabadulni. Végeredményben tehát rekonstruálja az általa ismert osztályozó rendszer szóanyagát, minden jószándéka ellenére. A másik veszély abban mutatkozik meg, hogy mechanikusan a dokumentumszöveg vagy referátumszöveg kifejezéseit fogja átvenni a rendszerbe. Ez azért hiba, mert akkor már egyszerűbb lenne az így nyert kifejezéseket statisztikai módszerrel elemezni. Ennek a módszernek éppen az az előnye a statisztikai módszerrel szemben, hogy a szöveg meghatározó voltát az osztályozó szakember szaktudásával akarjuk korrigálni. Ha a szöveg válik minden esetben a szóalkotás alapjává, akkor azokhoz a kifejezésekhez jutunk vissza, amelyektől éppen meg akarunk szabadulni. A szöveget nem lehet — nem is szabad — negligálni. A szöveg azonban csak inspiráljon. Legyen ellenőrzés alatt a gyűjtés időszakában is, az osztályozó szakembernek pedig legyen kellő autoritása, bátorsága ahhoz, hogy a szövegek eseti szóhasználatától szükség esetén eltérjen.

A fenti két hátrány együtt úgy fogalmazható meg, hogy a módszer kulcskérdése az osztályozók jó megválasztása. Ismert tény viszont, hogy jó osztályozó szakember nagyon kevés van, sokoldalú osztályozó még kevesebb.

E nehézségekkel szemben állnak a módszer kétségtelen előnyei:

a) Mindenekelőtt a gyorsaság. Egyszerűsége miatt a szógyűjtés egy-két hónap alatt vagy még rövidebb idő alatt elvégezhető.

b) Viszonylag alacsony költségek jellemzik.

c) Végül nagy előnye a szakmai kívánalmak maximális érvényesíthetősége már a szógyűjtés szakaszában is. Alkalmazkodni lehet a visszakereső rendszer kívánalmaihoz, különleges célkitűzésekhez, a felhasználói igényekhez.

A tapasztalati szógyűjtés egyik változatát képviselik azok a módszerek, amikor nem dokumentumszöveg vagy referátum, hanem a visszakeresési témák alapján kerül sor — a fentiekhez hasonló módon — a szóanyag összegyűjtésére. Az eljárást úgy lehet véghezvinni, hogy gyakorlati szakemberektől, kutatóktól, a szakterület vezetőitől elkérjük témáik 25–30 soros leírását, ugyanúgy, mintha témafigyelés céljára vagy visszakeresési célokra írnák le. Az osztályozó szakember dolgozza ezeknek a témáknak átültetése az általa javasolt osztályozási kifejezéseké.

Ha a visszakereső rendszer erősen felhasználói orientációjú — szemben a referátumokra épülő tudományra orientált eljárással —, akkor a keresőprofilok elemzését nem is mellőzhetjük. Kérdéses azonban, hogy önmagában az elég-e. A visszakereső rendszert ugyanis ezzel a napi aktualitások irányába toljuk, évek múltán azonban ennek megmutat-

koznak a hátrányai is, amikor a felhasználónak változnak a szempontjai. Teljes egészében tehát a keresőprofilok elemzésére nem lehet hagyatkozni, mint kiegészítő módszer azonban figyelemre méltó.

2.4 MEGLÉVŐ OSZTÁLYOZÁSI RENDSZEREK ÉS SZÓKÉSZLETEK FELHASZNÁLÁSA

Nincs olyan szakterület, amelynek ne lenne valamilyen osztályozási rendszere. Ha másként nem, akkor valamilyen egyetemes osztályozási rendszer részeként. Jogos tehát a kérdés: ezek szóanyagát miért ne lehetne a tezauruszépítésben felhasználni?

A szógyűjtésnél megszokott eljárás az a módszer, amely meglévő szójegyzékek átvételén alapszik.

A szóba jöhető források: meglévő tezauruszok, határterületi tudományok tezauruszai, osztályozó módszerek, tárgyszójegyzékek, indexek, kézikönyvek, alapvető monográfiák tárgymutatói, szakszótárak, szaklexikonok és enciklopédiák, terminológiai szabványok, szakmai nomenklatúrák, szakmai értelmező szótárak és egyéb szójegyzékek, pl. termékjegyzék, árjegyzék stb.

A felsorolt források különböző értékűek. Legcélravezetőbbek az osztályozási célra összeállított szójegyzékek, illetve indexek, mivel ezek általános célja megegyezik a készítenendő tezaurusz általános szempontjaival. A többi forrás más célt szolgál, átvételük csak fokozottabb kritikával képzelhető el.

Természetes, hogy egyetlen szójegyzékre a tezauruszépítésben nem lehet hagyatkozni, hiszen egyszerűbb lenne – a szóanyag megfelelő volta esetén – az adott rendszert átvenni. A több forrásból származó induló szóanyag

- nem fedi le egyenletesen a szakterületet;
- a különböző részanyagok nem képviselnek azonos specifikussági szintet;
- az idegen nyelvű források más szakmai gondolkodásmódot tükrözhetnek, más terminológiai rendszerben realizálódhatnak csak;
- nem idomulnak a konkrét célkitűzésekhez, sőt gyakran éppen más rendszerek, funkciók speciális kívánalmait tükrözik;
- nem minden műszó alkalmas egyben osztályozásra is (ez különösen a tárgymutatók, szótárak szóanyagára vonatkozik);
- nem elég korszerűek, hiszen a szaktudományok megelőző állapotának, tartalmának kifejezői. Négy-öt éves elmaradás egy szóanyagot avulttá és hiányossá tehet.

Ez a módszer a fenti, zavaró körülmények miatt fokozottabban előtérbe állítja a szógyűjtést követő elemző fázist, amelyben az egyenetlenséget ki lehet küszöbölni.

Mielőtt a számba jöhető forrásokból a szógyűjtés elindul, valamennyi forrás szóanyagát előzetesen értékelni kell. Az értékelés célja megállapítani, hogy

- melyik témákat részletezi inkább, s melyeket fed le elnagyoltabban;
- a szóanyag milyen bontottsági, részletező, specifikus szintet képvisel a többihez képest;

- a kifejezések hosszúsága általában mekkora: egytagú, kéttagú vagy több tagú összetettebb kifejezések jellemzik-e;
- mennyire korszerű a szóanyag, és milyen szakmai színvonal tükröződik benne;
- nem valamilyen speciális szempontból készült-e, amely a teaurusz szempontjából torzulásokat eredményez (pl. ha a kereskedelmi szempontok dominálnak, akkor a szóanyag elsősorban a terméknevezésekre helyezi a hangsúlyt, a technológiai szempontokat mellőzi; vagy például egy számítástechnikai szóanyag összeállítása nem lehet ügyviteltechnikai szempontú stb.).

A szóanyag specifikussága egyszerű módon ellenőrizhető. Egyik eljárás az lehet, hogy a szóanyagból mintát veszünk, 25–30 kifejezést. Ezeknek kiépítjük a hierarchiáját két-három fokozatig. Az így kialakult fogalmi családfákat vetjük egybe és igyekszünk megállapítani, hogy pl. az egyik forrásból származó alacsony hierarchia csúcsán lévő fogalmak azonos szintűek-e a másik forrás hierarchiájának alsó szintjével. A fokozati eltérések akár számszerűen is kifejezhetők.

Másik eljárás: az egyes fogalmakat, amelyek a mintában szerepelnek, ETO-számokkal fejezzük ki. Az ETO-számok hossza, illetve összetett volta mutatja az egyes források szavainak, kifejezéseinek specifikusságát.

Annak eldöntésére is több módszer kínálkozik, hogy az egyes szakterületeket egyenletesen, illetve a kívánalmaknak megfelelően fedik-e az egyes források. A szógyűjtés indulásakor már rendelkezésünkre áll a teaurusz szakcsoportjainak jegyzéke vagy nagyvonalú szakrendszere. Ennek hiányában az ETO szakrendszerét vehetjük alapul. A források szóanyagát vagy az abból származó mintát szakcsoportba rendezzük és megállapítjuk, hogy az egyes források mennyi kifejezést eredményeznek az egyes csoportokban.

Ha a szóba jöhető forrás szóanyaga megnyugtatónak mutatkozik, akkor a gyűjtési munka megindulhat.

A gyűjtést lehet a szóanyag előzetes kijelölésével indítani. Nem biztos, hogy a forrásjegyzékek minden kifejezésére szükség van. Tehermentesíthetjük a későbbi szakaszokat azzal, ha előzetes válogatást végzünk, különösen, ha a forrás nem fed teljesen a megkívánt tartalmat, vagy a szótárak, mutatók stb. szorúlnak előzetes szelekcióra.

Az idegen nyelvű forrásokat kijelölés után fordíttatni kell.

Az összegyűjtött szóanyag először betűrendezésre kerül, aztán írjuk rá ezeket az elemző kartonra. Ennek során az azonos kifejezések már az elemző kartonokon nem ismétlődnek. A kartonon valamennyi forrást feltüntetjük.

Ha a forrásjegyzék nemcsak kifejezéseket, hanem kapcsolatokat is tartalmaz, ezeket a szógyűjtés időszakában még lehetőleg ne vegyük figyelembe, hiszen szógyűjtést végzünk, nem elemzést. Ha az elemzés fázisában mégis kíváncsiak lennénk, hogy a forrásjegyzék milyen kapcsolatrendszert használt a kifejezésekhez, az elemző karton forráshivatkozása alapján az eredeti könnyen előkereshető.

Érdemes külön foglalkozni a szakirodalmi tájékoztatás céljára készült indexeken alapuló szógyűjtési módszerekkel. Az indexek közül elsősorban a permutált indexeket tartjuk szem előtt, de más szerkezetű indexművek is tartalmazhatnak alkalmas szójegyzékeket és megfelelő kiegészítő táblázatokat, pl. a citation indexrendszer is kiegészül alkalmazható keyword mutatóval.

Mindenekelőtt azt kell tisztáznunk, hogy az indexeken alapuló módszer átmenetet képez a statisztikai módszer és az itt tárgyalt eljárás között. A statisztikai módszernél tettünk is erre utalást.

A permutált indexek lényegében természetes nyelvű automatikus osztályozást valósítanak meg. Rendszerint a dokumentumok címében előforduló releváns szavakat sorolják betűrendbe egy ún. keresőoszlopban, kiírva a releváns szó elég terjedelmes szövegkörnyezetét. Maga az index automatikusan áll elő.

A további menetben azonban — egyelőre legalábbis — le kell mondanunk az automatikus szövegkezelésről, mert mint utaltunk rá, a magyar nyelvben ennek jelenleg még nincsenek meg a feltételei. Az angol nyelvre a Harvard egyetemen hatékony algoritmust dolgoztak ki automatikus szövegfeldolgozás céljaira. A módszer — Susumo Kuno gépi nyelvészeti analízise — jól alkalmazható az angol nyelvre. Többek között részletes tőszótár és végződésszótár áll az elemzés segítségére. A Cornell egyetemen pedig eredményes kísérletek folynak a tezauruszkészítés szinte teljes automatizálására. E kísérletek vezetője G. Salton professzor, az ún. SMART rendszer megalkotója.

A magyar nyelv e tekintetben lényegesen nehezebb problémákat vet fel, mint az angol: egyrészt gazdag morfológiával rendelkezik, másrészt a szótövekben gyakran történik a ragozás során változás. Még ha rendelkezésünkre állna is egy, a magyar szóképzésre összeállított tőszótár és toldaléktár, szintaktikai elemzés nélkül nem tudná kiszűrni a ragozás során megszorodott homonimiaproblémákat. Pl. a „falat” szó — a szintaktikai környezettől függően — lehet alanyesetű vagy tárgyesetű főnév.

A statisztikai módszerek alkalmazásának tehát komoly akadályai vannak. Egyelőre nincs más megoldás, mint permutált indexsorokból (és más forrásokból) *kézi erővel* összegyűjteni a tezaurusz induló szóanyagát. A permutált indexekben az előkészítés szempontjából a következő sajátosságokra ajánlatos figyelemmel lenni:

- a kulcsszavak az ún. keresőoszlopban fordulnak elő (így az elemzést elég a keresőoszlopra korlátozni);
- a kulcsszavak előfordulnak önállóan és szóösszetételben (pl. beton, betonoszlop, betonelem);
- a ragozott (toldalékolt) alakok nem kerülnek mindig a betűrendben a ragozatlan alak mellé;
- „kulcsszavaknak” kell tekinteni a „kulcskifejezéseket”, tehát lehetnek több tagúak is (pl. esztergagép, esztergapad karbantartása);
- egy-egy szó többször előfordulhat, az indexekből tehát gyakoriságuk is megállapítható.

A szöveg-előkészítés során a kulcsszavak az alábbi formára hozhatók:

- *főnéveknél* egyes számú, alanyesetű forma, kivéve, ha a szó eleve csak többes számban használatos. Minden más jel, rag levágandó;
- összetett szó esetén eldöntendő, hogy az összetétel marad-e vagy elemeire bontandó;
- az igét lehetőleg főnevesíteni kell (a szöveg átjavításával);
- összetett kifejezések előfordulásakor három eset lehetséges:
 - a) meg akarjuk tartani a több tagú kifejezést;

b) meg akarjuk tartani a kifejezés elemeit vagy az elemei közül valamelyiket;

c) mind a több tagú kifejezés, mind elemei külön-külön is kellenek.

Több tagú kifejezésekben a megtartandó elemeket megjelöljük, a közbeékelődött irreleváns szavakat pedig áthúzással töröljük. Végül megállapítjuk a kifejezések gyakoriságát és a gyakorisági számot a keresőoszlop bal oldalára írjuk. Ha összetett kifejezésről van szó, az összetétel elemeinél a gyakoriságot úgy számoljuk, hogy az önállóan előforduló esetek számához hozzáadjuk az összetett kifejezésben előforduló esetek számát. Ha csak a több tagú kifejezés marad, akkor gyakoriságnak azt számítjuk, ahányszor az adott kifejezés teljes egészében előfordul.

Képzett szavak esetében nehéz eldönteni, hogy mit tekintünk külön kigyűjtött szónak. Általában kövessük azt az eljárást, hogy külön esetnek számítsuk a képzett szavakat, ha azok különböző fazettába tartozhatnak, pl. válasszuk szét a beton és a betonozás szavakat, mert az egyik anyag, a másik művelet.

Ha a kulcsszó azonos alakú, de különböző jelentésű, kíséreljük meg elválasztását már a szöveg-előkészítés idején.

Több tagúként meghagyott kifejezések esetén tudnunk kell, hogy a szavak permutációja következtében elő kell kerülnie az összetétel összes helyén. A gyakoriság megállapításánál a konkrét szöveghely legyen mérvadó, tehát úgy tekintjük az esetet, mintha nem fordulna elő máshol.

Ha a kulcsszó számadat, töröljük. Jegyezzük fel viszont külön cédulára, hogy mely kulcsszavak mellett fordulnak elő szám adatok.

Az így előkészített szövegből lehet az elemzőlapokat kitölteni.

2.5 ELMÉLETI MÓDSZER

Ez a módszer az eddigi szógyűjtési eljárásokkal szemben — amelyek mind induktív utat jártak — deduktív szógyűjtési mód. Logikailag a szakterület felosztását kell elvégezni. A felosztás az osztályozással ellentétes logikai eljárás. Amíg az osztályozás a fogalmak osztályba sorolását jelenti, addig a felosztásnál mindig az egészből kell kiindulni, megállapítani közvetlen alárendelt (vagy rész-) fogalmait. Majd a kapott alárendelt, illetve részfogalmakat ismét felosztjuk, és az eljárás addig folytatódik, amíg a kellő részletességű kifejezésekig el nem jutunk. A kifejezetten hierarchikus osztályozó rendszerek nagyjából ezen az úton jönnek létre.

A felosztás nemcsak a dokumentumosztályozó munkában alkalmazható logikai módszer, hanem áthatja az egész tudományos gondolkodást. A tudományos rendszerezések alapja legtöbbször a deduktíve alkalmazható logikai eljárás, a felosztás. Éppen ezért a felosztásnak hatalmas elméleti irodalma van. A marxizmus klasszikusai közül Engels is behatóan foglalkozott a problémával. Az idevágó megállapítások idézése helyett csak annyit kívánunk leszögezni, hogy más kérdés egy tudomány felosztását, rendszerezését jól megadni, és ismét más kérdés praktikusán jó, visszakeresési célokra kiépített szöanyag-

hoz eljutni. Ha pl. egy mérnöknek az a feladata, hogy display terminálok méretezését végezze el, vajmi kevés segítséget jelent számára az a tudat, hogy pontosan ismeri a display vagy a méretezés helyét az emberi tudás egyetemes összefüggéseiben.

Mint szógyűjtési módszer azonban teljesen korrekt, elfogadott ez az eljárás. Generikusabb fogalmakat produkál a többi módszernél, a szakterület általában hézagtalanabbul, egyenletesebben fedi le. Kiegészítő, illetve ellenőrző szógyűjtési módszerként is jelentős, mivel a többiektől elvileg eltérő eredményeket ad.

Minden felosztáshoz szükség van egy ún. felosztási alapra. Olyan szempontra, amely alkalmas arra, hogy segítségével egy fogalmat alárendelt vagy részfogalmaira bontsunk. Az embereket például fel lehet osztani nemük, anyanyelvük, állampolgárságuk stb. szerint. Ha változik a felosztási alap, változnak a kapott fogalmak is. Alaptörvény, hogy *ugyanazon felosztásban* a felosztási alapot következetesen érvényesíteni kell, az alapokat nem lehet cserélgetni. Az emberek *nem* oszthatók fel ugyanazon a szinten nőkre, svájci állampolgárokra és német anyanyelvűekre. A felosztás akkor tökéletes, ha a kapott kifejezések egymást kölcsönösen kizárják, és a felosztás útján kapott fogalmak terjedelme éppen kiadja a felosztott fogalmat. Ez a „próba” igen fontos, mert a felosztások esetén nagyon sok hibát lehet elkövetni — a problémák általában nem olyan világosak, mint a fenti példában.

Az elméleti módszer alkalmazásánál — mint a teaurusz hierarchikus részeivel kapcsolatban említettük — nem lehet azt megkövetelni, hogy ugyanaz a felosztási alap vonuljon végig a fogalmi rendszer valamennyi szintjén. A fenti szigorúság csak ugyanarra a felosztási szintre vonatkozik. Ha egyszer jól elkészítettünk egy felosztást, akkor a *további* bontásnak már lehet más szempontja is. Ha pl. a gépjárműveket felosztottuk személy- és teherautókra, nem szükséges, hogy ugyanazt a felosztási alapot (ti. hogy mit szállít) erőltessük tovább. Lehet a felosztási alap pl. a hajtómű, a motor, a márka, a nagyság vagy valami más.

A felosztási alapok száma elvben végtelen lehet. Annak eldöntése, hogy a lehetséges felosztások közül melyek a legfontosabbak és ezek milyen sorrendben kövessék egymást, a szaktudomány problémája. Ha nem tudunk lényeges felosztási szempontokat találni, akkor valószínűleg a kapott fogalmak sem lesznek valóban relevánsak. Ezért a felosztás szempontjainak jó megválasztása a vizsgált szógyűjtési módszer sikerének kulcsa.

A felosztási szempontok között ritkán lehet abszolút sorrendiséget megállapítani. Gyakran nem elég egyetlen felosztást elvégezni, többet kell, valamennyi számításba jöhető lényeges ismérv alapján. Így elérjük, hogy felosztásunk nem lesz egysíkú, egydimenziójú, ami nagyon rontja a kereső rendszerek hatékonyságát. Több rendszert kapunk tehát. Ezek közül a teaurusz hierarchikus részébe feltehetően egyet — bár az nem kötelező — jelölünk ki. Csakhogy mint szógyűjtési módszernek a teaurusz valamennyi részére kell a fogalmat produkálni.

Annak eldöntésére, hogy a lehetséges felosztási szempontok közül melyeket érdemes érvényesíteni, praktikus segítség is van, nevezetesen, ha igényeket, azaz konkrét keresőtémákat, tájékoztatási kéréseket vizsgálunk meg. Ha ilyen nincs, akkor az érdekeltektől

kérünk ilyen témákat. Azok a dimenziók, amelyekben a kérdések felmerülnek, támpontokat adnak a felosztási alapok meghatározására. Az előbbi példánál maradva: keresnek-e úgy, hogy a gépjárművek melyik országból származnak? Keresnek-e aszerint, hogy a gépjármű hány éve van forgalomban? Nagyság, teherbírás szerint? Tulajdonos szerint? Motor-típus szerint? Színük szerint? Lehetne a példákat sorolni tovább. A kérdés minden szempontja egyben egy-egy felosztásnak felelhet meg. A különböző felosztások alapja pedig kialakulnak egymást nem feltétlenül kizáró fogalmaink is, s így összefüggérendszerünkben sem vagyunk egysíkúságra kárhozthatva.

Mint kiderül, a módszer mind az alárendelő, mind a részfogalmak megállapítását lehetővé teszi. Ha ehhez hozzávesszük a más-más felosztás útján nyert fogalmakat, be kell látnunk, hogy a szógyűjtést követő elemző fázist ez a módszer teszi a legegyszerűbbé. A legfontosabb relációk szinte készen vannak. Előnyös abból a szempontból is, hogy a fogalmak eleve kifogástalan, legelfogadottabb alakban kerülhetnek meghatározásra, a gyűjtött szóanyagnak a legkisebb lesz a „hordaléka” (a később eldobandó szó).

Számos előnye mellett azonban ez a módszer nyújt legkevesebb védelmet az egyéni véleménnyel szemben. Tapasztalat szerint alig van két szakember, aki meg tudna egyezni saját tudománya felosztásában vagy akár felosztásaiban. A fogalomanyag tehát egyéni véleményt tükröz majd. A többi módszer fogalomanyaga is a szakmai szerzőktől származik, de kollektíva szóanyagát elemezzük, amelyből a merőben egyedi szóhasználatot azonnal kizárhatjuk.

Hátránya még az is, hogy néha nem lehet a kellő részletekig eljutni. A felosztás túlságosan következetes logikája néha mesterkélt vagy szükségtelen fogalmakat eredményez. Az ideális felosztás teljesen logikus, a valóság azonban nem mindig az. Valóban elméletivé sikerülhet az eredmény – most már a szónak rosszabb értelmében –, elszakadhat a szóanyag a való élettől, és ha még egyéb hibák is előadódhatnak, akkor elsősorban az üres logikai konstrukció érvényesül. A visszakereső rendszernek éppen ez a feladata, hogy a terminológia helyenkénti pontatlansága, logikátlansága ellenére is eredményesen működjön.

2.6 MÓDSZEREK KOMBINÁCIÓJA

A módszerek leírása során kiderült, hogy mindegyiknek vannak előnyei a többivel szemben, de vannak olyan következményei, amelyeket szívesen kiküszöbölnénk.

Az egyes módszerek hátrányainak kiiktatására, illetve az előnyök érvényre juttatására a szógyűjtés során érdemes a módszereket együttesen alkalmazni. Ha a gyűjtési módszer az induktív eljárások valamelyikére akar alapozni, akkor a deduktív szógyűjtési módszer felel meg leginkább kiegészítő módszerként. Ha a választott fő gyűjtési technika az elméleti módszer lesz, kiegészítő módszerként a tapasztalati vagy statisztikai eljárást válasszuk. Minél különbözőbb módszereket kombinálunk, a gyűjtés annál eredményesebb lesz.

hoz eljutni. Ha pl. egy mérnöknek az a feladata, hogy display terminálok méretezését végezze el, vajmi kevés segítséget jelent számára az a tudat, hogy pontosan ismeri a display vagy a méretezés helyét az emberi tudás egyetemes összefüggéseiben.

Mint szógyűjtési módszer azonban teljesen korrekt, elfogadott ez az eljárás. Generikusabb fogalmakat produkál a többi módszernél, a szakterület általában hézagtalanabbul, egyenletesebben fedi le. Kiegészítő, illetve ellenőrző szógyűjtési módszerként is jelentős, mivel a többiekétől elvileg eltérő eredményeket ad.

Minden felosztáshoz szükség van egy ún. felosztási alapra. Olyan szempontra, amely alkalmas arra, hogy segítségével egy fogalmat alárendelt vagy részfogalmaira bontsunk. Az embereket például fel lehet osztani nemük, anyanyelvük, állampolgárságuk stb. szerint. Ha változik a felosztási alap, változnak a kapott fogalmak is. Alaptörvény, hogy *ugyanazon felosztásban* a felosztási alapot következetesen érvényesíteni kell, az alapokat nem lehet cserélgetni. Az emberek *nem* oszthatók fel ugyanazon a szinten nőkre, svájci állampolgárokra és német anyanyelvűekre. A felosztás akkor tökéletes, ha a kapott kifejezések egymást kölcsönösen kizárják, és a felosztás útján kapott fogalmak terjedelme éppen kiadja a felosztott fogalmat. Ez a „próba” igen fontos, mert a felosztások esetén nagyon sok hibát lehet elkövetni — a problémák általában nem olyan világosak, mint a fenti példában.

Az elméleti módszer alkalmazásánál — mint a teaurusz hierarchikus részeivel kapcsolatban említettük — nem lehet azt megkövetelni, hogy ugyanaz a felosztási alap vonuljon végig a fogalmi rendszer valamennyi szintjén. A fenti szigorúság csak ugyanarra a felosztási szintre vonatkozik. Ha egyszer jól elkészítettünk egy felosztást, akkor a *további* bontásnak már lehet más szempontja is. Ha pl. a gépjárműveket felosztottuk személy- és teherautókra, nem szükséges, hogy ugyanazt a felosztási alapot (ti. hogy mit szállít) erőltessük tovább. Lehet a felosztási alap pl. a hajtómű, a motor, a márka, a nagyság vagy valami más.

A felosztási alapok száma elvben végtelen lehet. Annak eldöntése, hogy a lehetséges felosztások közül melyek a legfontosabbak és ezek milyen sorrendben kövessék egymást, a szaktudomány problémája. Ha nem tudunk lényeges felosztási szempontokat találni, akkor valószínűleg a kapott fogalmak sem lesznek valóban relevánsak. Ezért a felosztás szempontjainak jó megválasztása a vizsgált szógyűjtési módszer sikerének kulcsa.

A felosztási szempontok között ritkán lehet abszolút sorrendiséget megállapítani. Gyakran nem elég egyetlen felosztást elvégezni, többet kell, valamennyi számításba jöhető lényeges ismérv alapján. Így elérjük, hogy felosztásunk nem lesz egysíkú, egydimenziójú, ami nagyon rontja a kereső rendszerek hatékonyságát. Több rendszert kapunk tehát. Ezek közül a teaurusz hierarchikus részébe feltehetően egyet — bár az nem kötelező — jelölünk ki. Csakhogy mint szógyűjtési módszernek a teaurusz valamennyi részére kell a fogalmat produkálni.

Annak eldöntésére, hogy a lehetséges felosztási szempontok közül melyeket érdemes érvényesíteni, praktikus segítség is van, nevezetesen, ha igényeket, azaz konkrét kérés-témákat, tájékoztatási kéréseket vizsgálunk meg. Ha ilyen nincs, akkor az érdekeltektől

kérünk ilyen témákat. Azok a dimenziók, amelyekben a kérdések felmerülnek, támpontokat adnak a felosztási alapok meghatározására. Az előbbi példánál maradva: keresnek-e úgy, hogy a gépjárművek melyik országból származnak? Keresnek-e aszerint, hogy a gépjármű hány éve van forgalomban? Nagyság, teherbírás szerint? Tulajdonos szerint? Motor-típus szerint? Színük szerint? Lehetne a példákat sorolni tovább. A kérdés minden szempontja egyben egy-egy felosztásnak felelhet meg. A különböző felosztások alapja pedig kialakulnak egymást nem feltétlenül kizáró fogalmaink is, s így összefüggérendszerünkben sem vagyunk egysíkúakra kárhozthatva.

Mint kiderül, a módszer mind az alárendelő, mind a részfogalmak megállapítását lehetővé teszi. Ha ehhez hozzávesszük a más-más felosztás útján nyert fogalmakat, be kell látnunk, hogy a szógyűjtést követő elemző fázist ez a módszer teszi a legegyszerűbbé. A legfontosabb relációk szinte készen vannak. Előnyös abból a szempontból is, hogy a fogalmak eleve kifogástalan, legelfogadottabb alakban kerülhetnek meghatározásra, a gyűjtött szóanyagnak a legkisebb lesz a „hordaléka” (a később eldobandó szó).

Számos előnye mellett azonban ez a módszer nyújt legkevesebb védelmet az egyéni véleménnyel szemben. Tapasztalat szerint alig van két szakember, aki meg tudna egyezni saját tudománya felosztásában vagy akár felosztásaiban. A fogalomanyag tehát egyéni véleményt tükröz majd. A többi módszer fogalomanyaga is a szakmai szerzőktől származik, de kollektíva szóanyagát elemezzük, amelyből a merőben egyedi szóhasználatot azonnal kizárhatjuk.

Hátránya még az is, hogy néha nem lehet a kellő részletekig eljutni. A felosztás túlságosan következetes logikája néha mesterkélt vagy szükségtelen fogalmakat eredményez. Az ideális felosztás teljesen logikus, a valóság azonban nem mindig az. Valóban elméletivé sikerülhet az eredmény – most már a szónak rosszabb értelmében –, elszakadhat a szóanyag a való élettől, és ha még egyéb hibák is előadódhatnak, akkor elsősorban az üres logikai konstrukció érvényesül. A visszakereső rendszernek éppen ez a feladata, hogy a terminológia helyenkénti pontatlansága, logikátlansága ellenére is eredményesen működjön.

2.6 MÓDSZEREK KOMBINÁCIÓJA

A módszerek leírása során kiderült, hogy mindegyiknek vannak előnyei a többivel szemben, de vannak olyan következményei, amelyeket szívesen kiküszöbölnénk.

Az egyes módszerek hátrányainak kiiktatására, illetve az előnyök érvényre juttatására a szógyűjtés során érdemes a módszereket együttesen alkalmazni. Ha a gyűjtési módszer az induktív eljárások valamelyikére akar alapozni, akkor a deduktív szógyűjtési módszer felel meg leginkább kiegészítő módszerként. Ha a választott fő gyűjtési technika az elméleti módszer lesz, kiegészítő módszerként a tapasztalati vagy statisztikai eljárást válasszuk. Minél különbözőbb módszereket kombinálunk, a gyűjtés annál eredményesebb lesz.

Az adaptációs módszer egyaránt jól kiegészíthető mind az elméleti, mind a közvetlen szógyűjtések valamelyikével. Ha rendszerünk generikusabb célkitűzésű, akkor inkább az elméleti módszerrel egészüljön ki, ha inkább individualizáló osztályozásra törekszünk, akkor inkább statisztikai vagy tapasztalati gyűjtéssel.

A szógyűjtő munkát másként is lehet szervezni, nemcsak a fent vázolt módszerek alapján. A tezaszuszkészítést magát is fel lehet osztani nagyobb témákra. A témánkénti, szakcsoportonkénti szógyűjtés a módszerek együttesét kívánja meg, egyidejűleg a közvetlen gyűjtést különböző forrásokból, meglévő szólisták adaptációját. Ebben a szervezésben több vagy valamennyi módszer együtt él. Előnye az, hogy elég áttekinteni a tezaszus egyik témájának szóanyagát, a gyűjtőmunka során mindig látni lehet, hol tart a szógyűjtés és milyen a szóanyag minősége.

Egyetlen hátránya ennek a racionális megoldásnak az, hogy szakcsoportonként más minőségű szóanyagot kaphatunk. Az elemző szakasznak, amely a gyűjtést követi, megőőnek a feladatai, mert fokozottabban előtérbe kerül az egységesítés.

OSZK

Nemzeti Széchényi Könyvtár

„Természetes törekvése az emberi elmének, hogy amit megismert, rendezni akarja. Az emberi és természeti jelenségek milliói szétguruló töredékek maradnának a megismerés után is, ha nem tudnók őket, kapcsolatuk minősége szerint, valami egységesen felfogható egészszé forrasztani.”

(Horváth János)

3. RELÁCIÓK

3.1 ÁLTALÁBAN

A szógyűjtést akkor tekintjük befejezettnek, ha a teaurusz várományos kifejezései, az összegyűjtött fogalmak lefedik a teaurusz tartalmát egy nagyjából előre meghatározott mélységi szinten. Egy ismeretág vagy tudományág fogalmakban reprodukált képét kaptuk meg. Csakhogy a tudományra nemcsak fogalmi rendszere jellemző, hanem meghatározott viszonyok is a fogalmak között. A relációk éppen annyira fontos részei a tudománynak, mint a fogalmak. *A relációkban* — logikailag — *ítéletek fogalmazódnak meg*. A szaktudományok nyelvén a relációk tételek, amelyekben a fogalmak közti kapcsolatok testesülnek meg. Amikor tehát a teaurusz szerves részeként jelöljük meg a fogalmak mellett a köztük fennálló relációkat, akkor a teauruszban a tudomány tartalmához alkalmasabb, mélyebb, hívebb kifejezésének igényét fogalmazzuk meg.

Ha valóban igaz, hogy a relációkban ítéletek öltenek testet, akkor a relációknak messzemenőig azokhoz az ítélettípusokhoz — relációtípusokhoz — kellene igazodniuk, amelyek a szaktudományt jellemzik. Bármennyire formalizálnánk is azonban egy-egy tudomány lehetséges viszonyait, egy teauruszban leírhatatlanul sok relációt kellene felvenni. Ha viszont megelégednénk a formális logikai absztrakt relációval, akkor távol kerülünk a konkrét tudománytól, a teaurusz tényleges feladataitól. A teauruszépítés tervezési szakaszaiban kell ennek az ellentmondásnak a feloldásáról döntenünk: elég konkrét legyen, de ugyanakkor elég általános is a relációrendszer. Ne legyen túlságosan bonyolult, hiszen ez a visszakereső rendszer technikai megvalósítását veszélyeztetné. Másrészt ne legyen túlságosan leegyszerűsíthető se, mert ebben az esetben gyengíti a visszakereső rendszer hatékonyságát.

Az itt tárgyalt relációk valamilyen átmenet — szükség diktálta kompromisszumot — jelentenek a két szélsőség között.

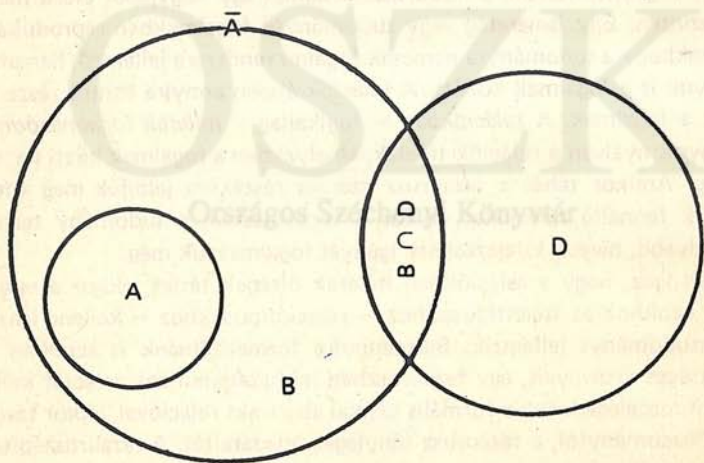
Második megjegyzésünk a relációkhoz, hogy a tárgyalásmód egyszerűsítése miatt olyan kapcsolatokat is ebben a fejezetben tárgyaltunk meg, amelyek nem a szaktudomány

tartalmának kifejezésére szolgálnak, de egyöntetűséget, következetességet és egyszerűsítést visznek a tezauszba. Közvetve ezek a kapcsolatok is tartalmi összefüggéseket tükröznek azzal, hogy azok tisztább beépítését teszik lehetővé a tezauszba.

A relációkat sokkal tömörebben is lehetne tárgyalni megfelelő jelölések bevezetésével. Az elterjedt jelöléseket is megadjuk a fejezetekben, de mégis inkább a szóbeli kifejtésre törekszünk, hogy a tartalmi részletek megtárgyalása közérthető legyen.

3.2 RELÁCIÓELMÉLETI ALAPFOGALMAK

Leibniz óta szokás a fogalmak terjedelmét körökkel, körök által határolt területekkel jelölni. A halmazelméletben hasonló módszert alkalmaznak a halmazok közötti kapcsolatok kifejezésére: az ún. Venn-diagramokat. (Elnevezése John Venn angol matematikus nevét idézi.) Az alábbi ábrán igyekszünk összefoglalóan ábrázolni a fogalmak terjedelme közti viszonyokat. (A fogalmat felfoghatjuk mint a fogalmi jegyek halmazát, de odaképzeltük egyúttal az általunk jellemzett objektumok – speciálisan dokumentumok – halmazát is.)



Ha A jelöl egy fogalmat, akkor \bar{A} jelöli a tagadását, a komplementerét. A komplementer A-nak ellentéte, olyan dolgok megnevezését jelenti, amelyek nem (részben sem) tartoznak A fogalmi körébe. Az osztályozásban jelentése: kivéve. Kereshetünk pl. meghatározott dokumentumokat különböző nyelveken, kivéve a francia nyelvűeket.

A francia nyelvű dokumentumok tagadása az összes többi dokumentumot jelenti. (Szokásos jelölés még: $\sim A$.)

$A \cap B \cap D$ metszetviszony (interszekció, a jegyek logikai sorozata) két fogalom

körének közös részét fejezi ki. A \cap jel kiejtése — alak után —: kapu. Ilyen viszony létesíthető pl. a „fehérjetartalmú élelmiszerek” és a „halak” deszkriptorokkal jelölt fogalmak között. (Nem minden halfajta élelmiszer!)

A $B \cup D$ egyesítést, uniót, a fogalmi jegyek logikai összegét jelöli. Alakjáról a \cup jelet kupának mondjuk. A két fogalom együttes terjedelmét jelöli oly módon, hogy az így képzett fogalom egyesíti magában mindazokat a fogalmi jegyeket, amelyek vagy B-re vagy D-re jellemzőek (vagy mindkettőre egyszerre). A „harmadik generációs számítógépek” és az „IBM számítógépek” fogalmak logikai összegébe beletartozik minden harmadik generációs számítógép, akár IBM gyártmányú, akár nem, és beletartozik minden IBM gyártmányú számítógép, akár harmadik generációs, akár nem. Nyilván beletartoznak az IBM harmadik generációs számítógépei is.

Az $A \subset B$ azt az ítéletet jelöli, hogy A teljes terjedelmével beletartozik B-be, tehát valódi része. (Inklúzió, szubszumció, része, alárendeltje.) Pl. a „főnév” fogalom ilyen módon tartalmazza a „tulajdonnév” fogalmát. A \subset jel neve: kampó.

Ha $A \subset B$ és $B \subset A$, akkor ez csak úgy állhat fenn, ha A és B egyenlőek. Ezt ekvivalenciaviszonynak nevezzük. Pl. az „üres halmaz” és „nullhalmaz” fogalma egyenértékű.

A $B - D$ jelölést két fogalom, B és D különbségét, D-nek B-re vonatkozó relatív komplementerét jelöli. Jelenti B terjedelmének azt a részét, amely nincs benne D-ben. A gépjárművek és a tömegközlekedési eszközök különbségét azok a gépjárművek jelentik, amelyeket nem használnak személyszállításra, hanem pl. teherfuvarozásra.

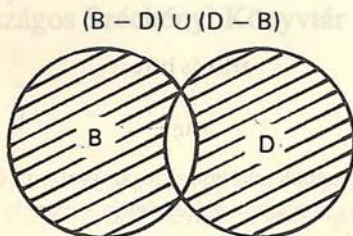
A teljesség kedvéért megemlítjük az ún. szimmetrikus különbséget. Jele:

$$B \oplus D$$

és jelentése:

$$(B - D) \cup (D - B).$$

Ábrázolva:



Az osztályozási munkában ez a reláció aligha fordul elő. Ha igen, más, egyszerűbb módon fejezendő ki.

Minden fogalom egyének, individuumok, egyedi tárgyak stb. sokaságának feleltethető meg — azoknak az objektumoknak, amelyekben az adott fogalom realizálódik. Az asztal pl. az összes létező asztalt megnevezi, tehát „összegyűjti” ezeket. Ebben az értelemben minden fogalom egyben osztályt képvisel, az általa gyűjteménybe foglalt dolgok osztályát. Maguk a dolgok, tárgyak az osztály elemei. Ezt a következőként fejezzük ki.

$$a \in A.$$

Az \in jel a görög kis ábécé epszilonjának ún. groteszk alakja. Ha egy egyed nem eleme az osztálynak, akkor így jelöljük:

$$a \notin A.$$

Az MTA pl. eleme a tudományos intézmények osztályának.

Az itt tárgyalt kapcsolatokon kívül — amelyek a legfontosabbak — még számos másfajta lehetséges. Ha két fogalom között valamilyen viszony van, azt általánosan úgy fejezzük ki, hogy

$$aRb,$$

és azt jelenti: a az R relációban áll b-vel.

Mint a jelölésből kitűnik, a „relációkalkulus” eredetileg objektumokra vonatkozik, tehát dolgokra, tárgyakra stb. Minden nehézség nélkül kiterjeszthető azonban a jelölés fogalmak közti relációkra is, hiszen adott esetben ezeket is tekinthetjük objektumoknak. Ebből nem származik zavar, de ha mégis előfordulna, hogy ez a kiterjesztés zavart okozna, akkor külön keresünk megoldást.

A relációk lehetnek *szimmetrikusak*, ha ugyanaz a reláció oda és vissza is igaz, azaz, ha

$$aRb$$

és

$$bRa$$

egyidejűleg fennáll. Az „egyetemes bibliográfia” asszociatív viszonyban van a „bibliográfiai kontroll”-al, ugyanígy a bibliográfiai kontroll asszociál az egyetemes bibliográfiával.

Valamely reláció aszimmetrikus, ha a reláció csak egyirányú, visszafelé már nem érvényes, vagy visszafelé *más* reláció érvényes. A puli alárendeltje a kutyanak, visszafelé nem igaz, a kutya nem alárendeltje a pulinak.

A reláció tranzitív, ha

$$aRb \text{ és } bRc$$

esetén

$$aRc$$

mindig fennáll. Ha a szerszámgépek alárendeltje az eszterga és az esztergák alárendeltje a másoló eszterga, akkor a szerszámgépek alárendeltje a másoló eszterga is.

A reláció nem tranzitív, ha

$$aRb \text{ és } bRc$$

esetén

$$aRc$$

nem szükségszerűen áll fenn. Ha pl. a szerszámgépek alárendeltje az eszterga és az esztergák *része* a szegnyereg, akkor a szerszámgépeknek *nem* alárendeltje a szegnyereg.

A reláció egyenértékű, ha egy fogalom mindig csak legfeljebb egy fogalommal áll az adott kapcsolatban, több értékű, ha egy fogalom több fogalommal áll ugyanolyan kapcsolatban (pl. egy fogalomnak több alárendeltje lehet). Jellemző a relációra az is, hogy a

neki megfelelő fordított (inverz) reláció egy vagy több értékű. Így állnak elő a fogalmak közötti bonyolultabb „oda-vissza” megfeleltetések (egy-egy, egy-több, több-egy, több-több), ami analóg a nyelvészet homonima-szinonima problémáival.

A relációkat fel lehet írni táblázatos formában is. A táblázatok elkészítésének leg-egyszerűbb és további elemzéseket lehetővé tevő módja a mátrixreprezentáció. A fogalmak felelnek meg a mátrix sorainak és oszlopainak, ezek állnak a sor- és oszlopfőekben. A mátrix cellájában 1-et írunk, ha a sor fogalma és az oszlop fogalma között az adott reláció fennáll, és 0 kerül a cellába, ha a reláció nem áll fenn.

Igy egy fogalom-fogalom mátrix keletkezik. Ezt azért kell kiemelni, mert a visszakereső rendszerek modellálásában más típusú mátrixok is szerepet játszanak, pl. az 1. fejezetben tárgyalt dokumentumfogalom-mátrix.

Legyen pl. három kifejezésünk: síkidomok, háromszögek, derékszögű háromszögek, jelölje ezeket rendre a, b és c. A köztük levő alárendelési reláció: b alárendeltje a-nak, c alárendeltje b-nek és – a tranzitivitás elvének megfelelően – a-nak is. A megfelelő mátrix:

	a	b	c
a	0	1	1
b	0	0	1
c	0	0	0

A tájékoztatási gyakorlatban rendszerint csak a *közvetlen* alárendeltséget jelöljük. A közvetlen alárendeltség:

	a	b	c
a	0	1	0
b	0	0	1
c	0	0	0

A mátrixból leolvasható a reláció iránya is: a-nak alárendeltje b, tehát a sorának és b oszlopának találkozási pontjában 1 áll. Fordítva a reláció nem igaz, ezért b sorának és a oszlopának cellájában 0 áll.

Több reláció mátrixát összevonhatjuk egyetlen mátrixban is. Ilyenkor a zérus elem fejezi ki a reláció hiányát, s a zérustól különböző pozitív egész számmal fejezhetjük ki a reláció típusát. Ez a jelölés azonban csak akkor szerencsés, ha egyetlen cellába csak egyetlen típus jelét kell beleírunk. A bonyodalmak elkerülése végett mi inkább relációként írjuk fel a mátrixot.

Hasonló módon szokás a reláció erősségét kifejezni 1-től különböző, nem feltétlenül egész számokkal. Ezzel jelölhető pl. a korrelációs viszony gyengébb vagy erősebb volta.

A mátrix elemzése egyéb tanulságokkal is szolgál. Jellemzi a 0-tól különböző elemek száma a fogalom „hajlamát” arra, hogy más fogalmakkal gyakran vagy ritkábban

lépjen kapcsolatba. A sorok jellemzőek arra, hogy a fogalomból kiindulva más fogalmak melyek és mennyien vannak, míg az oszlopából leolvasható, hogy ő mely és hány fogalomból származtatható az adott reláció szerint.

A relációk kifejezésére van egy másik jól használható matematikai apparátus is: a gráfok. A gráfelmélet nyelvén lényegében ugyanazokat a problémákat lehet megfogalmazni, amilyenekkel a relációelmélet foglalkozik. Különösen jól használható eszközt jelentenek a gráfok, ha nem túlságosan sok elem közötti kapcsolatokat akarunk szemléletesen kifejezésre juttatni. Ezek a kapcsolatok akár több reláció együttesét is jelenthetik az elemeknek ugyanazon a halmazán. Minthogy a gráfok és az általunk vizsgált speciális mátrixtípusok között kölcsönösen egyértelmű megfeleltetést lehet létrehozni, a relációk egyikféle ábrázolási módjáról mindig egyértelműen át lehet térni a másikra és viszont. Így mindig azt a formát választhatjuk, ami az adott probléma jellegének legjobban megfelel.

A gráf lényegében pontok (az ún. *csúcsok*) és a köztük lévő kapcsolatot kifejező összekötő vonalak (az ún. *élek*) együttese:

$$G = \langle C, E \rangle,$$

ahol C a csúcsok, E az élék halmaza. Természetesen nem szükségszerű, hogy minden csúcsot minden csúccsal él kössön össze. Ez az eset — az ún. teljes gráf esete — mátrixrepresentációban annak felel meg, amikor a mátrix minden cellájában 1 áll. Az érdekes éppen az, hogy melyik csúcsot melyik csúccsal köti össze él (mátrixrepresentációban: mely cellákban áll 1).

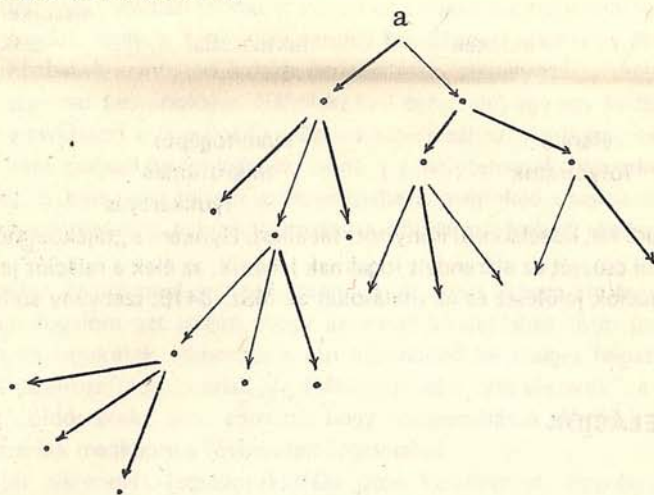
A gráf csúcsainak is, éleinek is jelentés tulajdonítható — esetünkben a gráf csúcsai feleltethetők meg a fogalmaknak, az élék a köztük fennálló relációknak. Ha csak a kapcsolat fennállását akarjuk kifejezésre juttatni, akkor nem szükséges az élék között különbséget tenni. Ha a reláció típusát is meg akarjuk különböztetni, akkor — úgy képzelve, hogy minden relációtípust más-más színű éllel jelölünk — ún. *színezett gráfokat* kell alkalmaznunk. Az élék valóságos színezése helyett azonos betűvel vagy azonos számjeggyel jelölhetjük az azonos „színű” éleket. Két-két csúcsot elvileg több él is összeköthet — ilyenkor keletkezik az ún. *multigráf*. A teauruszok gyakorlatában ennek alkalmazására csak ritkán van példa.

Ha a reláció nem szimmetrikus — márpedig általában nem az —, akkor közönséges élék helyett ún. irányított éleket használhatunk, vagyis közönséges vonalak helyett nyilakat. Az ilyen gráfot *irányított gráfnak* nevezzük. Az egy-egy csúcsból kiinduló és oda beérkező nyilak száma alapján beszélhetünk bármely csúcs kibocsátási, illetve befogadási fokáról. Láttuk, hogy mindez a mátrixról is közvetlenül leolvasható.

A gráfok közül az alkalmazások szempontjából különös figyelmet érdemelnek az ún. *irányított fák*. Egy irányított gráfot akkor nevezünk irányított fának, ha

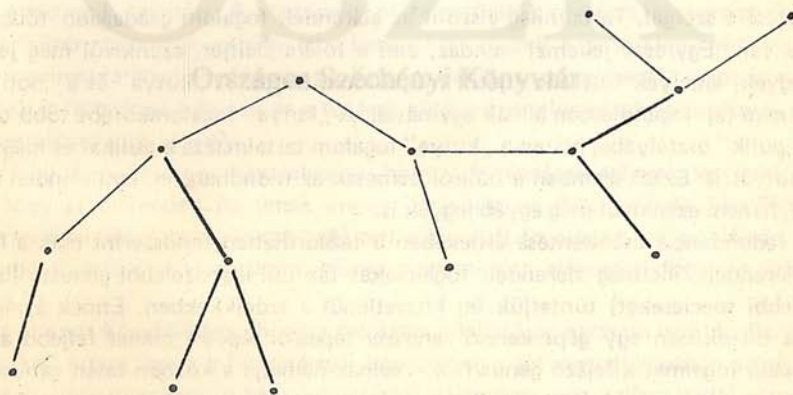
- van egy kitüntetett csúcsa, amelynek a befogadási foka zérus. Nevezzük ezt a csúcsot a fa gyökerének és jelöljük pl. a -val,
- minden más csúcs befogadási foka 1,
- nincs a gráfban ciklus, vagyis bármely csúcsból elindulva és a nyilak mentén haladva soha nem juthatunk vissza egy csúcsba, ahol már jártunk.

Az alábbi gráf pl. irányított fa:



Az irányított fa szerteágazó utak halmaza. Azokról a csúcsokról, amelyeknek kibocsátási foka nagyobb, mint egy, azt mondjuk, hogy elágazások.

A fa fogalmához úgy jutunk, hogy eltekintünk az élek irányításától. Az irányított fa ennek megfelelően mindig fát alkot, de megfordítva ez nem igaz. A fának nincs gyökere, illetve minden pontja gyökérnek tekinthető. Az alábbi ábrán egy fát látunk:



A gráfban irányított éleket kell használnunk, ha a kapcsolat egyirányú, pl. alárendelés esetén. Ha a reláció oda-vissza igaz, akkor eltekinthetünk az élek irányításától.

Az alábbi példa legyen egy ismert, egyszerű eset a fentiek illusztrálására. A „tájékoztató szolgáltatások” fogalom felbontása egyfajta alárendelési reláció szerint:

Tájékoztató szolgáltatások

szétsugárzó			visszakereső
gyors tájékoztatás	indexek	faktográfiai	szakirodalmi
kurrens			
bibliográfia			
	referáló	számítógépes	
	folyóiratok	mikrofilmes	
		lyukkártyás	

A fenti struktúra fát, közelebbről irányított fát alkot. Gyökere a „tájékoztató szolgáltatások”. A gráf többi csúcsát az alárendelt fogalmak jelentik, az élek a relációt jelzik.

A tezauszrelációk jelölését és az utalásokat az MSz. 3418. szabvány szabályozza.

3.3 TEZAURUSZRELÁCIÓK

3.31 NEM-FAJ VISZONY

Ha A fogalom teljes terjedelmével beletartozik B -be, azaz, ha ACB fennáll, akkor B -t A fölérendeltjének mondjuk, A -t pedig B alárendeltjének. A logikában szokás a genus és species elnevezést is használni. (A tájékoztatásban az „apa” és „fia” is elterjedt.) Az alárendelt fogalom speciális esete a fölérendeltnek, köztük terjedelmi és tartalmi különbségek vannak. Terjedelmileg az alárendelt „kisebb”, az individuumok szűkebb körének megnevezésére szolgál. Tartalmilag viszont az alárendelt fogalom gazdagabb: több jellemző jegye van. Egyrészt jellemzi mindaz, ami a fölérendeltjét, ezenkívül még jellemzik olyan jegyek, amelyek fölérendeltjétől megkülönböztetik. A „kutya” és a „puli” fogalmak pl. nem-faj kapcsolatban állnak egymással. A „kutya” fogalomkörébe több tartozik, mint a „pulik” osztályába, hiszen a „kutya” fogalom tartalmazza a pulikat és még a többi fajtájú kutyát is. Ezzel szemben a puliról nemcsak az mondható el, ami minden kutyára jellemző, hanem ezenkívül még egyéb jegyek is.

A redundancia csökkentése érdekében a tezauszban rendszerint csak a legközelebbi fölérendelt, illetőleg alárendelt fogalmakat (az ún. legközelebbi genus, illetőleg a legközelebbi speciest) tüntetjük fel közvetlenül a szócikkekben. Ennek az információnak a birtokában egy gépi kereső rendszer lépésről lépésre mehet feljebb az egyre általánosabb fogalmat kifejező genus felé, s felhasználhatja a közben talált genusok kapcsolatrendszerét is. Több-kevesebb lépés után természetesen eljutunk olyan általános fogalmakig, amelyek már nem sorolhatók még általánosabb fogalmak osztályába. Némi-leg hasonló a helyzet az egyre speciálisabb fogalmak felé haladva is: itt is eljutunk egy olyan szintig (az egyéb ágon hamarabb, a másikon később), ahol már nincs értelme a további specializálásnak. Vagy olyan speciális fogalmakhoz jutunk, amelyeknek még kialakult terminológiája se igen van (az ismeretág fejlődésével természetesen tovább bomolhat a fogalom — ezt a tezausz továbbfejlesztésénél figyelembe kell venni), vagy

az adott kereső rendszer részletezési elvei szabnak határt a további bontásnak (gyorsan fejlődő ismeretág esetében időnként ezeket az elveket is módosítani kell).

Előfordul, hogy a fogalmi hierarchia következetessége vagy áttekinthetősége érdekében a közbeeső szinteken kisebb pontosításra, szabályozásra vagy éppen módosításra szorul a szakmai terminológia. Néhol ki kell egészíteni egy-egy közbeeső terminus beiktatásával a rendszert a fogalmi fa világos kialakításához. Az ilyen beavatkozásnak természetesen nem szabad mesterkéltnek lenni, s a helytelennek, következetlennek ítélt terminológiát is bele kell építeni a tezauruszba a megfelelő utasításokkal, hogy a kereső rendszer hatékonyságát a logikai struktúra következetesebbé tétele emelje, és ne csökkentse.

Minden fogalomnak több fölérendelt és több alárendelt fogalma lehet. A több fölérendelt fogalom azt jelenti, hogy az adott kifejezéshez több úton is eljuthatunk, a tudományos ismeretek lebontása során különböző lehetséges felosztások csomópontjában áll. A „szappan” fogalomnak pl. fölérendeltjei a „mosószer”, a „kapilláraktív anyagok”, az „oldószer” stb., aszerint, hogy felhasználásuk, hatásuk vagy tulajdonságaik alapján akarjuk megkapni a fölérendelt fogalmakat.

Több alárendelt fogalom kétféle úton keletkezhet. Egyrészt úgy, hogy ugyanazon felosztás több alárendelt fogalmat eredményez, másrészt úgy, hogy különböző felosztások lehetségesek. Ha két (vagy több) alárendelt fogalom ugyanannak a genusnak azonos módon van alárendelve, akkor egymáshoz képest *mellérendelők*, „testvérek”. Ha két (vagy több) alárendelt fogalom a genus különböző felosztásai útján keletkezett, akkor kollaterális viszonyban állnak egymással. Egy adott szempont szerinti felosztásban a kapott mellérendelő fogalmak terjedelmileg kizárják egymást. A kollaterális kapcsolatban álló speciesek között ezzel szemben metszetviszony áll fenn, ezek a fogalmak részben átfedhetik egymást. A „katonák” pl. feloszthatók egyrészt fegyvernemek szerint: gyalogos, tüzér, páncélos, repülő stb. katonákra, s ezek a genuszt egymást kizáró alárendelt fogalmakra bontják. Másrészt a katonák lehetnek ténylegesek, sorozottak és tartalékosok. A tartalékos katonák és a tüzérek különböző felosztások eredményei, egymással metszetviszonyban vannak.

A logikában általános követelmény, hogy a felosztások teljesek legyenek. Ez azt jelenti, hogy az alárendelt fogalmak uniója (terjedelmeik összege) adja ki a fölérendelt fogalom terjedelmét. A tezauruszba beépített alárendelt fogalomra ezt az előírást nem lehet mindig szigorúan érvényesíteni, hiszen nem minden alárendelt fogalom válik deskriptorrá.

Az viszont követelmény, hogy a felosztás a lehetőség határain belül logikus legyen. Ezen azt kell érteni, hogy a felosztásnak legyen meg a jól meghatározható alapja. Ebből következik, hogy speciesekre való felosztásnál szigorúan meg kell maradni az adott fogalmi kategóriában, vetületben, „fazettában”, ha úgy tetszik. Eszköz alárendeltje csak eszköz lehet, eljárás eljárás és így tovább.

A fogalmak felosztásának egyik gyakorlatilag is jól használható módja az ún. dichotomikus felosztás. Ez olyan felosztást jelent, amikor a genuszt felosztjuk egyik speciesére és a speciesnek a genusra vonatkozó komplementerére. Az élelmiszereket feloszthatjuk gyümölcsre és nem gyümölcsökre. A közlekedési eszközöket közúti járművekre és nem

közúti járművekre. Ez a felosztás igen fontos, mert ha valamely genus alárendeltjei közül csak valamelyik fontos, akkor ezzel az eljárással megkaphatjuk a fontosnak ítélt alárendeltet, ugyanakkor a felosztást a logika szabályainak megfelelően tudjuk elvégezni.

A visszakereső rendszerek szempontjából az alá-fölérendelési relációk kiépítését a biztonságos keresésre való törekvés teszi indokolttá. Utaltunk már arra, hogy a keresőkép és a dokumentumoknak megfeleltetett deszkriptorsorozat között ritkán van teljes egyezés. Ha a megadott deszkriptor alapján a visszakeresés eredménytelen vagy nem kielégítő, mindig meg lehet ismételni a keresési műveletet a deszkriptorhoz legközelebb eső fölérendelt fogalmak szerint, illetve az alárendelt deszkriptorok alapján.

Ezt az ismétlést az eredmények értékelése alapján a kereső rendszer automatikusan is elvégezheti. Maguk a visszakereső programok úgy készíthetők el, hogy a tezauszba beépített relációkat automatikusan figyelembe vegyék, és ennek alapján menet közben módosítsák, kiszélesítsék a keresőképet. Az első keresés ilyenkor közvetlenül az ember által megadott keresőkép deszkriptorai alapján megy végbe. Az eredmény megfelelő értékelése alapján a gépi program megállapíthatja, hogy milyen mértékben szükséges bevonni a további keresésbe a fölérendelt és/vagy alárendelt deszkriptorokat, és így tovább.

Természetes, hogy a kiválasztott dokumentumok ily módon különböző mértékben fogják csak a keresőképet visszatükrözni. Az a dokumentum, amit nem a kereső deszkriptor alatt, hanem — mondjuk — a genusa alatt találtunk, már kevésbé lesz releváns. A relevancia mértékét — amelyet a keresőképpel való egyezés mértéke fejez ki — súlyszámokkal lehet finomítani. A súlyszámok azt mutatják, hogy a kereső deszkriptorok közül — a tapasztalatok szerint — mennyivel lép fel teljes egyezés, hány esetben egyezik csak az alá-, illetve fölérendelt deszkriptor és hány esetben nincs találat egyáltalán.

Egyre jobban elterjednek az információkeresésben az ún. interaktív rendszerek. Ezek bizonyos esetekben az emberre bízák a döntést: terminál írógépeken tájékoztatást adnak a folytatási lehetőségekről, s ugyanott várják a begépelte választ. Az ember ezáltal közvetlenül irányíthatja a keresés menetét.

3.32 EGÉSZ-RÉSZ VISZONY

A tezauszalkészítésben az egész-rész viszonyt gyakran a fölé-alárendelési reláció analógiájának tekintik. Holott nem az, sőt, a két relációt azonos módon kezelni a visszakeresés zajtényezőjének növelése nélkül nem lehet. A fölé-alárendelési viszonyt a fogalmak terjedelme (köre) és tartalma közötti viszonyként definiáltuk. Az egész-rész viszony esetében nincs szó a két fogalom terjedelmi vagy tartalmi kapcsolatáról. (Tartalmi kapcsolaton a meghatározó jegyek, ún. differencia specifikák közti viszonyt értjük.) Az asztalnak alárendeltje az íróasztal, része pl. az asztalfiók, zárszerkezet stb.

Az egész-rész viszony leggyakrabban — fizikai objektumok esetén — közvetlen *fizikai, szerkezeti* kapcsolat. A részfogalmak azokat a részeket jelentik, amikből az egész felépül, az egész a részek összességét. Az egész más minőségű, bonyolultabb tárgy, mint részei. A más minőséget az adja, hogy az egész több, mint részeinek összege. A virág

egyrészt részeiből áll: gyökérzet, szár, levelek, ezenkívül még abból a minőségből, ami a részeket virággá teszi.

A részfogalmak ismét feloszthatók további részfogalmakra. A folyamatot elvileg a tovább nem bontható homogén alkatrészekig lehet folytatni. Azonban ugyanúgy, mint a genus-species relációnál, a tezauszban mindig csak a legközelebbi szintig bontjuk az egészet. További bontást már a deskriptorra emelt részfogalmaknál kell eszközölni. Formális hasonlóság tehát fennáll az analógiaként felhozott fölé-alárendelési relációval.

Folyamatok részei részfolyamatok. A tervezésnél pl. rész a pénzügyi tervezés, a tezausz-készítés folyamatának része a szögyűjtés.

Az egész-rész reláció jelenthet időbeli, illetve térbeli tagozódást, felbontást. Folyamatok időbeli felbontása nem mindig esik egybe a más szempontú részfolyamatok rendszerével, hiszen pl. ha a folyamatot a tevékenység tárgya alapján bontjuk fel, időben párhuzamos tevékenységeket is kaphatunk.

Intézmények részeit szervezeti szempontból lehet megkapni. Az ENSZ-nek része ebben az értelemben a Biztonsági Tanács.

Elvont tudományos műszavak, terminus technicusok esetén néha nehéz eldönteni, hogy alárendelési vagy részviszony esete áll-e fenn. A számelmélet pl. milyen viszonyban áll a matematikával? Rész-e? És az alkalmazott matematika? Ilyen esetekben valóban az előző fejezetben a fogalmak tartalmának és terjedelmének viszonyáról mondtak a mérvadók. Ha alárendeltje, akkor a genustól valamilyen jellemző jegyben különböznie kell, de ami a genus jellemző jegye, annak mind jellemeznie kell az alárendeltet. A puli teljes kutya, csak fajtája szerint más, mint a többi. Ezzel szemben a számelmélet nem teljes matematika, hanem része, amely tárgya szerint különbözik pl. a valószínűségelmélettől, matematikai analízistől. Tehát részfogalom. Az alkalmazott matematika viszont teljes matematika egy meghatározott szempontból, az alkalmazás oldaláról tekintve. Tehát nem rész-, hanem fajfogalom.

A részfogalmaknak egyidejűleg több egésze, és természetesen minden egésznek több része lehet. A műszaki tudományok területén ennek különböző hangsúlya van, hiszen alkatrészek, gépelemek szinte önálló életet élnek, és ugyanazon szerkezet, elem, alkatrész sok mindennek lehet része. A kondenzátorok pl. a legtöbb elektromos szerkezetnek részei. Éppen ezért szabványosítják is az alkatrészeket, hogy sokoldalú felhasználásuk biztosítva legyen. A tezauszban csak a legfontosabb „egész” fogalmakat szabad felvenni, amelyek felvétele a tezausz célja szempontjából indokolt.

A fogalmaknak részfogalmakra bontása több különböző szempontból történhet. Egy berendezés feldarabolható pl. különböző funkciókat ellátó szerkezeti részeire, továbbá valamely tulajdonság, pl. hőhatásnak kitett és ki nem tett részekre, megint más szempontból stabil, illetve mobil részekre és így tovább. A folyamatok részeit megkaphatjuk a tevékenység tárgya, a bizonyos módszereket alkalmazó részek szerint, de intenzitásuk stb. szerint is. A tezauszépítésben sohasem a lehetséges, hanem a szükséges darabolásokat kell elvégeznünk.

A visszakeresés mechanizmusában az egész-rész viszony másként hat a végeredményre, mint amit a fölé-alárendelt deskriptoroknál láttunk. A kereső engedélye nélkül veszélyes dolog megengedni egy fogalomról való automatikus továbblépést egészeihez,

illetve részeihez. Éppen ez a tény indokolja, hogy a teauruszok szempontjából élesen válasszuk szét ezt a két kapcsolattípust. A hangszóró a rádió része. Ha valaki a hangszóróval foglalkozó irodalmat keresi, nem biztos, hogy elégedett lesz, ha megkapná a rádiók dokumentumait is. Ezt a relációt tehát másként érvényesítjük. A legegyszerűbb megoldás az, ha a teauruszban foglalt információkat közvetlenül a keresőkép összeállításához használjuk fel. A keresőkép leírásánál szerepeltetni lehet azokat a rész vagy egész deszkriptorokat, amelyek a kérdéssel funkcionális összefüggésben vannak. Ha pl. valakit a gépkocsik biztonsági szempontjai érdekelnek, akkor a gépkocsik részfogalma, a fékberendezések részt vehetnek a keresésben.

3.33 ROKON KAPCSOLATOK

Az eddig tárgyalt két relációtípus teljes egyértelműséggel megállapítható minden teaurusz esetében. Vannak azonban a deszkriptorok között más tartalmi kapcsolatok is, amelyek esetenként roppant fontosak, de meghatározásuk sokkal nehezebb, mint az előbbieké. Számos teaurusz az egyéb kapcsolatokat gyűjtőfogalommal „rokon kapcsolatoknak” nevezi, gyakorta közelebbi vagy mélyebb kifejtés nélkül. Talán azért van ez, mert formailag is, tartalmilag is több típusú kapcsolatot foglal magában, és valóban nehéz megmondani adott esetben, hogy a kapcsolat milyen. A kapcsolattípusok megállapítása sem könnyű, hogy ti. egy konkrét eseten kívül ugyanazon típusú kapcsolat következetesen érvényesíthető-e vagy sem. Mivel azonban ezek a relációk a szaktudomány mélyebb tartalmi összefüggéseit tükrözik, elhagyásuk jelentős információvesztéssel járna, így károsan befolyásolná a teauruszt.

A kapcsolatok jellegét mi sem kívánjuk meghatározni. Megközelítőleg arról van szó, hogy más-más szakembernek egy deszkriptorról más-más kapcsolatok, más-más deszkriptorok jutnak eszébe, ez az asszociáció mégsem szubjektív, hanem a szaktudomány tartalmi összefüggéseinek következménye. Ahogyan egy orvosnak egy betegségről eszébe jutnak annak tünetei, vizsgálati módszerei, gyógyszerek, terápiás eljárás stb., ugyanúgy bizonyos összefüggések minden szakterületen kézenfekvőnek látszanak, és semmiképpen nem azonosíthatók az eddig tárgyalt relációkkal. Az alábbi táblázat vázlatosan feltárja ezeket az összefüggéseket, feltüntetve azokat a vonzási pályákat, amely utakon ezek az asszociációk keletkezhetnek. A táblázattal a teaurusz összeállítóinak szeretnénk némi segítséget nyújtani a lényeges kapcsolatok kiemelésében.

A mátrixról sok minden leolvasható. A fontosabb általános összefüggések:

a) Produktumról, termékről lehet asszociálni — ha van — a produktummal együtt járó melléktermékre és arra a folyamatra, amellyel a termék készül. A termék való valamire, asszociált kapcsolatban van tehát a termék céljával. A produktumnak vannak jellegzetességei, tulajdonságai, amelyek kézenfekvően együtt járnak a produktummal. Végül asszociált kapcsolatban lehet hasonló produktumokkal. Példák:

növényi balzsam — gyanta (termék — melléktermék)

műméz — savas invertálás (termék — folyamat)

butadién polimerek — hidegtűrés (termék — tulajdonság)

	Termék	Melléktermék	Módszer, eljárás	Folyamat, művelet	Feltétel, állapot, befolyásoló tényező	Ellentét	Hatás, ellenhatás	Ok	Cél, rendeltetés	Eszköz	Tulajdonság, minőség	Elmélet	Korrelatív	Hasonlóság
Termék	1		1						1		1			1
Melléktermék		1						1			1			
Módszer, eljárás	1		1	1					1	1		1		
Folyamat, művelet	1	1	1		1		1		1	1		1		
Feltétel, állapot befolyásoló tényező			1	1		1	1	1		1	1			1
Ellentét			1					1			1			
Hatás, ellenhatás				1	1	1		1	1	1		1		1
Ok					1	1	1				1	1		
Cél, rendeltetés	1		1	1		1				1				
Eszköz	1		1	1			1		1					1
Tulajdonság, minőség	1	1			1	1			1					1
Elmélet			1	1	1				1					
Korrelatív													1	
Hasonlóság	1				1	1		1			1			

látcső – nagyítás (termék – cél)

b) A melléktermékek asszociálnak az okra, amely létrehozta őket és tulajdonságokra, amely tulajdonságok alapján további kapcsolataik írhatók le. Példák:

gázvíz – szénleptárlás (melléktermék – ok)

melasz – káliumtartalom (melléktermék – tulajdonság)

c) Módszer, eljárás leggyakrabban asszociál termékre, amelynek előállítására használatos, folyamatokra, amelyekben alkalmazzák, azon feltételekre, amelyek esetén alkalmazható, eszközökre, végül arra a tudományos elméletre, amelynek értelmében egy adott szituációban alkalmazható. Néhány példa:

mátrixmódszer — modell (módszer — produktum)

dedukció — felosztás (módszer — folyamat)

stb.

d) Folyamat, művelet asszociál a produktumra, melléktermékekre, módszerre, alkalmazási feltételeire, a hatásra, amelyet a folyamat a tárgyakon kifejt, a célra (amely lehet a produktum), eszközeire, pl. berendezésekre és arra a tudományos elméletre, amelynek értelmében alkalmazható. Néhány példa:

gázlecsapolás — szivattyú (folyamat — eszköz)

hidegalakítás — ridegség (folyamat — tulajdonság)

stb.

e) Feltétel, állapot, befolyásoló tényezők asszociálnak módszerekre, folyamatokra, hiszen legtöbbször ezek feltételei. De asszociálnak ellentétes feltételekre, a hatásokra, amelyek elérése kívánatos az adott feltételek mellett, az állapot okaira, eszközökre, amelyekkel adott állapot elérhető, tulajdonságokra és hasonló feltételekre. Néhány példa:

hőcserélés — sugárvédelem (feltétel — cél)

érdekltség — anyagi ösztönzők (feltétel — eszköz)

stb.

f) Az ellentétes deskriptorok rendszerint asszociatív kapcsolatban állnak egymással, a szembeállítás tudatosítása gyakran jár együtt okok megfogalmazásával és a tulajdonságok, módszerek kiemelésével. Példa:

indukció — dedukció, ellentétük kiemeli a következtetés irányát, módszereit, tulajdonságait.

g) Hatás (ellenhatás) asszociál folyamatra, hiszen a hatás megvalósulása maga is folyamat, feltételekre, ellentétes hatásokra, a hatás okaira és céljára, eszközeire, valamely elméletre és hasonló hatásokra. Néhány példa:

magnetosztatikus — mágneses tér (hatás — eszköz)

prézelés — nyomás (hatás — hasonlóság)

stb.

h) Okok, célok asszociálnak folyamatokra, feltételekre stb. Példa:

képlékenységi — lágyítás (cél — folyamat)

magassági szélmérés — léggömb (cél — eszköz)

stb.

i) A cél, rendeltetés — eszköz kapcsolatokat a leggyakrabban kiemelt, fontos relációként építik be a tezauszba. A kapcsolat azt mutatja, mi mire való, illetve fordítva, mihez milyen eszközök, feltételek szükségesek. Néhány példa:

tezausz — osztályozás

tezaurusz — információátvitel

szeműveg — látásjavítás

Hasonló módon lehetne végigmenni a mátrix minden esetén, de az eddigi példák-ból látható, hogy milyen típusú relációk tartoznak az ún. rokon kapcsolatok körébe. Az esetek között természetesen átfedések is vannak. Hatás lehet folyamat, melléktermék, néha a termelési folyamat egyik célja stb.

A tezaurusznak nem mindig válik javára a rokon kapcsolatok túlságosan tág körű kiépítése, különösen, ha a relációk erőltetettek, ha mindenáron találni akarunk a deszkriptorokhoz kapcsolt fogalmakat. A túlzásoktól az a megszorítás menthet meg, hogy a rokon fogalmaknak is deszkriptoroknak kell lenniük. Említettük az 1. fejezetben, hogy a tezaurusz relációrendszerét a deszkriptor között kell kiépíteni, ezért nem szabad olyan kapcsolt fogalmakat megengedni, amelyek a tezaurusz más helyén nem válnak önálló deszkriptorrá, azaz vezéreszkriptorrá.

Ha a rokon kapcsolatokat közelebbről megvizsgáljuk, két egymástól jól elkülönülő típusokat vehetjük észre.

Az első típus az, amelyik oda-vissza igaz, tehát a reláció szimmetrikus, aRb és bRa egyidejűleg fennáll. Ilyen minden esetben a korrelatív viszony, amely csupán mennyiségi kapcsolatot (gyakorlati együttes előfordulást) jelent. Ilyen — jellege szerint — a hasonlóság és az ellentétes viszony is. Az ilyen típusú relációk a gráfban irányítás nélküli éllel jelölhetők.

Mátrixreprezentációban szimmetrikus reláció esetén

$$a_{ij} = a_{ji} = 1$$

azaz a mátrix i -edik sorának és j -edik oszlopának eleme egyenlő a j -edik sor és az i -edik oszlop elemével.

A másik eset az egyirányú reláció, amelynek megfordítása nem igaz. A reláció aszimmetrikus, a gráfban irányított él reprezentálja.

Gyakorlati értelme a kétféle asszociatív reláció elkülönítésének az, hogy pontosabban tudjuk követni a deszkriptorok kapcsolatának útjait. A pszeudomorfóziáról mindenki kristályokra gondol, de a kristályokról nem jut eszébe feltétlenül a pszeudomorfózia. A pszeudomorfózia kizárólag a kristályok tulajdonsága, érvényessége tehát ebben a konkrét kapcsolatban van. Ezzel szemben ahogyan az explicit alakról eszünkbe jut az implicit alak, ugyanolyan módon jut eszünkbe az implicitről az explicit alak.

A továbbiakban az egyirányú kapcsolatot „jellemző”-nek mondjuk és megkülönböztetjük a kétirányú, szimmetrikus relációtól — az utóbbiakat tekintjük szorosabb értelemben asszociálnak, asszociatív relációnak.

Az egyirányú kapcsolat arra figyelmeztet, hogy egy általánosabb deszkriptor használata egy másik, a kívánt témához pontosabban illeszkedő deszkriptorral helyettesíthető. Ez a deszkriptor egyértelmű függőségi viszonyban van azzal a deszkriptorral, amelyhez hozzá van kapcsolva (pl. a „nagyolvasztó” egyértelműen a „kohászat” témakörébe tartozik, annak, hogy így mondjuk: „segédfogalma”). Az oda-vissza asszociatív kapcsolatban lévő deszkriptor kiegészítő információt ad a másik deszkriptorhoz, ketten (vagy többen) együttesen határozzák meg a koordinált indexelés elveinek megfelelően.

A tezauszra épülő kereső rendszer gyakorlati tapasztalatai elsősorban ebben a két relációtípusban értékesíthetők. A keresés pontosságának és teljességének esetenkénti értékelése, a zaj okának és a várt dokumentumok elmaradása okának elemzésével lehet egyre inkább közelíteni az optimális megoldáshoz. Feltétlenül kívánatos a számszerű, statisztikai adatok gyűjtése és automatikus értékelése is a gyakorlati *használat* folyamán, deskriptorként, dokumentumonként és témánként külön-külön.

3.34 SZINONIMÁK ÉS KVÁZISZINONIMÁK. A HOMONIMIA

A tezauszba szabályozott kifejezések, deskriptorok kerülnek, hogy formailag is, tartalmilag is biztosítsák a keresőnyelvben az egyértelműséget. A szemantikai egységesítés szempontjából két kapcsolattípus elemzését kell elvégezni. Az első a tartalmi egybeesés — a szinonimák elemzése, a másik a formai egybeesés — a homonimák elemzése és ki-küszöbölése.

Az egynyelvű tezauszokra vonatkozó nemzetközi szabvány (ISO 2788–72) egyik definíciója a tezauszt olyan szótárnak nevezi, amely a természetes nyelv és egy szakterület szabályozott, egyértelmű nyelve között biztosítja a kapcsolatot. Ebből a szempontból fontos a szinonima- és homonimakapcsolatok gondos kimunkálása. Sokan — a tezauszoknak ezt a feladatát szem előtt tartva — a tezauszok relációit két fő típusra bontják: az első a természetes nyelv és a tezausz deskriptorai közötti kapcsolatok (ide a szinonimiát és részben a homonimiát sorolják). A többi kapcsolat deskriptorok egymás közötti relációi, mint az alá-fölérendelés, része-egésze kapcsolatok stb.

Szinonimának nyelvi szempontból azokat a szavakat vagy szerkezeteket, kifejezéseket kell tekinteni, amelyeknek alakja, nyelvi burka, kifejezése eltérő, de jelentésük megegyezik, vagyis ha tartalmilag egymást maradéktalanul helyettesíteni tudják.

A tezauszépítésben a szinonimaproblémát úgy lehet megoldani, hogy a szinonimák egyikét — a legelterjedtebb, leghabzóosabb, nyelvíleg legkifejezőbb változatot — deskriptorrá emeljük, a többi változatról utalunk rá. Egyszóval a szinonimák valamelyike helyettesíti a többi, s a többi használatát megtiltjuk. Ha ezt nem tennénk, akkor ugyanazokat az információkat mindegyik helyen külön tárolni kellene.

Szinonimának tekintjük az idegen és magyar nyelvű alakját ugyanannak a fogalomnak: refrakció — fénytörés. Köznapi és tudományosabb elnevezéseket: cethal — bálna. Régi és újabb elnevezés: függvény — leképzés. Eredeti (akár idegen, de elterjedt) elnevezést vagy körülírást ha ez semmiképpen sem kerülhető el: Pancsa Sila — Nemzetközi kapcsolatok öt alapelve (vagy nem körülírt, de kevésbé használt változatban: bandungi alapelvek). Valaminek eleve több neve lehet: Gauss-eloszlás — normális eloszlás. Szinonimának tekintjük az azonos jelentésű alakokat, amelyeket azonos szó különböző változataiban fejezünk ki: permutáló index — permutált index. Különböző helyesírással írt szavak: thesaurus — tezausz.

A szinonimák között típusokat lehet megkülönböztetni az egyezés mértéke szerint. Azoknak a szinonimáknak, amelyekről teljes bizonyossággal nem állapítható meg ez a viszony, kváziszinonimáknak tekintjük. Ide azok a szinonimák kerülnek, amelyek feltéte-

lesen szinonimaként kezelhetők, ha a kereső rendszeren kívül nem is tekinthetők egyenrangú kifejezéseknek. A kereső rendszer realizálása során ezeket éppen úgy tiltott fogalomnak tekintjük, mint a valódi szinonimákat.

A *homonímia* bizonyos értelemben ellentétes jelenség a szinonimával. Azonos hangalak hordoz több jelentést. Az „ár” szó pl. három dolognak, illetve jelentésnek a neve. Szűrőszerszám, áradat (e tekintetben is van konkrét és absztraktabb jelentése), és jelenti az áruk pénzben kifejezett értékét is.

A „horgony” jelenti a hajóknak a vízfenékhez rögzítő szerkezetét, másrészt jelenti elektromos berendezésekben az időleges rögzítés szerkezetét. A homonimák gyakorta jelentésátvitel útján keletkeznek. Valamely új dolgot hasonlósági alapon egy, már létező dologról neveznek el. Az intermedier a genetikában a szülők sajátosságának átmeneti, köztes megjelenését jelenti az utódokban, a kémiai technológiában a több szakaszos előállítási folyamat közbülső termékét jelöli.

Ha valamely deskriptor nem egyértelmű, akkor különbözőképpen járhatunk el:

a) Ha a homonimiában szereplő minden jelentés szükséges, akkor annyiszor vesszük fel a kifejezést deskriptorként, ahány jelentése van, de a deskriptor után értelmezőt teszünk, amely mutatja érvényes jelentését. Pl:

entropia (információelmélet)

entropia (termodinamika)

entropia (filozófia)

Az értelmezőt zárójelbe tesszük. Maga az értelmezés lehet előre megszabott rövidítés is.

Az értelmező azonban a legtöbb esetben elmaradhat, ha a teaurusz témaköre eleve kizárja a többértelműséget. Pl.: egy matematikai teauruszban felesleges lenne a *gyök* kémiai jelentését egyáltalán figyelembe venni. (Fordítva némileg nehezebb a helyzet.)

b) Ha csak valamelyik jelentésű alakja kíváncsú deskriptorként, akkor ezt a megszorítást ugyanolyan módon értelmezővel jelölhetjük. Pl:

integráció (gazdasági)

c) Hasonló a helyzet akkor is, ha a deskriptorként használt fogalomnak nincs szükség a teljes jelentéstartalmára. A deskriptort esetenként is felbonthatjuk értelmezőkké, s vagy arra utalunk az értelmezővel, hogy milyen jelentésű részeket foglal magába, vagy arra, amit nem tartunk kíváncsúnak. Az utóbbira példa:

indikátor (kémia, kivéve adszorpció)

A deskriptor relációrendszere minden esetben a már értelmezővel pontosított jelentéstartalmú deskriptorra vonatkozik.

A harmadik példa már utal arra, hogy a teauruszban mesterségesen is lehet jelentésmegkülönböztetést tenni. Erre akkor kerül sor, ha valamely deskriptor használata túl gyakori és túlságosan sok dokumentumot választana ki a visszakeresés során. Értelmezővel egyértelműen szétválaszthatjuk a jelentéseket. Pl:

gázolaj (termék)

gázolaj (üzemanyag)

Ehhez a megoldáshoz csak szükség esetén szabad folyamodni. Segíthet a terhelésen az alárendelt, illetve részdeszkriptorok jó kiépítése is, olyan osztályozási utasítással, hogy a túl gyakori deszkriptor helyett, ha csak lehet, az alárendelteket, illetve részfogalmakat kell helyettesíteni. A másik mód az ún. fazetták (vetületek) kiépítése. Mesterséges jelentésmegkülönböztetéshez akkor folyamodunk, ha egyúttal a deszkriptor fogalomrendszere is világosan elkülöníthető, nem kell tehát több helyen lényegében ugyanazokat az információkat tárolni.

3.35 KÜLSŐ KAPCSOLATOK

A tezauszznak meghatározott tartalmi köre van, de számos szállal kötődik más, a tezausz számára közvetlenül el nem érhető szakterülethez. Mint ahogyan nem képzelhető el izolált tudomány, ugyanúgy nem létezhet tezausz sem tudományközi vagy interdiszciplináris kapcsolatok feltüntetése nélkül. Azonban ha valamennyi külső kapcsolatot részleteiben feltüntetnénk, akkor a tezausznek nem lennének tartalmi határai. A külső kapcsolatok deszkriptorait tehát egyfelől limitálni kell a valóban fontos kapcsolatokra (amely kapcsolatok feltüntetése kívánatos a tezausz szempontjából), másrészt lehetőleg átfogóan, generikus deszkriptorokkal fejezzük ki őket, hogy ne duzzadjon fel emiatt a tezausz deszkriptoranyaga perifériális fogalmakkal. Pl.: az élesztő deszkriptor pl. kapcsolódik – külső kapcsolat szerint a cukorgyártáshoz (innen nyerik), az élelmiszeripar egyes ágaihoz és a gyógyászathoz egyaránt. Ezek közül elég, ha csak azt tüntetjük fel, ami az adott tezausz számára fontos. A külső kapcsolatokat különböző szinten kell feltüntetnünk. Pl. a hematoxilin külső kapcsolata lehet a textilfestés vagy átfogóbban a textilgyártás, a textilipar is, a „hullámcsapda” alkalmazási területe a híradástechnika. Ezt egy híradástechnikai tezauszban nem szabad külső kapcsolatként feltüntetni, hiszen a szakcsoportba tartozás a tezausz tartalmi körén belül kifejezi ezt a kapcsolatot. Fizikai tezausz a gamma-sugárzásnál külső kapcsolatként megjelöli az anyagvizsgálatot, gyógyászatot. Kémiai tezausz a celofánnál megjelöli a csomagolótechnikát, matematikai tezausz a trendszámításnál megjelöli a statisztikát vagy a futurológiát és így tovább.

A külső kapcsolatok kifejezésére a tezauszban új reláció bevezetését kell megengedni.

3.4 A DESZKRIPTOROK SZÓALAKJA

A deszkriptorok helyesírása lexikográfiai szempontból is megkövetelik az egységesítést. Ebben a tekintetben csupán általános irányelvek adhatók, valamennyi konkrét esetre alkalmazható utasítások azonban nem.

A deszkriptorok helyesírására vonatkozóan az érvényes akadémiai előírások mérvadóak.

A deskriptor szófajai között szinte kizárólag főnevek fordulnak elő. Ha tevékenységet, folyamatot vagy tulajdonságokat kell kifejeznünk, alkalmazzuk — ha csak lehet — a főnevesítés módszerét. (Mégmunkál helyett megmunkálás, hengerel helyett hengerlés, fest helyett festés stb., másrészt fényáteresztő helyett fényáteresztés vagy fényáteresztőképesség, gyors helyett gyorsaság stb.)

A főnevek lehetnek tulajdonnevek is (Kohó- és Gépipari Minisztérium).

A jelzőket lehetőleg ne hagyjuk meg önállóan, csak szókapcsolatban. Még akkor se, ha sokféle szerkezetben szokott megjelenni. Nem sok értelme lenne ugyanis „aktív”, „periodikus” stb. alakú deskriptoroknak, arról nem is szólva, hogy elemezhetetlenek ebben az alakban. De az „aktív szén” már elemezhető.

A jelzők ennek megfelelően több szerkezetben, deskriptorban megjelennek. Ezek a szokásos sorrendben úgyis egymás közelébe kerülnek, például „reaktív fegyverek” után „reaktív motorok”.

A jelzős szerkezeteket természetes sorrendben írjuk. Tehát „ogam írás” és nem „írás, ogam”.

A rövidítések és feloldott alakjaik úgy kezelendők, mintha szinonimák lennének. Az elfogadottabb változatot deskriptorosítjuk, a teljes, feloldott változatról utalót készítünk. Ha a deskriptor a rövidített alak lesz, akkor a rövidítés után a feloldást is megadjuk. Pl:

MKE (Magyar Könyvtárosok Egyesülete)

Deskriptor kezdődhet egyetlen betűjellel. Ha az idegen betűjel pl. (γ -sugárzás), akkor a jelet oldjuk fel:

γ [gamma]-sugárzás

Ha a deskriptor számmal kezdődne, akkor a számot hátra írjuk. Pl. 21–30. kA zárati áram helyett Zárati áram 21–30 kA.

Több tagú deskriptorok besorolási helye a betűrendben, a természetes sorrendben leírt deskriptor első szava. Ezen a helyen szerepel a deskriptor szócikke, tehát szinonimái, relációrendszere stb. Az összetett deskriptor többi tagjának helyéről akkor utalunk, ha a használata valószínű. Az inverz alakok — kéttagú deskriptor esetén — vagy a permutált alakok — több tagú deskriptor esetén — tehát úgy kezelendők, mintha szinonimák lennének. Az utalót azonban csak akkor készítjük el, ha az invertált, illetve permutált alakváltozatok is értelmes, használt variánst nyújtanak.

Szélsőséges példát hozva: „nagyfeszültségű osztott gyűjtősínes rendszer” ciklikusan permutált alakváltozatai a következők lennének:

Osztott gyűjtősínes rendszer, nagyfeszültségű;

Gyűjtősínes rendszer, nagyfeszültségű osztott;

Rendszer, nagyfeszültségű osztott gyűjtősínes;

A deskriptort csak a nagyfeszültségű alá soroljuk be, a többi változatról akkor készítünk utalót, ha a kérdés abban a formában is felmerülhet. A példa esetén utaló kell az

Osztott gyűjtősínes

Gyűjtősínes rendszer,

változatokról, de nem szükséges a Rendszer kezdőbetűről, mivel nem valószínű, hogy bármilyen témát valaki a „Rendszer” alatt keressen, még ha a tezausz a rendszerelmélet tárgykörében készülne is.

A tezausz nyomtatott változatában a deskriptort minden szócikkben nagybetűvel vagy tipográfiaileg megkülönböztethető módon fogjuk írni, a kapcsolatban lévő deskriptorokat és utalókat kisbetűkkel (vagy kisebb betűkkel).

3.5 FAZETTÁK (VETÜLETEK)

A fazettáknak hasonló szerepük van, mint a szakcsoportjelölésnek, csak hogy nem tematikailag, hanem általánosabb fogalomkategóriaként osztja csoportokba a deskriptorokat. A fazetta a fogalmi metszetet, a témához való viszonyulás általános típusát adja.

A fazetták felhasználásán

- alapulhat rajtuk önálló osztályozási eljárás;
- nem fazettás osztályozási rendszerben fazettás műveletek bevezetése finomíthatja az osztályozást;
- a deskriptor relációrendszerének meghatározásában és a deskriptorelemzésben szellemi segédeszköz lehet;
- végül elősegíti a túlságosan gyakran használt deskriptorok tehermentesítését azzal, hogy e deskriptorok a fazetták szerint szétbonthatók.

Egy deskriptor egyidejűleg több fazettába tartozhat. A kalapács egyrészt tekinthető végterméknek, másrészt eszköznek. A titrálás lehet egy tanulmány tárgya, ahol erről a módszerről van szó, és lehet más témájú tanulmányban módszereljárás, amelyben valamit titrálással végeznek.

A tezauszban a deskriptor mellett fel lehet tüntetni a deskriptor vetületét. Az osztályozásban – amikor a fazetták használatára kerül sor – ugyanazon fazettákkal végzünk műveleteket, a tezausz egyéb információitól függetlenül. Ez már osztályozási munka: a fazetta konkrét esetét kell kiválasztani, nem általában a deskriptor lehetséges fazettáját.

Általánosságban az alábbi fazetták, vetületek használatosak.

- Objektum, dolog,;
- Folyamat, cselekvés, tevékenység;
- Eszköz, eljárás, módszer;
- Jellemző jegy, tulajdonság, sajátosság;
- Cél, produktum, termék, rendeltetés, eredmény;
- Ok, feltétel;
- Elméleti kifejezés.

„Általában a jó regulátornak lényeges jellemzője, hogy meggátolja a sokféleség áramlását a zavaroktól a lényeges változókhoz.”

(W. Ross Ashby)

4. A TEZAUROSZKÉSZÍTÉS MENETE

A tezauszok készítését — egy lehetséges változat szerint — nagy vonalakban az alábbi főbb szakaszokra lehet bontani:

1. A célok meghatározása és a tervezés szakasza.
2. Az induló szóanyag összegyűjtése.
3. A szóanyag elemzése és a legfontosabb deskriptorok kiválasztása.
4. A legfontosabb deskriptorok relációinak gráfszerű kifejtése.
5. A deskriptorcikkek kialakítása.
6. A tezausz fő részének szerkesztése.
7. A tezausz hierarchikus részének szerkesztése.
8. Ellenőrzés, modellkísérletek.

A nyolc munkaszakasz — szintén csak nagy vonalakban — logikai és időbeli egymásutániséget jelent. Az egyes munkaszakaszok műveleti bontásban azonban párhuzamos munkavégzést is lehetővé tesznek, illetve a konkrét célkitűzésektől függően módosulhatnak. Modellkísérletekre például sor kerülhet akár az induló szóanyag összegyűjtése után is, hiszen számos visszakereső rendszer működik ún. ellenőrizetlen kifejezésekkel (uncontrolled vocabulary).

Az alábbiakban egy lehetséges konkrét tezauszkészítési munkamenetet írunk le, részletes műveleti jegyzékkel.

A munkamenet leírása elég szűkszavú, ezért az egyes részek után némi magyarázatot adunk. Hangsúlyozzuk, hogy az itt vázolt megoldás *csak egy a lehetséges alternatívák közül*, amelyen esetenként változtatni lehet. A kitűzött cél realizálható más sorrendben is.

Ismétlődő műveleteket csak egyszer írunk le. Az elemzőkártyákkal végzett műveletek pl. annyiszor ismétlődnek, ahány kártya van. Utalunk azonban arra, hogy milyen műveleti lépéseknek kell ismétlődnie.

A műveleti leírás a tezauszkészítés manuális változatára vonatkozik, a számítógépek belépésére csak a tezauszkészítés végén kerül sor. A műveleti bontás azonban kellően részletes, így eldönthető, hogy mely fázisokban lehet számítógépi segítséget igénybe venni.

A munkamenetben helyenként vannak feltételes lépések is, amelyek egyes esetekben átugorhatók.

A műveleti sorrend nagyjából időrendet is jelent, de az egyes munkaszakaszok sok helyen párhuzamosan vagy adagokban is végezhetők. Ez általában a helyi adottságoktól függ, ezért nem vehetjük mindenütt számításba.

A továbbiakban sorra vesszük az egyes szakaszokat, megadjuk a műveleti bontást, és konkrét megjegyzéseket fűzünk az egyes műveletek elvégzéséhez, sorrendjéhez stb.

4.1 A CÉLOK MEGHATÁROZÁSA ÉS A TERVEZÉS SZAKASZA

Ez a munkaszakasz főleg az első teaurusz kidolgozása idején jelent nagy idő- és energiabefektetést, amíg a kellő teauruszépítési tapasztalatok össze nem gyűlnek. A későbbi teauruszok kidolgozása idején az itt felsorolt tennivalók némileg módosulnak: önálló új koncepció kidolgozása helyett az előző teauruszok készítésével és használatával kapcsolatos tapasztalatokra kell nagyobb figyelmet fordítani.

Az egyes munkafázisok ezen a szakaszon belül:

4.1.1 A teaurusz fő szakterületének és felhasználási lehetőségeinek tisztázása. Milyen szomszédos szakterületekre kell tekintettel lenni? Milyen dokumentumtípusokra és milyen szolgáltatásokra terjedhet ki a teauruszra épülő kereső rendszer?

4.1.2 A teaurusz általános szerkesztési elveinek meghatározása.

4.1.3 Az anyagi és szellemi erőforrások számbavétele.

4.1.4 A munkában részt vevő személyek kiválasztása, feladatkörük megállapítása, a módszerek és a felelősség meghatározása.

4.1.5 A teaurusz tartalmi körének részletes kidolgozása és a tartalmi arányainak meghatározása. A fő és határterületei szakcsoportok meghatározása.

4.1.6 A teaurusz szerkezetének részletes megtervezése. Döntés: a teaurusz fő-része milyen adatokat tartalmazzon, milyen relációk megkülönböztetése indokolt, milyen legyen a szakrendszer, a fazettaszerkezet, milyenek legyenek a kódok és a jelölésrendszer, mely fokozatokban szükséges a számítógép igénybevétele.

4.1.7 A teauruszkészítés részletes munkamenetének kidolgozása, munka-, idő- és költségigényének felmérése.

4.1.8 Az elkészült tervezet megvitatása, esetleges módosítások végrehajtása, jóváhagyása.

4.1.9 A tervezet sokszorosítása, szétosztása a résztvevők között és — szükség esetén — a képzés megszervezése.

Megjegyzések:

Külön műveletként szerepel (4.1.6 és 4.1.7) a teaurusz részletes megtervezése és a részletes munkamenet kidolgozása. Ez a két fázis részben párhuzamosan is végezhető, ha

4.1.2-ben eléggé világosan tisztázódtak az alapelvek. A részletes tervezés erősen összevont munkafázis. Lényegében a felsorolt témában (szakrendszer, fazetták, relációk stb.) tanulmányok elkészítését jelenti, melyek kellő részletezettség esetén alkalmasak döntések meghozatalára is.

4.2 AZ INDULÓ SZÓANYAG ÖSSZEGYÚJTÉSE

A munka gyorsítása érdekében a szógyűjtés munkafázisa elkezdődhet a tezaurusz általános tervének elkészülte után, a részletes terv megszületése előtt. Ebben a fázisban ugyanis már tisztázottak mindazok a szempontok, amelyek a szógyűjtéshez szükségesek.

Az egyes munkafázisok ebben a szakaszban:

4.2.1 A szógyűjtés lehetséges forrásainak számbavétele, a szóba jöhető források kritikai elemzése.

4.2.2 Döntés a feldolgozásra érdemes források kiválasztásáról és a feldolgozás módszereiről (gépi vagy kézi feldolgozás, teljes vagy részleges stb.).

4.2.3 A források beszerzése, xeroxozása, megbízatások elkészítése.

4.2.4 Szógyűjtő bizonylat tervezése és sokszorosítása, illetve — ha ennek előfeltételei megvannak — a gépi feldolgozás előkészítése.

4.2.5 A források szó- és kifejezésanyagának előzetes válogatása. Döntés: a válogatás szempontjainak szabályozása.

4.2.6 A szógyűjtés teljes körű megszervezése.

4.2.7 Az idegen nyelvű források fordítása.

4.2.8 A fordítás szakmai ellenőrzése és javítása.

4.2.9 A különböző forrásokból származó anyag betűrendbe osztása.

4.2.10 A duplikátumok kiiktatása, a kiiktatott bizonylatokról a szükséges adatok rávezetése a megmaradó bizonylatra.

Megjegyzések:

A 4.2.5 és 4.2.6 munkafázis felcserélhető. Mi feltételeztük, hogy van olyan szakember, aki az idegen nyelvű szóanyagból is el tudja végezni a kijelölést. Mindenesetre ez a megoldás gazdaságosabb és gyorsabb is, hiszen feleslegessé teszi sok-sok idegen kifejezés fordítását.

A 4.2.9 és 4.2.10 műveletre azért kerül sor, mert az induló szóanyag általában több forrásból származik, így elkerülhetetlen, hogy egy-egy kifejezés többször szerepel. A duplumpéldányokat tehát ki kell iktatni, viszont a megmaradt egyetlen kartonon jelölni kell a többi forrást is. Ha — fordítások esetén — idegen nyelvű megfelelők is szerepelnek a cédulákon, akkor ezeket is átvezetjük a megmaradt kartonokra.

Két dologra nincs mód ebben a fázisban. Először: nem lehet gyakoriságot számolni, mivel viszonylag kevés forrás felhasználása esetén ezek a várhatóan alacsony számok nem jellemeznének semmit, és nem a fogalomra, hanem a kifejezésformára vonatkoznak. A feltüntetett források száma egyébként mutatja a gyakoriságot, illetve — pontosabban — azt, hogy hány forrásból származik ugyanaz a szóalak. Másodszor: nem lehet ezen a szinten a szinonimákat kiiktatni, mivel a betűrend erre nem alkalmas. Mód van viszont arra, hogy egyes kifejezések különböző összetételekben szereplő változatait észleljük, ha a betűrendben összekerülnek. (Pl. egymás mellé kerül: munkaerő, munkaerő-szükséglet, munkaerő-kapacitás, munkaerő-tervezés, munkaerőképzés stb.) Összevonásokra még sincs mód, mivel ezt a lényegében technikai jellegű munkafázist középkáder végzi, s ezt a szelekciót aligha lehet rábízni.

4.3 A SZÓANYAG ELEMZÉSE ÉS A LEGFONTOSABB DESZKRIPTOROK KIVÁLASZTÁSA

Ennek a szakasznak lényegében az a célja, hogy a tezaurusz magját s egyúttal a profilját is meghatározza.

Az egyes munkafázisok:

4.3.1 A bizonylatokra a deskriptorok ráírása — idegen nyelvű forrásanyag esetén —, az idegen nyelvű megfelelő beírása.

4.3.2 A szakcsoport megállapítása, a gyűjtőlapoknak szakcsoportok szerinti szétosztása, és a szakcsoportjelzet beírása (osztályozás). Döntés: a feleslegesnek ítélt és a több szakcsoportba besorolható gyűjtőlapok különválasztása.

4.3.3. A kihagyásra javasolt fogalmak betűrendes listájának elkészítése.

4.3.4 A több szakcsoportba sorolható fogalmak felülbírálata, betűrendes listájuk elkészítése, és a döntésnek megfelelően a szakcsoportjelzetek beírása.

4.3.5 A több szakcsoportba sorolt fogalmak gyűjtőlapjának sokszorosítása a szakcsoportok számának megfelelő példányban, egy-egy szakcsoportszám aláhúzása és a gyűjtőlapoknak beillesztése saját szakcsoportjába.

4.3.6 Szakcsoportonként a kellő bontottság ellenőrzése (350—400 gyűjtőlapnál több ne kerüljön egy szakcsoportba).

4.3.7 A fazetták megállapítása és beírása. Ha nagy a szakcsoport, a szakcsoporton belül végezzük el a fazetták szerinti bontást.

4.3.8 A szakcsoporton vagy a fazettákra bontott szakcsoporton belül betűrend kialakítása, és mindegyik csoportról gépelt lista készítése.

4.3.9 A csoportokról készült listák sokszorosítása, azok szétosztása, illetve szétküldése a rokon területek elemzői és külső szakértők számára.

4.3.10 A listák véleményezése: reprezentálják-e kellően a szakcsoportot, megfelelő-e a bontás, eléggé egységes-e a terminológia? Módosító javaslatok beküldése.

4.3.11 A beérkezett javaslatok elbírálása. Egyedi kiegészítések, javítások, elhagyások elvégzése. A kihagyott fogalmakról betűrendes lista készítése.

4.3.12 Döntés: szükség van-e egyes szakcsoportok anyagának nagyobb mérvű kiegészítésére? Ha igen, a megfelelő források és módszerek meghatározása és vissza 4.2.5-be.

4.3.13 Döntés: van-e teljesen használatlan szóanyag. Ha igen, vissza 4.2.1-be.

4.3.14 A szakcsoport központi fogalmait képviselő legáltalánosabb kifejezés-anyag kiválasztása.

Megjegyzések:

A 4.3.6 műveleti sorszámmal jelölt tevékenység azt a célt szolgálja, hogy a szakcsoportba sorolás eredményeként átfogható mennyiségű cédula kerüljön valamennyi szakcsoportba. A cédulák további elemzése során ugyanis az elemző szakembernek át kell tekintenie az egész szakot. Általában nem is kerül 3–400-nál több karton egy-egy elemzőhöz. Ha kivételesen mégis előfordul, akkor szakcsoporton belül a fazetták nyújtanak további bontási lehetőséget. (Annak nincs akadálya, hogy az elemzők ezt megtegyék minden szakcsoport esetében. Ez egyéni munkastílus kérdése.) A szakcsoport cédulamennyiségének jobb áttekintésére szolgál a következő munkaszakasz.

A szakcsoporton, illetve a fazettákon belüli betűrend (4.3.8) nem lehet mereven előírt utasítás. Célja egy-egy keresett szó cédulájának gyorsabb fellelése. De az elemzés során ez a rend úgyszólván többször megbomlik.

Ügyelni kell arra, hogy a több – vagy minden – szakcsoportba tartozó, vagy azokat érintő ún. általános deskriptorok külön csoportot alkossanak (4.3.4). Ennek elemzését a teaurusz teljes tematikáját áttekintően ismerő szakember végezheti el, pl. a területtel ismerős osztályozásban jártas könyvtáros vagy teaurusz-szakértő. Mivel az általános szakcsoport kifejezéseivel az összes többi szakcsoport érintkezik, a megfelelő lista minden elemző számára szükséges. Legcélszerűbb szellősen írt jegyzék formájában készíteni, hogy a jegyzékbe újabb kifejezések is beírhatók legyenek.

4.4 A LEGFONTOSABB DESZKRIPTOROK RELÁCIÓINAK GRÁFSZERŰ KIFEJTÉSE

Ennek a munkaszakasznak az a célja, hogy a teaurusz magját képező deskriptorállomány relációrendszerét összefüggéseiben kidolgozza.

Az elemzésnek azzal kell indulnia, hogy a szakember a biztosan feleslegesnek ítélt kifejezéseket kiiktatja – ezek nem vesznek részt a további elemzésben. A kiiktatott kifejezéseket azonban egy jegyzékre írja, vagy a kartonokat félreteszi. Ezek a kartonok a

későbbiekben — pl. a tezaurusz továbbfejlesztése során — hasznosak lehetnek. Előfordulhat, hogy az elemzés későbbi menetében néhány kifejezést ő maga is szükségesnek ítél.

Az egyes munkafázisok:

4.4.1 A 4.3.14-ben kapott legáltalánosabb fogalmak elemző lapra írása. A fogalom generikus felbontásának irányított gráffal (nyilakkal) való felrajzolása: alárendeltjei (speciális esetei) a szükségesnek tartott hierarchiaszintig. Fölrendeltje vagy fölrendeltje egyetlen hierarchiaszintig. Az alá-fölrendelésnél mindig fazettán belül kell maradni. (A nyilak pl. kék színnel rajzolandók.) A nyilak iránya az alárendelt felé mutat.

4.4.2 A fogalom felbontása egész-rész relációban, szintén a szükségesnek tartott szintig. (A nyilak pl. piros színnel rajzolandók.) A nyilak a részfogalmak felé mutatnak. Eredmény: színezett gráf, amely egyrészt alá-fölrendelési, másrészt rész-egész reláció szerinti bontást fejez ki. A 4.4.1 és 4.4.2 műveletek megismétlődnek a 4.3.14-ben kiválasztott valamennyi fogalomra.

4.4.3 A gráfba belerajzoljuk az egyéb relációban álló kifejezéseket, most már a lapon szereplő valamennyi fogalomra. A nyíl iránya az egyéb relációkat realizáló fogalmak felé mutat. (Alkalmazzunk új színt, pl. zöldet.)

4.4.4 A gráf kifejezésanyagának egybevetése a gyűjtött szóanyaggal (kartonokkal). Az eltérések és azonosságok megállapítása.

4.4.5 A gráfon olyan kifejezés szerepel, amely nincs a kartonokon. Döntés: szükséges-e vagy kiiktatandó. Ha szükséges, akkor karton készül róla. A karton eddig kitöltetlen rovatainak kitöltése: szakcsoportba sorolás, fazetták megállapítása stb. Ha a döntés nemleges, a szó a gráfról törlendő, rákerül a kiiktatott szavak jegyzékére.

4.4.6 A kartonon szerepel olyan kifejezés, amely hiányzik a gráfról. Döntés: szükséges-e vagy törlendő. Ha szükséges, akkor rárajzolandó a gráfra. Ha törlendő, rákerül a kiiktatott szavak jegyzékére.

4.4.7 Ha a beiktatott fogalmak más szakcsoporthoz is tartoznak, ezek eljuttatása az illetékes szakcsoportok elemzőihez. Az új kifejezésekkel ismétlődik 4.4.4.

4.4.8 A gráfok lapjainak sokszorosítása, eljuttatása minden elemzőhöz és más szakértőkhöz. A gráfok észrevételezése, értékelése, javaslatok a módosításra.

4.4.9 A javaslatok értékelése, a szükséges módosítások átvezetése. Kialakul a gráfok végleges formája.

Megjegyzések:

Ez a munkaszakasz tulajdonképpen deduktív módszer bevezetése az elemzés során. Az elemzőt arra kényszeríti, hogy ne hagyatkozzék az elemzés során valamilyen mechanikus elemzési algoritmusra, hanem szaktudását aktivizálja. Egyben a szógyűjtésnek elegendhetlen ellenőrzése is. De az elemző szakértő számára is segítséget jelent: maga építi fel azt a vázat, amely áttekintést biztosít számára az anyagról egész elemző munkája során. Azzal, hogy a szakcsoport jelzései közül kiemeli a legáltalánosabb fogalmakat, majd megkísérli ezeknek a fogalmaknak a felbontását felülről lefelé — függetlenül attól, mi sze-

repel a többi cédulán — viszonyítási alapot teremt a szóanyag értékeléséhez. A módszer önálló szógyűjtési eljárásnak is felfogható.

Az eljárás során a tovább-bontást először az alárendelt fogalmakkal kezdjük, és a rendszer egésze szempontjából érdemes. Sok fogalom felosztása többféle logikai alapon elvégezhető, bár a legtöbb esetben a fogalmaknak csak egy kis hányadára jellemző egy-egy adott szinten a felosztás szempontjainak keveredése. A szerkezetet általában *irányított fa* formájában lehet felrajzolni. Ezt követően hasonlóan kell eljárni a „rész” fogalmakkal, amelyekhez képest a felbontásra kerülő fogalom „egész”. Ez szintén hasonló szerkezetet eredményez, amit célszerű az áttekinthetőség biztosítása érdekében más színű vonalakkal feltüntetni. Ezután — szintén más színt alkalmazva — a kifejezésekhez hozzákötődők az egyéb kapcsolattal fogalmak. Végül a felbontásra kerülő fogalom fölérendeltjét, illetve egészét kell felrajzolni, ha ilyen egyáltalán létezik és beletartozik a szakcsoportba.

A fastruktúra felrajzolásánál célszerű szigorú logikai elveket követni, különben sem a feldolgozott szóanyag aránytalanságai, sem következetlenségei nem derülnek ki. Ezt a megszorítást tükrözi az az előírás, amely kimondja, hogy a felosztásnál fazettán belül kell maradni.

Ha a fastruktúrák elkészülnek a legfontosabb fogalmakra, ezzel rendelkezésünkre áll a szakcsoportok fogalomrendszerének egy olyan váza, amelyen a kartonok kifejezései elhelyezhetők, és fordítva, amelyek megkönnyítik a kartonok rovatainak kitöltését. Egybe kell tehát vetni a fastruktúrák fogalmait a kartonok fogalmaival. A különbségeket ki kell küszöbölni. Ha a fastruktúrákon szerepel olyan kifejezés, amely hiányzik a kartonokról, akkor ezekről kartont szükséges lefektetni — vagy ki kell iktatni őket. Ha kartonok kifejezései hiányoznak a faszerkezetből, akkor beépítésükről gondoskodunk vagy törlésükről. A kiiktatott kifejezéseket — esetleges későbbi használat érdekében — szintén listázni kell. Ha az újonnan kapott fogalmakat egyben az általános fogalmak közé is be kell iktatni, akkor erről külön jegyzék szükséges, amelyekkel az általános deskriptorok csoportja kiegészül.

4.5 A DESZKRIPTORCIKKEK KIALAKÍTÁSA

Ennek a munkaszakasznak az a célja, hogy a teljes kifejezésanyagot hozzáillessze a gondosan kialakított teauruszmaghoz.

Az egyes munkafázisok:

4.5.1 Szinonimák és kváziszinonimák összefüggése. Forrás: kartonok, gráfok. Döntés: a szinonima fokozatáról. A döntés eredményeként a szinonimákat és kváziszinonimákat beírjuk a kartonok megfelelő rovatába.

4.5.2. Kváziszinonimáról jegyzék. Célja: későbbi elemzés a teaurusz továbbfejlesztésekor. A jegyzék tartalma: kváziszinonima, deskriptora, szakcsoportjelzete.

4.5.3 A deszkriptor szóalakjának meghatározása. Döntés: a szófaj, helyesírás, egyes szám-többes szám kérdésében. Döntés: több tagú deszkriptor esetében az inverz és permutált alakváltozatokról, amellyel a deszkriptorra való utalás helyeit jelöljük ki. Kiépítjük az értelmezőket.

4.5.4 Szakcsoportjelzetek és fazettajelek ellenőrzése, esetleg módosítása.

4.5.5 Főlé- és alárendelt fogalmak rávezetése a kartonokra. Hierarchiaszint mind fölfelé, mind lefelé csak egyfokozatú. Ha több genus van, aláhúzással jelölendő az ún. fő genus. Ha több szempontú felbontást alkalmazunk az alárendelt fogalmaknál, aláhúzással jelöljük a fő felosztás kifejezéseit. Célja: az aláhúzás adja azt a generikus kapcsolatot a lehetséges többféle közül, amelynek alapján a deszkriptor a teaurusz hierarchikus részében szerepelni fog.

4.5.6 Egész-rész kifejezések rávezetése a kartonra. Ismétlődik a 4.5.5 pont, az aláhúzással együtt.

4.5.7 Rokon fogalmak rávezetése. Döntés a kapcsolatok egyirányú vagy kétirányú voltáról.

4.5.8 Külső kapcsolatok (felhasználás) meghatározása.

4.5.9 A kitöltött karton ellenőrzése. A „Megjegyzés” rovat kitöltése szükség esetén.

4.5.10 A kartonok és gráfok lapjai visszakerülnek.

4.5.11 A kartonok szakfordítókhoz juttatása. Idegen nyelvű ekvivalensek beírása.

Megjegyzések:

A kartonok kitöltése a már rendelkezésre álló „kitöltési utasítás” szabályai definíciói értelmében mennek végbe. A kitöltés kritikus pontját az ún. kváziszinonimák jelentik. Ha ugyanis ezek a teaurusz tiltott fogalmai lesznek, akkor ugyanúgy minősítendőek, mint a szinonimák. Ha megjelennek deszkriptorként is, akkor nem lehetnek szinonimák kvázi sem. Ezért a kváziszinonima rovatba a fogalmakat kell beírni, amelyeket egyelőre szinonimaként szükséges kezelnünk, de a teaurusz későbbi fejlesztése során talán módosulhat a szerepük. Egyébként ami a kváziszinonima rovatba kerül, azt a teauruszban — amíg annak későbbi revíziójára nem kerül sor — szinonimaként kell kezelni, tehát tiltott kifejezésként, amit a deszkriptor helyettesít.

Mivel elképzelhető, hogy később a kváziszinonimákkal további elemzések válnak szükségessé, érdemes ezeket külön jegyzékre írni.

Az alá-főlérendelési rovat, illetve az egész-rész rovat kitöltése során a felosztás polihierarchikus. Számítva arra, hogy a későbbiekben ki kell építeni a teaurusz hierarchikus részét (4.7.), amely viszont csak monohierarchiát tűr meg, a felosztások közül ki kell emelni az ún. fő felosztást, mint legfontosabbat. Ennek nincs jelentősége a teaurusz betűrendes részében, és kizárólag azt jelenti, hogy a teaurusz hierarchikus részében a deszkriptor a kiemelt felosztásban jelenik meg.

A 4.5.7 alatt leírt rovatok kitöltése nem lehet mesterkélt, tehát szükséges a rovatok kitöltéséhez mindenáron ragaszkodni.

4.6 A TEZAUROSZ SZERKESZTÉSE

Lényegében ebben a szakaszban alakul ki a tezaurusz főrésze (betűrendes rész).

Az egyes munkafázisok:

4.6.1 A cédulák szoros betűrendbe sorolása.

4.6.2 Ugyanazon kifejezés esetleg több kartonjának összevonása egyetlen kartonra.

4.6.3 Szinonima- és kváziszinonima-utalások kialakítása. A szinonimák és kváziszinonimák betűrendi helyére utalót iktatunk. Ha ezen a helyen megtaláljuk a szinonimát deszkriptornak javasolva, akkor döntés arról, melyik változat legyen deszkriptor és melyik szinonima. A fogalombokokat egyeztetjük és a kívánatos változtatásokat elvégezzük.

4.6.4 Alá-fölrendelési reláció kiépítettségének ellenőrzése. Ellenőrzés: az alárendeltnek van-e kartonja. Ha igen, szerepel-e ott az éppen elemzés alatt álló deszkriptor fölrendeltként? Ha nincs az alárendeltnek kartonja, dönteni kell, maradjon, vagy ki kell iktatni. Ha marad, kartont nyitunk számára, a kartont kitöltjük. Ha kiiktatásra kerül, feljegyezzük a kiiktatott fogalmak jegyzékére. Ha az alárendelt helyén szinonima-utalót találunk, akkor az alárendelt szóalakját megváltoztatjuk az utalónak megfelelően. Ugyanúgy járunk el a fölrendeltek esetében is.

4.6.5 Egész-rész relációk utalásrendszerének kiegészítése. A 4.6.4 ponttal azonos módon járunk el.

4.6.6 Külső kapcsolatok relációrendszerének ellenőrzése. Eljárás azonos a 4.6.4 ponttal.

4.6.7 Rokon fogalmak relációrendszerének ellenőrzése. Az eljárás azonos a 4.6.4 pontban foglaltakkal.

4.6.8 A redundancia csökkentése, a tezaurusz tömörítése. Minden deszkriptort megkísérlünk kifejezni a többi deszkriptor koordinációjával. Ha ennek folytán egy deszkriptort kiiktatunk, helyére utalás kerül. Az utalás tartalma: kiiktatott deszkriptor, az utalás jele, lásd (kombinációban) azok a deszkriptorok, amelyek koordinációja kifejezi az adott deszkriptort, köztük esetleges logikai jelek.

4.6.9 A tezaurusz betűrendes részének ellenőrzése.

4.6.10 Formai deszkriptorok kialakítása.

Megjegyzés:

A tezaurusz szerkesztése a szinonimák utalásrendszerének kiépítésével kezdődik el. Ez a munkaszakasz megelőzi a többi. A későbbiekben ugyanis, amikor az alá-fölrendelési stb. relációk fogalmait ellenőrizzük, már szükség van annak ismeretére, hogy valamely fogalom melyik szóalakja deszkriptor és melyik tiltott használatú, azaz szinonima. A szinonima-utalók beépítése a lehető legegyszerűbben megy végbe. A betűrendbe tett kartonok közül elő kell venni a legelső. Ha azon szinonima vagy kváziszinonima

rovatban szerepel egy kifejezés, azt egy kartonra kell felírni utalás formájában. Pl.

haszon *lásd* nyereség,

illetve jellel:

haszon L nyereség

A cédulát azonnal beosztjuk a megfelelő betűrendi helyre a szinonima alapján. Nyilván a szinonimáról kell eljutni a deszkriptorhoz és nem fordítva. Példánkban tehát a karton a „haszon” betűrendi helyére kerül. A későbbi szerkesztési fázisok bármelyikében is előfordul a haszon kifejezés, már tudni lehet, hogy helyette a nyereség deszkriptort kell használni.

Megtörténhet, hogy amikor az utalót be akarjuk építeni, annak helyén megtaláljuk a szinonimáját, amelyet egy másik elemző deszkriptornak minősített. Ez esetben döntünk, hogy valóban melyik alakváltozat maradjon deszkriptor és melyik legyen szinonima, az utalást ennek megfelelően végezzük el és írjuk át a cédulákat.

A teauruszok a szinonimákat nemcsak a szinonima helyén tüntetik fel, hanem a deszkriptor után is, közvetlenül utána, jelezve, hogy a deszkriptor mely kifejezések helyett szerepelhet. Erre szükség van a teaurusz kiírásánál, ebben a munkaszakaszban azonban nincs teendő, hiszen a szinonima szerepel a deszkriptor kartonján, a megfelelő rovatban. Az utalás alakja:

Nyeresség H haszon (H = helyette, helyettesít)

Az alá-fölérendelési kapcsolatokat (4.6.4) úgy ellenőrizzük, hogy céduláról cédulára haladva megállapítjuk: a fölérendelt, illetve alárendelt deszkriptorként feltüntetett fogalmak szerepelnek-e a kartonok között. A különböző relációk a teauruszban felvett deszkriptorok között létesítenek finoman kimunkált kapcsolatrendszeret. Az viszont nem lehet, hogy egy deszkriptor relációba kerüljön olyan kifejezés, amely nem szerepel a teauruszban. A teaurusz ebből a szempontból nem lehet nyitott. Ha tehát pl. egy alárendelt deszkriptor nem található meg a maga helyén vezérdeszkriptorként, be kell iktatni, illetve dönteni kell, hogy szükség van-e rá vagy sem. Ha a döntés az utóbbi, akkor mint alárendelt deszkriptort is ki kell iktatni. Ha viszont felvesszük a deszkriptorok közé, akkor elemzését is el kell végezni (vissza 4.5.1-hez).

Valószínű, hogy az így beiktatott új deszkriptorok kapcsolatrendszere jórészt felderíthető magából az eddigi kartonokból, ezeket legfeljebb ki kell egészíteni. Túl szigorú tehát az az előírás, hogy újrakezdődjék egy elemzési fázis. Lényegében erről van szó, de nem merev értelemben. Ha ugyanis újra megkapják az elemző szakértők az új kartonokat, ismét újabb és újabb kifejezések kerülnek be a rendszerbe, és a folyamatnak nincs vége. Ezeket az új elemzéseket tehát már a meglévő szóanyag segítségével kell elvégezni.

Lényegében hasonlóan járunk el a többi relációval is. Amíg azonban a logikai és egész-rész viszonyok kimunkálása a teaurusz deszkriptorainak belső, tartalmi, lényegét, nélkülözhetetlen kapcsolatrendszerét tárja fel, addig a többi kapcsolat (asszociált, jellemző, felhasználási terület) már lazább, képlékenyebb, elmosódottabb viszonyokat fejez ki. Sőt, ezeken keresztül ki is léphetünk a teaurusz tartalmi köréből és távoli szakterületekre is eljuthatunk. Sok lenne tehát a teaurusz tartalma szempontjából a külső, távolabbi

kifejezés. Éppen ezért ezek részletes elemzését nem kell elvégezni, ha ilyenek újonnan való beiktatásáról lenne szó. Ha nem ezt tesszük, akkor – a „minden mindenhol összefügg” elv értelmében – számtalan sok szakterület deskriptorait is elemezni kellene. Márpedig a munkálatokat valahol el kell határolni. (Fejezetünk mottója – analógiaként – erre is rámutat.)

Sarkalatos része – önmagában nagyon nehéz munkafázis – a 4.6.8 sorszám alatt leírt eljárás. Erre akkor kerülhet sor, amikor a betűrendes rész elkészült. Előfordul ugyanis, hogy a deskriptorok között *számos olyan van, ami más deskriptorral kifejezhető*. Egyszerű példával: tegyük fel, hogy szerepel deskriptorként a biológia, kémia, csillagászat, biokémia, asztrokémia stb. A biokémia előállítható a biológia és a kémia, az asztrokémia a csillagászat és kémia metszeteként. Az a fajta terjengősség, amelyet a probléma takar, sohasem iktatható ki teljes mértékben a tezaurszból. Csökkentésére azonban törekedni kell. Azért van erre szükség, mert az osztályozás és a keresőprofil leírása számára ez a terjengősség alternatív megoldásokat rejt magában, következtetlenségekkel járhat. Ha például az egyik esetben a biokémiát mint „biokémiát” osztályozza a feldolgozó, a másik esetben mint „biológia és kémia”-t ugyanígy következtetlen lesz, mint a keresőprofil leírója, aki ugyanazt a témát egyszer így, máskor úgy jelöli. Végző soron ezek az inkonzekvenciák a visszakeresés határfokát csökkentik, sőt, félre is vihetik a keresést. A probléma egyik megoldása az osztályozási szabályok és a keresőtémák leírási szabályainak kidolgozása. Ugyanannál a példánál maradva: biztosan mást jelent a „biológia vagy kémia”, mint a „biológia és kémia” és nem biztos, hogy az utóbbi mindig biokémiát jelent. De ez az osztályozási szabályok előírásának dolga. Magában a tezaurszban más megoldás kínálkozik az ilyenfajta redundancia csökkentésére.

A feladat úgy végezhető el, hogy a deskriptorokat megkíséreljük kifejezni a többi deskriptorral. Mi történne, ha az illető deskriptor nem lenne és ki kellene fejezni jelentését a többi deskriptorral. Az elemzés eredményét érvényesíteni kell a tezaursz tömörítésében.

Az eljárást azonban nem szabad túlfeszíteni. Ha túlságosan sok a magas szinten koordinált deskriptor, ez hibalehetőségeket rejt magában, növeli az információs zajt.

Technikai kivitelezését tekintve ez az elemzés egy újabb, a szinonimákéhoz hasonló utalórendszer beépítését jelenti a tezaurszba. Példák az utalókkal együtt:

Ferromágneses vékonyréteg

LL Ferromágneses anyag \wedge vékonyréteg.

Az LL utalás jelentése: lásd logikai kifejezésben. A \wedge jel helyett a logikai kapcsolat az „és” szócskával is helyettesíthető. Az utalás inverze:

Ferromágneses anyag

HL Ferromágneses vékonyréteg.

A HL jelentése: helyettesít logikai kifejezésben. A példából világos, hogy a „ferromágneses vékonyréteg” tiltásra került, mert kifejezhető a „ferromágneses anyag” és „vékonyréteg” deskriptorok metszeteként.

Más példa:

Felsőoktatási intézmény

LL egyetem V főiskola,

verbálisan: LL egyetem vagy főiskola. Az utalás inverze:

Egyetem

HL felsőoktatási intézmény.

4.7 A TEZAUROSZ HIERARCHIKUS RÉSZÉNEK SZERKESZTÉSE

Ez a munkaszakasz már a tezaurusz gépi realizációját készíti elő.

Az egyes munkafázisok:

4.7.1 A gráfok lapjairól a genus-species relációk kimásolása.

4.7.2 A – gépkezelés esetén már lelyukasztott – deskriptorokból az alá-fölérendelési reláció szerint több fokozatú hierarchia kiírása a kölcsönös egymásra hivatkozások alapján. Polihierarchia esetén a kiírás az ún. fő felosztás szerint történik.

4.7.3 A 4.7.1 és a 4.7.2 pont szerinti hierarchia. Eredmény: több hierarchia (szakcsoportonként néhány).

4.7.4 A csúcsokban lévő deskriptorok elemzése: ezek között hierarchia kialakítása. Eszköze a 4.1.5-ben kialakított szakcsoportjegyzék, illetve szakrend. Eredmény: a részhierarchiák egyetlen hierarchikus szerkezetté állnak össze.

4.7.5 A kialakult hierarchia kódolása.

4.7.6 Osztályozási szabályok meghatározása.

Megjegyzések:

A műveletek leírása és a munkafázisok ismertetése során sehol nem térünk ki az esetleges számítógépes segítségre.

A szakirodalom e tekintetben háromszintű megoldást ismer.

Az első a teljesen automatikus úton előállított tezauruszok, illetve osztályozási megoldások szintje. Az eljárás természetesen szövegrészek automatikus analízisén nyugszik. A szövegek szavainak tőalakra hozása után szóstatisztikai vizsgálatokat folytatnak, és minden szóra, illetve az egymás melletti szópárookra a gyakoriságból súlyszámot generálnak. A tapasztalati úton megállapított gyakorisági intervallumba tartozó – nem túl ritka és nem túl gyakori – kifejezéseket relevánsnak tekintik. A szavak és szóösszetételek környezetének vizsgálatából bizonyos relációkat valószínűsítene. A számos elemzésnek alávetett jegyzéket végül listázzák, illetve a dokumentumokat minősítik, osztályozzák azokkal a szavakkal és kifejezésekkel, amelyeket a dokumentumból nyernek.

Ez a teljesen automatikus eljárás, amelynek vázlatát itt adtuk, a magyar nyelv esetében egyelőre nem járható, mivel hiányoznak a nyelvi analízisre vonatkozó számítógépi programok és például a ragozott szavak tövesítése is megoldatlan. Egyébként a teljesen automatizált rendszerek 30–40 deskriptorral osztályoznak egy-egy dokumentumot, magas redundanciával és mind az osztályozásnál, mind a visszakeresésnél csak valószínű, nem biztos eredményekre törekcsenek.

A félautomatikus rendszerek lényegében az előbbi utat követik, azzal a különbséggel, hogy az egyes fázisok között szójegyzékeket listáznak, és abból emberi elemzéssel szűrik ki a nyilvánvaló tévedéseket.

A harmadik megoldás — talán ez volna a leginkább járható út számunkra — a rendezési műveleteket, a hivatkozási rendszerek beépítését és a megfelelő ellenőrzéseket számítógéppel végeztetik el.

Ennek megfelelően az ebben a fejezetben leírt munkamenetben az alábbi konkrét esetekben kerülhet sor számítógépi segítségre:

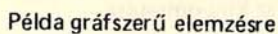
- osztályozás és a fazetták beírása után az induló szóanyag automatikusan sorolható szakcsoportokba,
- az elemzett szóanyag betűrendbe sorolását el lehet végeztetni géppel,
- az elemzett szóanyagba számítógéppel lehet beépíteni a szinonimautalókat,
- az elemzett szóanyag relációinak ellenőrzését lehet automatizálni,
- az elemzett szóanyag relációi alapján emberi ellenőrzés céljára fastruktúrák építhetők — ameddig az egymásra hivatkozások ezt lehetővé teszik — akár a genus-species, akár az egész-rész kapcsolat szerint.

4.8 ELLENŐRZÉS, MODELLKÍSÉRLETEK

Ez a szakasz fejezi be a tezauszépítést — s egyúttal kiindulópontja a tezausz ciklikus javításának, tökéletesítésének.

A főbb munkafázisok:

- 4.8.1 Kísérleti osztályozásra dokumentumanyag kijelölése.
- 4.8.2 Keresési témák kérése vagy gyűjtése.
- 4.8.3 A 4.8.1 és 4.8.2 kísérleti osztályozása. A tapasztalatok, esetek, bizonytalan megoldások lejegyzése.
- 4.8.4 A kísérleti dokumentumanyag alapján a visszakereső rendszer teljes modelljének kialakítása.
- 4.8.5 Kísérleti visszakeresés az adott keresőtémák alapján. Esetek, hibák egyenkénti lejegyzése.
- 4.8.6 Az eredmények értékelése. Javaslatok a tezausz módosítására.
- 4.8.7 A tezausz módosítása 4.8.6 alapján.
- 4.8.8 A végleges tezausz kinyomtatása.



1. példa elemzett deskriptorra

Fogalom:	MÁTRIX
Szakterület:	matematika
Részterület:	mátrixelmélet
ETO:	
Szinonima:	
Kváziszinonima:	táblázat
Fölé:	
Alá:	transzportált mátrix

minormátrix
kvadratikus mátrix
konjugált mátrix
projektomátrix
adjungált mátrix
hipermátrix
kőbös mátrix
végtelen mátrix
reciprokmátrix

Összetevői:

vektor
sorvektor
oszlopvektor
skalármennyiségek

Együtt:

mátrixelmélet
mátrixfüggvény
mátrixpolinom
mátrixmodell

Külső kapcsolat:

térgeometria
lineáris algebra
lineáris programozás
differenciált egyenletek
statisztika

Fazetta:

x

2. példa elemzett deskriptorra

Fogalom:	KVADRATIKUS MÁTRIX
Szakterület:	matematika
Részterület:	mátrixelmélet
ETO:	
Szinonima:	négyzetes mátrix
Kváziszinonima:	determináns

Főlé:

Alá:

Összetevői:

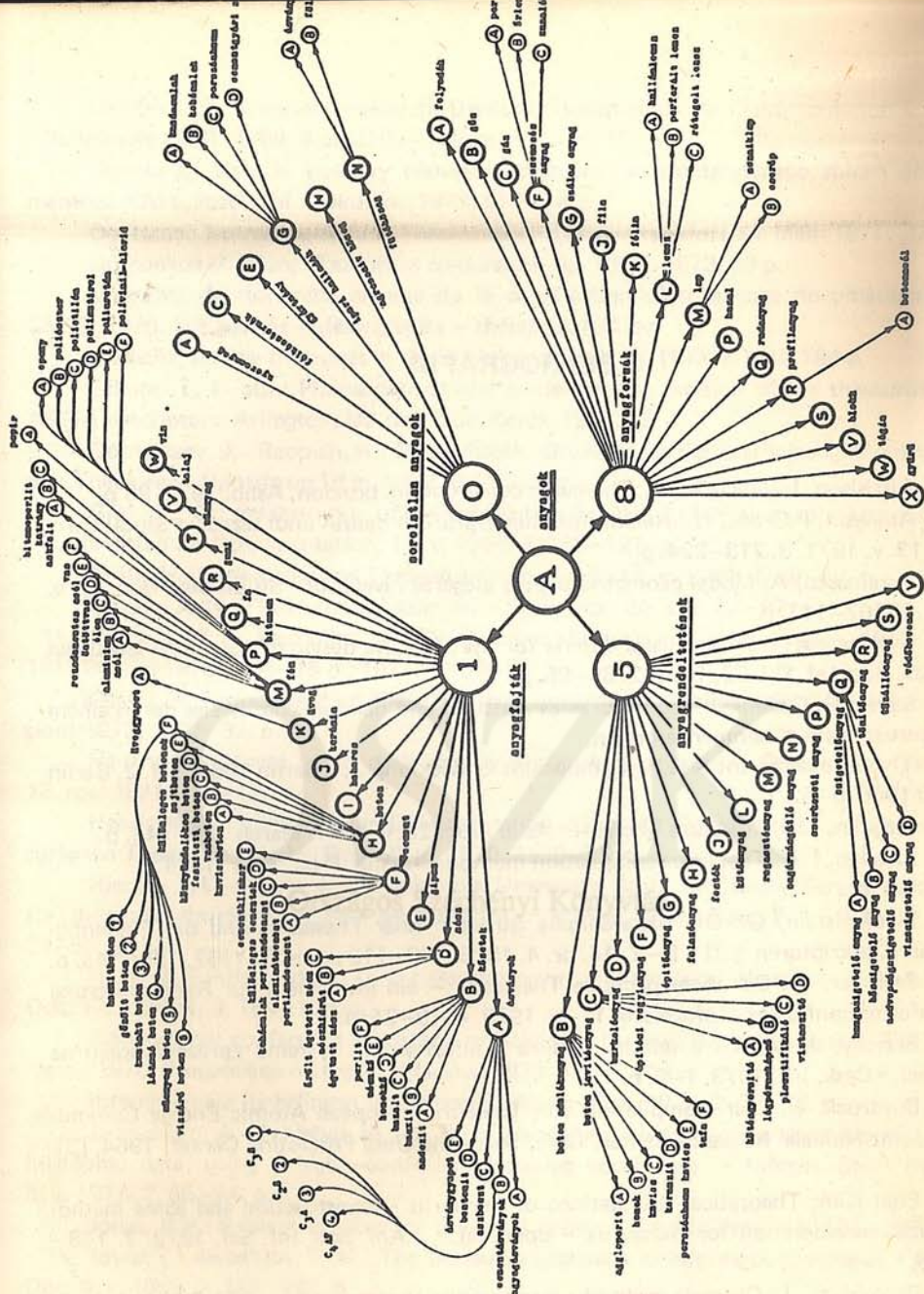
Együtt (azonosság):

Külső kapcsolatok:

mátrix
diagonálmátrix
egységmátrix
permutáló mátrix
szimmetrikus mátrix
hermitikus mátrix
háromszögmátrix
ciklikus mátrix
kontinuáns mátrix
vektor
sorvektor
oszlopvektor
skalármennyiségek
együtthatók
mátrixelmélet
mátrixpolinom
mátrixfüggvény
lineáris egyenletek
lineáris algebra
statisztika
lineáris programozás

OSZK

Országos Széchényi Könyvtár



Példa grafikus táblázatra

Az Építéstudományi Intézet tematikai információs rendszerének tezauruszából

BIBLIOGRÁFIA

- Aitchison, J.—Gilchrist, A.: *Thesaurus construction*. London, Aslib. 1972. 95 p.
- Atanasin, P-Grecu, N.: *Relatii intre descriptori in cadrul unui teaur.* = *Studii Cerc.* Doc. 13. v. 1971. 3. 213—224. p.
- Avramescu, A: *Model geometric pentru alegerea nivelurilor* = *Studii Cerc Doc.* 12. v. 1970. 2. 107—117. p.
- Avramescu, A: *Probabilistic criteria for the objective design of descriptor languages* = *J. Am. Soc. Inf. Sci.* 22. 1971. 2. 85—95. p.
- Bauer, G. *Gedanken zur Theorie des Aufbaus und der Funktionsweise der in einem Thesaurussystem zusammengeforsten.*
- Thesaurusarten In: R.G.W. Symposium Entwicklung . . . Berlin 1966. Tel. 2. Berlin 211 D 1966. 5—52. p.*
- Blagden, 1. F. *Structure Thesauri* = *Aslib Proc.* 23. v. 1971. March 139—143. p.
- Blagden, 1. F. *Thesaurus compilation methods.* = *Aslib. Proc.* 20. v. 1968. 8. 345—359. p.
- Blaukenstein, G.: *Die hierarchische Structur einer Thesaurus und die Eindentigkeit der Deskriptoren* = 211 D—Z; 14. nr. 4. 1976. 110—113 p.; nr. 5. 1967. 138—143. p.
- Brendler, G.: *Der mehrspachtige Thesaurus — ein instrument zur Rationdlisierung des Informationsflusses* = *Informatik* 17. jg. 1970. 4. 19—24. p.
- Brzicky, J. *Ulozen a udrzba tezauru v informacnim systemu zprecovávanem na pocitaci.* = *Cesk. Inf.* 1973. 1—9. p.
- Buntrock. H.: *Zur Kompilation von Thesauri.* European Atomic Energy Community, Joint Nuklear Research Center, Ispra, Scientific Data Processing Center. 1964. [3]. 16 p.
- Ehai Kim: *Theoretical foundations of thesaurus — construction and some methodological consideration for thesaurus — updating.* = *J Am. Soc. Inf. Sci.* 1973. 2. 148—156. p.
- Csernűj, A. J.: *Obscsaja metodika posztroenija thezsursuszov.* = *NTI Szerija* 2. N. 5. 1968. 9—32. p.
- Cujbá, A: *Thesaurul național politematic — National polytematic thesaurus.* = *Probl. Inform. Doc.* 7. 1973. 3. 190—209. p.
- Cujba, A.: *Tezaurul politematic al Ins. Central de Doc. Tehica. Metoda de elaborare.* = *Probl. Inform. Doc.* 6. v. 1972. 5. 273—281. p.

- Dahlberg, J.: Thesaurus research: Deutsche Gellenschaft für Dokumentation E. V. = Bulletin de l'AID. 1966. 4. no. 103–106. p.
- Danuta s.: Metody budowy tesarusa: problem reprezentatywnego zbiorn dokumentow. = Akt. Probl. Inf. Dokumm. 1972. 5. 7–14. p.
- Deutschen Normenausschuss: Thesaurusrichtlinien. Frankfurt am Main 1971. 20. p.
- Domonkos Miklósné: Tezaurusz szerkesztés. Bp. KMK. 1973. 58 p.
- Éléments de recherche en vue de la constitution d'un langage documentaire = CNRS. 1970. 1. Résumés — descripteurs — thésaurus 190 lev.
- Gilchrist, A.: The thesaurus in retrieval. London, Aslib, (1971). VIII, 184 p.
- Gillum, T. L. stb.: Philosophy of and guidelines for revizion of the thesaurus of ASTIA descriptors. Arlington, US Def. Doc. Centr. 1961. 19.
- Gornostaev, J.—Rappich, H.: Methodische Grundfragen der Erarbeitung kompatibler Thesauren = Informatic 19 jg. 1972. 6. 23–29. p.
- Graf, K. C.: Erfahrungen und Erkentsnisse bei der Aufstellung von Thesauren. = Nachrichten für Dokumentation. 15. v. 1964. 3. 118–121. p.
- Gráfe, H.: Systematik und Thesaurus = Informatik, 16. v. 1969. 2. 49–50. p.
- Grosch, A. N.: Thesaurus consruction. = Spec. Libr. 60. v. 2. 87–92. p.
- Guéniot, Y. — Laureilhe, M.: De quelques thesauri. = B. Bibl. France. 1969. 5.: 181–202. p; 1970. 1.: 5–19 p.; 1971. 1.: 33–38. p.
- Guy L. — Reduction d'un thésaurus dans un système documentaire. = TA Informations 1973. 1. 27–37. p.
- Hauzner, I.: Vyvoj pojeti deskriptorovych jaryku a tezauru. = Ceskoslov. Inform. 13. roc. 1971. 2. 9–17. p.
- Henrichs, N.: Philosophische Dokumentation. Liberatur Dokumentation ohne strukturierten Thesaurus = Nachr. Dok. 21 jg. 1970. 1. 20–25. p.
- Hienes, T. C.—Harris, J. L.: Columbia University Scool of Library Service system for thesaurus development and maintenance. = Inform. Stor. Retr. 7. v. 1971/72. 1. 39–50. p.
- Holm, B. E.—Rasmussen, L. E.: Development of a technical thesaurus. = America Doc. 12. v. 1961. 3. 184–190. p.
- International conference on general principles of thesauri building. Warsaw, 23–27. March 1970. Proceedings og the . . . Warsaw, 1970. 186 p. 10 t.
- Internationale Richtlingen für Thesauri = Nachr. Doc. 1970. 2. 72–75. p.
- Jacquesson, A.—Schieber, W. D.: Term association analysis on a large file of bibliographic data, using a highly-controlled indexing vocabulary. = Inform. Stor. Retr. 9. v. 1973. 2. 85–94. p.
- Jones, K.P.: Basic structuces for thesaural systems. = Aslib. Proc. 11. 577–590. p.
- Joyce, T.—Needham, R. M.: The thesaurus approach to information retrieval. = Am. Doc. 9. v. 1958. 3. 192–197. p.
- Keen, E. M.: Thesaurus, Phrase and hierarchy Dictionaries, Part. VII. of. Cornell. Univ. Report. = Inform. Stor. Retr. 1968. jan.
- Lang, F.H.: Automatisierte Hertstellung von Thesauren und Begriffssystemen für Wörterbücher und Fachterminologien. = Nachr. Dok 24. jg. 1973. 6. 231–238. p.

Leontyeva, T. M.—Margaritov, V. B.: Statisticheskij analizi kalicsentvennue karakterisztiki elementov tezaurusznoj szisztemü. = Naucs, Tehn. Inf. Ser. 2. 1970. 10. 19—25. p.

Leski, K.: Normalizacja w budowie tezaurusow = Akt. Probl. Infr. Dok. 1974. 5. 3—8. p.

Lévery F.: Les problèmes posés par le vocabulaire documentaire et l'organisation des dictionnaires et thesaurus. = TA Informations 1972. 1. 1—8. p.

London, G.: A classed thesaurus as an intermediary between textual indexing and searching languages = Revue Internationale de los Documentation 32. v. 1965. 4. 145—149. p.

Löhr, H.—Göldner, R.—Poehlmann, E.—Haase, U.: Anpassung eines Arbeitsthesaurus an übergeordnete Thesauren. = Informatik. 19 jg. 1972. 6. 30—34. p.

Mandersloot, W. G. B.: Thesaurus Control: the selection, grouping and Cross — referencing of terms for inclusion in a coordinate Index World List. = J. ASIS 21 v.; 1970. 49—57. p.

Metodické pokyny na vypracovanie tezáuru. (Návrh.). Prel. Viera Simová. Slovenská technická knižnica. Bratislava, 1967. 27. p.

Methodische Rahmenregelung zum Indexieren mit Hilfe eines Thesaurus. Zentralinstitut für Information und Dokumentation der Deutschen Demokratischen Republik, Berlin. Berlin, 1973, 30 p.

Methodische Richtlinie für die Erarbeitung von Thesauren = Inform.-Dienst. Inform. u. Dok. Beil. 211 D. Mitt. Berlin 1972. jan. 31 p.

Model, Tr.: Thesaurus in der Dokumentation — Vielschichtigkeit des Begriffs und historische Entwicklung — Nachrichten für Dok. 17. v. 1966. 1/2. 5—13. p.

Majzisek, J.: Metodika tvorby tezáuru. 1—2. v. = Českoslov. Inform. 15 roc. 1973. 7/8. 53—71. p., 9. 20—37. p.

Molnár Imre: Az információkezelés fogalomrendszere. (Az információs szaktezaurusz fogalma, rendszere és felépítése) Bp. 1973. 213. p.

Molnár Imre: Az információs szaktezaurusz fogalma, rendszere és felépítése különös tekintettel az építésügyi ágazati szaktezaurusz kialakítására Bp. 1973. 391 p.

Moszkovics, V. A.: Disztributivnosztaticheskij, metod posztroenija tezauruszov: szovremennoe szosztovanieiperspektivi. Cs. 1—2. = Naucs. Tech., Inform. Szer. 2. 1972. 3.: 12—21. p. 4.: 15—24. p.

Moureau, M.: Problemes posés par la structure d' un thesaurus. = B. Bibl. Fr. 1968. 5. 201—210. p.

Mulvikill, J. G.—Brenner, E. H.: Faceted organization of a thesaurus vocabulary. In: American Doc. Inst. Annual Meeting Santa Monica Cal, 3—7. Okt. 1966. Proceedings 75—83. p.

Neguljaev, G. A.: Posztroenie tesaurusa élementarnüh vüszkazüvanij = Naucs. Tech. Inform. Szer. 2. 1973. 5. 8—15. p.

Neufeld, M. L.: Linguistic approaches to the construction and use of thesauri: a reviev. = Drexel Libr. Quart. 8. vol. 1972. 2. 135—146. p.

Neville, H. H.: Feasibility study of a scheme for reconciling thesauri convering a common subject. = J. Dok. 26. v. 1970. 4. 313—336. p.

Neville, H. H.: Thesaurus reconciliation = Aslib Proc. 24. v. 1972. 11. v. 620–626. p.
Nuyl, Th. W. te—Angell, S.: Examination of the validity of the conclusions arrived at in the Aslib Cranfield research project. — Two papers on thesaurus construction. Copenhagen, 1968. 25, 27 p.

Ovcsinyikov, V. G.: Tezauruszi nektorüe metodü ego nostroenija 2. Metodü nostroenija tezaurusza. = Naucsno Tehn. Inf. 1966. 9. 20–26. p.

Petőfi S. János: A tezaurusz-kérdés jelenlegi helyzete különös tekintettel a tudományos, műszaki-gazdasági tájékoztatásra. Bp. 1969. 167. p.

Poletlo, M.: Struktura tezaurusza INIS. = Akt. Probl. Inf. Dok. 1971. 6. 23–29. p.

Preisler, W.: Bedeutung, Aufbau und Prüfung eines Thesaurus = Informatik 17 jg. 1970. 6. 4–18. p.

Preisler, W.: Thesaurusarten und probleme ihrer Strukturierung. = Informatik 20 jg. 1973. 4. 9–16. p.

Preisler, W.: Zur Gestaltung eines Thesaurus. = Informatik. 19 jg. 1972. 1. 22–28. p.

Principes directeurs pour l'établissement et le développement de thesaurus monolingues destinés a la recherche documentaire = Cah. Doc. 26. a. 1972. 1/2. 1–33. p.; 36–50. p.

Rainey, L.: Experience with the new TEST Thesaurus and the new NASA thesaurus. = Spec. Libr. 61 v. 1971. 1. 26–32. p.

Robinson, F.: A computer based rebrieval system using a reactive thesaurus. = Inform. Star. Retr. 6. v. 1970. 2. 171–177. p.

Rolland, M. T.: Thesaurusproblema in Informationsverbundsystemen. München-Pullach, Verl Dokumentation, 1973, 130 p.

Rolling, L. N.: Compilation of thesauri for use in computer system. = Informat. Stor. Retr. 6. v. 1970. 4. 341–350. p.

Rolling, L.: Graphic display decives in thesaurus construction and use = Aslib. Proc. 23. v. 1971. 11. 391–594. p.

Roloff, Heinrich: Sachkatalogisierung auf neuen Wegen. Von der Facettenklassifikation zu Deskriptor und Thesaurus. Berlin. 1968. 56. p.

Rosenbaum, H. D.: Methodische grundsätze für die Erarbeitung mehrsprachiger Thesauren der Mitliedslander des RGW. = Informatik: 21 jg. 1974. 2. 17–21. p.

Rosenbaum, H. D.: Zur künftigen Entwicklung der Informations auf bereitung und Recherchesprachen für Textinformationen. = Informatik. 19. jg. 1972. 4. 38–41. p.

Rostron, R.M.: The construction of a thesaurus = A slib Proc. 20. v. 1968. 3. 171–187. p.

Sand, W.—Sawicky, L.: Prace nad budowa krajowego systemu tezaurusow. = Aktual Probl. Inform. Dok. 17. v. 1972. 5. 3–6. p.

Schön, J.: The thesaurus of Engineering and Scientific Terms = Nachr, Dok. 21 jg. 1970. 110–113. p.

Sikorski, L.: Die Struktur eines einheitlichen Thesaurus für verschiedene Industriezweige. = Informatik 18 jg. 1971. 6. 23–26. p.

Simescu, A.—Divna, A.: Asupra analizei unei multimi de documente in vederea obtinerii tezauruli = Studii Cerc Doc. 14 a. 1972. 2. 147–152. p.

Soergel, Dagobert: Klassifikationssysteme und Thesauri. Frankfurt/Main, Deutschen Gesellschaft für Dokumentation, 1969. 224 p.

Sparch, J. K.: Automatic thesaurus construction and the relation of a thesaurus to indexing terms. = ASLIB proc. 22. 1970. may. 226–228. p.

Spranger, U.: Zusammenführung verschiedensprachiger und unterschiedlich strukturierter thesauri unter einem gemeinsamen code = Informatic. 20 jg. 1973. 6. 15–17. p.

Stejder, J. A.: Tezaurusii v. informatik, teoreticeszkoj szemantik. = Naucs. Tecn. Inform. Szer. 2. 1971. 3. 21–24. p.

Starska, D.: Metody Budowy tezaurusa-problem reprezentatywnego zbioru dokumentow. = Aktual. Probl. Informat. Dok. 17. rok. 1972. 5. 7–14. p.

Surace, J.: The display of a thesaurus. Santa Monica, Rand Corp. 1970. 37 p.

Szepesiné Benda Mária—Barné Pollak Zsuzsanna: Tezaurusz és gyakorlati alkalmazás. Bp. 1971. 36. p.

Thomas, P. A.: Some problems in compliting a thesaurus of transportation and planning terms. = Aslib. Proc. 23 v. 1971. 11. 595–606. p.

Toma, E.: The structure of the EURATOM — Thesaurus. = J. Doc. 27. v. 1971. 4. 267–272. p.

Toma, E.: Studial statistic al structurii tezaurelar: diagnostic comparativ. = Stud. Cerc. Doc. 1972. 3. 217–229. p.

Toman, J.: Postup při sestavování tezauru. [Vyd] Státní technická knihovna, Ustředí vedeckých technických a ekonomických infomaci, Praha. Praha, 1968. 32 p.

Turski, W. M.: On a model of information rethrhval system based on thezaurus. = Inform. Stor. Retr. 7. v. 1971. 2. 89–94. p.

Vickery, B. C.: Thesaurus — a new word in dokumentation. = J. of. Doc. 16. v. 1960. dec. 181–89. p.

Varga Dénes: Információs tezauruszok készítésének módszertana. Bp. 1969. 156 p.

Wersing, G.: Eines neue Definiton von „Thesaurus“ = Nachr. Dok. 20. 1969. 2. 53–62. p.

Wollf-Tervonie, M.—Rimberl, D.: Computer aided automatic generation of a Structured documentary language. = J. of Doc. 27. 1971. jun. 111–124. p.



Ára: 10,— Ft

OSZK

Országos Széchényi Könyvtár