

A TEZAUROSZOK AUTOMATIZÁLÁSA. A GÉPI TEZAUROSZKÉSZÍTÉS ÉS HASZNÁLAT NÉHÁNY KÉRDÉSE*

Ungváry Rudolf

OSZK Fejlesztési Csoport

1. Bevezető

1.1 A gépi tezausz környezet

Az információkereső tezauszok elterjedése szorosan összefügg a szöveges adatok és bibliográfiai tételek gépi tárolásával és azzal az igénnyel, hogy ezeket a tételeket a tartalmukra vonatkozó kérdések alapján lehessen visszakeresni.

Az automatizált dokumentációs rendszerekben a hagyományos „bibliográfiai tétel” helyett helyénvalóbb az általánosabb értelmű „információtétel” [1] kifejezést használni, mivel a keresőszavakkal – legyenek ezek szabadon választhatók (ún. „szabad szöveges” rendszerekben) vagy kötöttek (szótárba, pl. tezauszba, tárgyszójegyzékbe, osztályozási táblázatokba stb. foglalt információkereső nyelvek) – korántsem csak a szorosan vett bibliográfiai tétel az egyetlen fajta visszakereshető egység. Ilyen önállóan kérdezhető egységek lehetnek a bibliográfiai tételek egyes részei (mint például a szerző neve, a kiadó, a dokumentum fajtája, a cím stb.), az annotáció vagy a referátum és ezek részei, sőt egyes dokumentációs rendszerekben az eredeti dokumentumszövegek és ennek részei is. Mindig az adott keresési cél határozza meg, hogy milyen fajtájú a keresett tétel, ezért néha külön névvel is megkülönböztetik ezeket a tételeket: a GOLEM-ben használt terminológia szerint például *célinformációkról* beszélnek.

A szöveges információtételek szervezett állományát szöveges vagy bibliográfiai *adatbázisnak* nevezzük. Általános értelemben minden szervezett és kezelhető adatállomány adatbázisnak tekinthető (ilyen értelemben „adatbázis”, „manuális adatbank” a könyvtári katalógus vagy a telefonkönyv adatainak összessége is), szűkebb értelemben azonban csak a géppel olvasható formában rögzített és számos – itt nem részletezett – követelménynek [VI] eleget tevő adatgyűjteményt nevezik adatbázisnak. Megkülönböztethetők az elsődleges dokumentumokat tartalmazó adatbázisok a másodlagos adatokat – például a bibliográfiai adatokat, referátumokat stb. – tartalmazó bibliográfiai adatbázisoktól, s más szempontból típusba sorolhatók aszerint, hogy szöveges és/vagy faktografikus jellegű adatokat tartalmaznak. Egyes szerzők szűkkeblűen [VI, 82–83] csak a „logikai” *szerkezettel* leírható faktografikus adatbázisokat tekintik adatbázisoknak (és nem használják a „faktografikus” jelzőt), a *tartalom* alapján – például deskriptorokkal, kulcsszavakkal, osztályozási jelzetekkel stb. – kezelt gyűjteményeket „szöveges adatállománynak” nevezik.

Azt a programrendszert, melynek segítségével az adatbázisokat kezelik és a szolgáltatást biztosítják, *adatbáziskezelő rendszernek* hívják. Bibliográfiai adatbázisok – vagy „szöveges adatállományok” – esetében különleges eljárások szükségesek a hatékony tartalmi információkereséshez, a „hagyományos” adatbáziskezelő rendszerek túlságosan erős kötöttségei általában nem teljesíthetők, illetőleg feleslegesek. Mivel itt a rendszer fő funkciója rendszerint nem az adatok karbantartása, hanem kész adatokból válaszinformációk kiválogatása, hangsúlyozottan *információkereső rendszerrel* szoktak

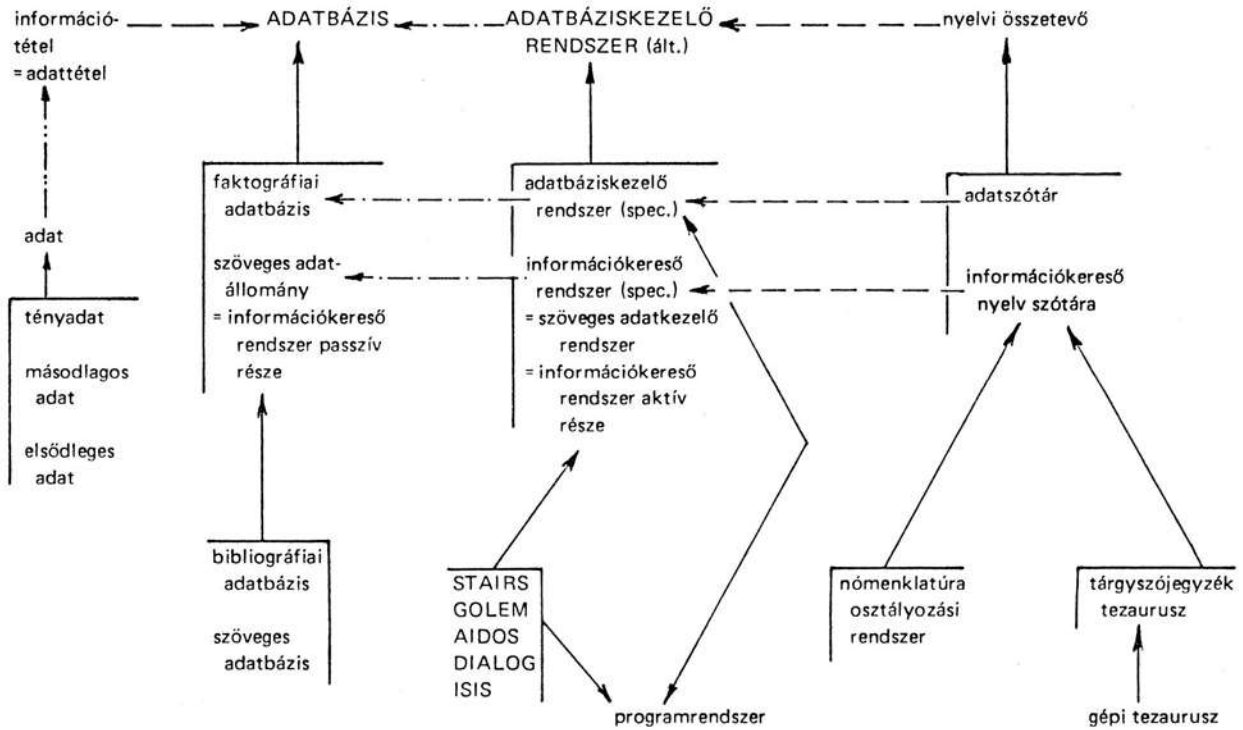
* E tanulmány előzményét a *Magyarországon készülő tezauszok a nemzetközi szabványosítás tükrében* című felmérés [XXIII] képviseli, melyet kiegészít a gépesítésre vonatkozó kérdésekkel.

beszélni [2]. Az adatok szerkezetének formális leírása helyébe itt kevésbé szigorúan formalizált eszközök általánosabb keresést lehetővé tevő felhasználása lép.

A tartalom szerinti keresés érdekében a kezelőrendszer nyelvi összetevőt – az információkereső nyelvet – kezelő résszel is rendelkezik. Az információkereső nyelv szavai – szótár (pl. teaurusz) esetén a lexikai egységek – biztosíthatják a tartalmi szempontok meghatározott jellegű és finomságú érvényesítését az információkeresés-

ben, rajtuk keresztül férhetünk hozzá az egyszerű vagy az összetett információtöttelekhez. Úgy is mondhatjuk, hogy a lexikai egységek behatárolják a lehetséges célinformációk körét, de egyúttal strukturálják is a visszakereshető információkat a tartalom szerint.

A programrendszer által használt – kezelt – teauruszt nevezik *gépi teaurusznak*. (A tárgyalat fogalmak közötti összefüggések az 1. ábrán láthatók.)



1. ábra A gépi teaurusz fogalmi környezetének címkézett, irányított gráfja

(—→ faj–nem reláció; - - - → rész–egész reláció; · · · · · → eszköz–rendeltetés reláció; - · - · - · → eredet–eredmény reláció)

A nyílakkal jelölt összefüggések olvasására néhány példa:

A TÁRGYSZÓJEGYZÉK és a TEAURUSZ nem más, mint az INFORMÁCIÓKERESŐ NYELV SZÓTÁRA, mely része az INFORMÁCIÓKERESŐ RENDSZERnek, s mint ilyen nem más, mint NYELVI ÖSSZETEVŐ. Az INFORMÁCIÓKERESŐ RENDSZER a SZÖVEGES ADATÁLLOMÁNY kezelésének eszköze.

1.2 A gépi teauruszkészítés

A gépi teauruszok adatállományát – a lexikai egységeket, a közöttük meghatározott relációkat és a lexikai egységeket jellemző egyéb adatokat (mint a szakcsoportjuk, a fogalmi kategóriájuk stb.) – számítógéppel olvasható adathordozón rögzítik. Érthető törekvés az, hogy a teauruszok készítését is gépi segítséggel végezzük. Az információkereső rendszerek egy része ugyan alkalmas arra, hogy karbantartsa és a keresésbe bevonja a meglévő gépi teauruszt, de magának a készítésnek és fejlesztésnek különleges adatfeldolgozási

feladatait nem lehet velük megoldani, vagy legfeljebb nagyon leegyszerűsített formában és drágán. A teauruszkészítés gyűjtő, válogató, átalakító és szerkesztő műveletei ugyanis azzal járnak, hogy minduntalan sok teauruszadatot – megnevezést, szakcsoportos besorolást, relációjelet stb. – kell törölni, bevinni, módosítani, átrendezni stb. Az információkereső rendszerek teauruszkezelő moduljaival a már kész gépi teaurusz egyes adatait lehet ugyan módosítani, de sok adat esetén a módosítás már nem végezhető el hatékonyan, elfogadható időn belül. Ezeket a teauruszkezelő programmodulokat ugyanis csak a kész gépi teaurusz kis mértékű

változtatásokkal járó aktualizálására méretezték; ráadásul számos – a készítéskor feltétlenül szükséges – teauruszadatot többnyire nem is tudnak nyilvántartani (pl. a lexikai egységek szakcsoportos besorolását). Úgy is mondhatjuk, hogy magával az információkereső programrendszerrel gazdaságtalan teauruszt készíteni: olyan, mintha traktort használnának személyszállításra. Szükség volt tehát önálló, csak a készítésre specializált programrendszerekre is.

Felhasználhatók ezek a programrendszerek a kézzel kezelt információtárak és katalógusok teauruszainak készítéséhez is, mivel ezeknek a készítése is nagy adatfeldolgozási és rendezési munkát igényel. A teauruszok a finom, részletes feltárás eszközei, és ezért – még a rendkívül szűk szakterületek kisebb teauruszai is – általában több ezer lexikai egységet tartalmaznak [3]. Ha lexikai egységenként átlagosan csak 5 relációt és annak 5 inverzét vesszük számításba, és teauruszonként átlagosan 5000 lexikai egységgel számolunk, a beépítendő adatelemek száma eléri az ötvenezret. Gépi segítség nélkül nagy nehézséget okoz annak állandó ellenőrzése, hogy minden kapcsolatnak bejegyezték-e az inverzét is, és nem tartalmaz-e ellentmondást a teauruszcikk (nem kapcsolódik-e például nemdeszkriptorhoz valamilyen csak deszkriptorok között lehetséges – pl. egész–rész, rokonsági stb. – relációban deszkriptor stb.).

A 60-as évek közepétől kezdve vannak adatok arról, hogy kézi, illetve gépi teauruszok készítéséhez speciális programrendszereket használtak, ezek azonban többnyire csak a manuális műveletek elvégzésében nyújtottak segítséget, amilyen a betűrendezés, a kapcsolatnyilvántartás stb. Készítésük ad hoc jellegéből fakadt, hogy mindig csak egyetlen teauruszhoz használták őket fel, a felhalmozott tapasztalatok alapján a programrendszer tökéletesítésére már nem került sor. Ezért a legtöbb rendszert ugyanazok a gyermekbetegségek jellemzik: nehézkesen használhatók, nem nyújtanak kifogástalan, szabványos teaurusz-kinyomtatásokat, számos hiba automatikus ellenőrzése és kiszűrése nem lehetséges stb.

Az újdonsült teauruszkészítő tehát mindig ugyanabba a problémába ütközött: az itt-ott valamikor használt programrendszerek beszerzése körülményes volt, ha sikerült beszerezni, kiderült, hogy a dokumentációjuk hiányos, legfeljebb az ismeri ki magát bennük, aki készítette őket, más igények kielégítésére készültek, nem alkalmazkodtak a kialakult nemzetközi és nemzeti szabványokhoz, többnyire speciálisabbak voltak annál, mint amit átalakítás nélkül fel lehetett volna használni. Ezért aztán vagy a kézi teauruszkészítést választotta, vagy maga is készítetett egy egyszerűbb vagy igényesebb adatkezelő rendszert a munkájához.

Ez a feladat egyrészt könnyebbnek látszott, mint valamilyen meglévő, de nehezen beszerezhető, rosszul dokumentált, körülményesen kezelhető szoftvertermék adaptálása, másrészt valóban sokkal egyszerűbb, mint

magának az információkereső programrendszernek a kidolgozása [4], melyből a 70-es évek eleje óta egyre több került kereskedelmi forgalomba. Ezért a programozók részéről sokszor nagy a kísértés, hogy inkább belevágjanak új, teauruszkészítést támogató programrendszer kidolgozásába, amely aztán semmivel sem jobb azoknál, amelyek helyett készültek.

Idővel mégis ismertté vált néhány programcsomag [VIII, XVI, XIX], melyet több teaurusz készítésére is felhasználtak és kereskedelmi forgalomban is beszerezhetőek. Ezek a 70-es évek végén készültek, és jellemző rájuk, hogy maguk a felhasználók az előállítói, szemben az információkereső rendszerekkel, melyeket a nagy számítógépgyártó cégek – mint az IBM, SIEMENS, ROBOTRON stb. – dolgoztak ki és dobtak piacra.

2. A kereskedelmi forgalomban kapható, Magyarországon ismertebb programcsomagok

2.1 Gépi teauruszt is használó információkereső rendszerek

Az ismertebb programcsomagok közül a GOLEM [V], a STAIRS [XVIII] és az AIDOS [5, I] tartozik az élvonalba. A teljesítmény – és nem utolsósorban a dokumentáltság – szempontjából a GOLEM [6] és a STAIRS felülmúlják az AIDOS-t. A továbbiakban csak az információkereső és szótárkezelő képességek szempontjából értékelem őket.

Az *információkereső funkciók* (pl. az automatikus hierarchiaszint emeléssel és süllyesztéssel végzett keresés, az automatikus nemdeszkriptor–deszkriptor utalás, a szótöredékekkel – csonkolással, „maszkolással” –, illetve a deszkriptorok rangsorolásával végzett keresés) szempontjából is fejlettebb a GOLEM és a STAIRS (a GOLEM-mel egyedül az automatikus hierarchiaszint változtatás nem lehetséges).

A *teauruszkezelő funkciók* (pl. szóhosszúság, szakcsoportos rendszer, relációk száma, teauruszadatok felvitelének módja stb.) szempontjából szintén a GOLEM és a STAIRS a rugalmasabb; e téren a GOLEM hátránya, hogy körülményesebb valamivel a teauruszadatok felvitele, és nincs lehetőség a szakcsoportos besorolásra.

A STAIRS hagyományos változatában nem használnak teauruszt, hanem szabadon választott szövegszavak, illetve töveik alapján végezhető a keresés. Ilyen „szabad szöveges” információkeresés elvileg a többi ismertített rendszerben is végezhető, de a STAIRS-nél kisebb határfokkal, és drágán. A STAIRS hagyományos változatának kipróbálására Magyarországon a 70-es években került sor. Legújabb, virtuális tárral működő változata – a STAIRS VS/TLS [XVIII] – többnyelvű teauruszok kezelésére is alkalmas rendszer. Magyarországi adaptálására eddig nem került sor.

Az ISIS [IX] lényegesen „kevesebbet tud”. A relációt legfeljebb nyilvántartja, de kezelni még a nemdeszkriptor–deszkriptor kapcsolatokat sem képes, és semmiféle szintaktikai relációt nem ismer [7].

A DIALOG információkereső rendszert [8] az NTMIK-ban fejlesztették ki; bár „ősének” a STAIRS hagyományos változata tekinthető, annak színvonalától elmarad. Tezauruszkezelő képessége az ISIS-nél valamivel jobb – automatikus nemdeszkriptor–deszkriptor utalás lehetséges –, de az AIDOS-énál jóval gyengébb (még a deszkriptorok közötti relációkat sem ismeri). Különlegessége, hogy – a STAIRS-hez hasonlóan – nyilvántartja és kezeli a túl általános jelentésű, ún. negatív deszkriptorokat (a rendszerben stop-paraméter a nevük). Magyarországi felhasználására várhatóan a közeljövőben kerül sor.

A hazai termékek közül a HIR [VII], a MIR [XIII] és a SZIV [XXI] – illetve ez utóbbi továbbfejlesztett változata, a PRESZIV – alkalmasak arra, hogy gépi tezauruszt kezeljenek. Ami a feldolgozható információ-tételek mennyiségét illeti, teljesítőképességük messze az előbb ismertett programrendszerek alatt marad. A SZIV teljesítőképessége a legkisebb, 10–20 ezer információtétel fölött már gazdaságtalan a használata. A HIR [VII] általános célú rendszer, eredetileg az országos jogszabálynyilvántartás céljaira készült, tezauruszkezelő és információkereső képességek tekintetében igényes rendszer. Különlegessége, hogy biztosítja néhány, a faktografikus adatbáziskezelő rendszerekre jellemző funkció ellátását is; a tárolt információtételeknek ugyanis lehet összetett, belső – „hierarchikus” – szerkezete, mely alapján ugyancsak végezhető keresés (ez egyébként a GOLEM-et is jellemzi). Gyakorlati teljesítőképességeit még nem próbálták ki.

A MIR [XIII] mikroszámítógépen tárolt szöveges és bibliográfiai adatbázisok kezelésére készült információkereső rendszer. Az ilyen programcsomagok esetén nem a gyorsaságnak van jelentősége – hiszen rendkívül kicsi, „magánjellegű” adatállományok kezeléséről van szó –, hanem a laikus felhasználót messzemenően támogató nyelvi kényelemnek. A MIR öt szemantikai relációjával, automatikus szinonimautaló és hierarchiaszintemelő képességeivel a tezauruszkezelés szempontjából a HIR-hez hasonlóan igényes rendszer.

A többi hazai információkereső programcsomag – mint a BIBDOSZ [XV], a BINÁR [X] és a KSH-ban adaptált TEXT-PACK nem kezel szótárral rendelkező nyelvi összetevőt, így tezauruszt sem. A BINÁR arra sem képes, hogy a felhasznált kulcsszavakat rendezetten tárolja.

Különleges rendeltetésű rendszer a LOGEL, mely eredetileg tudományszervezési és döntéshozzájárulási célokra készült [XI]. Kezelni lehet vele egyszerű szerkezetű – monohierarchikus, egyetlen relációt tartalmazó – tezauruszt is, s ezáltal bizonyos információkereső feladatok is megoldásíthatók vele.

2.2 Az információkereső rendszerek alkalmazásának hazai problémái az információkereső nyelv szempontjából

A BIBDOSZ története [9], akárcsak a számítástechnika hazai történetéről a közelmúltban megjelent tanulmány [II] – alátámasztja azt az érvelést, hogy megfelelő színvonalú szakmai háttér (s tegyük hozzá: anyagi ráfordítás) nélkül, csak néhány kiemelkedő tehetségű emberre támaszkodva nem érdemes hazai erőből nagyobb teljesítményű információkereső programrendszer készítésébe belefogni. Az ilyen rendszerekkel szemben támasztott összetett követelmények specializált tervezőintézetek – ún. szoftverházak – közreműködését igénylik, melyek már megfizették a tanulópénzt. Még a külföldi programcsomagok igényes, a bennük rejlő lehetőségeket teljesen kihasználó és ezáltal a felhasználó kényelmét maradéktalanul biztosító adaptálása is rendkívüli nehézségekkel jár, ahogy ezt a GOLEM és az ISIS tapasztalatai bizonyítják. „... Ha hazai eredményeket nem is sikerült elérni, a külföldi megértése és követése is elég nagy feladat, a jelenleginél nagyobb erőfeszítést igényel” [II].

Előfordul, hogy az alkalmazó vállalatokban van a hiba: elsősorban a pusztán működtetést tartják fontosnak, mintha valaki csak azzal törődne, hogy az autónak járjon a motorja és guruljanak a kerekei, az már nem számít, milyen az ülés, lehet egy deszkalap is. A nyelvi, tartalmi szempontok iránt gyakran érzéketlen rendszertervező figyelmen kívül hagyja a felhasználó szempontjait és kényelmét, lényegében tehát a mélyebb értelemben vett kereskedelmi szempontokat is.

A kereskedelmi forgalomban kapható külföldi programrendszerekkel kapcsolatban gyakran megtörténik, hogy beszerzésükkor *nem mérlegelik kellően információkereső nyelvi összetevőjüket*, így tezauruszkezelő képességüket sem. Így aztán csak a gyakorlati üzembehelyezés-kor derül ki – amikor az adatbázis feltöltéséhez az információkereső nyelvhez mélyebben értő szakemberek kezdenek hozzá –, hogy mennyire kezdetlegesek a tezauruszkezelő funkciók. Holott éppen ez a funkció, az információkereső-nyelvi összetevő a lényege annak a feladatnak, mely az ilyen rendszerek nevében is megfogalmazódik.

Az információkereső-nyelvi összetevő kitüntetett szerepet játszik a laikus felhasználó és az adatbázis közötti kapcsolatban, s kényelmére még nagyobb figyelmet kell szentelni, mint a faktografikus adatbázis adattípusainak – adatfajtáinak – jegyzékét tartalmazó és szemantikai szempontból sokkal egyszerűbb adatszótárra, ahol az adattípusok közötti relációkat nem a szótár – tehát a „nyelv” –, hanem maga a faktografikus adatbázis tartalmazza [VI, 275–302]. A faktografikus adatbázisokat elsősorban a rekordok közötti explicit kapcsolat jellemzi (az információtételeknek tehát összetett belső szerkezete van), szemben a bibliográfiai adatbázisokkal, melyekben

a rekordok közötti kapcsolatok nyelvi kapcsolatokon keresztül realizálódnak [VI, 82–83]. A természetes nyelvi összetevő fejlettsége tehát a szöveges információkereső rendszerek beszerzése és működtetése szempontjából nemcsak számítástechnikai és szervezéstechnikai

részletkérdés, hanem a dolog lényegét érintő probléma [10].

Az ismertebb információkereső rendszerek tulajdonosságait és főbb adatait az 1. táblázatban foglaltuk össze.

1. táblázat

Az ismertebb kereskedelmi forgalomban kapható információkereső rendszerek és tezauszépitést támogató programcsomagok

Név	Készítő	Felhasználó	Forrányelv	Gép	Megnevezések hossza	Szakcsoportos rendezés	Relációk száma	Automatikus nemdeszkriptor-deszkriptor utalás	deszkriptor-deszkriptor utalás	Szintaktikai reláció
STAIRS	IBM	KFKI	PL/1	IBM 370	korlátlan	nincs	korlátlan	van	van	van
GOLEM	SIEMENS	SZKI ÉTK	Assembler	Siemens	255	nincs	korlátlan	van	nincs	van
AIDOS	ROBOTRON	Ipari Informatikai Központ	PL/1	R 20	50 100	van	14	van	van	van
DIALOG	NTMIK	OMIKK	Assembler	ESZ-II.	63	nincs	1	van	nincs	van
ISIS	Unesco/ILO	SZÁMALK	PL/1	IBM 370	256	nincs	nincs	nincs	nincs	nincs
HIR	SZÁMALK	–	Fortran PL/1	ESZ-II. IBM 370	256	nincs	16	van	van	nincs
MIR	Informatika Gmk/Csáky B.	–	Assembler	VT20A	40	nincs	5	van	van	nincs
SZIV	SZKI	Zrinyi Katonai Akadémia	Assembler	R 20	220	nincs	60	nincs	nincs	nincs
BINÁR	KFKI	OMIKK KFKI	Fortran Assembler	R 40	korlátlan	nincs	nincs	nincs	nincs	nincs
LOGEL	INFELOR/ÉTI	ÉTI Ipari Minisztérium	Fortran	R 20 R 55	20 32	nincs	1	nincs	nincs	nincs
TSPR	Schlanger L.	Ipari Informatikai Központ, SZÁMALK, MÉM NAK stb.	PL/1	R 20 IBM 360	40	van	18	12 tezausz készült vele		
ASTUTE	EURONET	–	Cobol	IBM 360/370 Siemens	35	van	10			
SZVET	NTMIK	–	Assembler	ESZ-I.-II.	63	van	5			

2.3 A tezauszépitést támogató ismertebb programcsomagok

A kevés kereskedelmi forgalomban hozzáférhető külföldi programcsomag közül az ASTUTE az EGK Bizottságának számítóközpontjában készült, eredetileg az EURONET-ben használt többnyelvű tezauszok készítésének támogatására [VIII]. A SZVET az NTMIR-ben készülő tezauszok kidolgozását és karbantartását hivatott segíteni [XIX]. Beszerezhető hazai fejlesztésű, kipróbált programcsomag is, a TSPR, melyet 1975 óta folyamatosan használnak tezauszok készítésére, és részletesen dokumentálták [XVI, II].

A legfejlettebb közülük a TSPR. Tökéletesítéséhez nagyban hozzájárult, hogy eddig már több mint tíz tezausz készült el vele (illetve van készülőfélben). A TSPR-rel a lexikai egységek szakcsoportok, elvont fogalmi kategóriák és a gyűjtés forrása szerint csoportosíthatók egyszerre; az ASTUTE – akárcsak a SZVET –

egyszerre csak egyetlen szempont szerint rendezi a szavakat. Mindegyik rendszerben ellenőrizhetően szerkesztett azonosítási szám képviseli a lexikai egységeket, amely nagyon leegyszerűsíti a relációk bevitelét és módosítását, mert a megnevezések ismételt, hosszas leírása szükségtelen. Mind az egy-, mind pedig a többnyelvű tezauszok készítésére alkalmasak. Az ASTUTE egyszerre öt, a TSPR korlátlan számú nyelv tezauszát képes kezelni.

Az ASTUTE legnagyobb hátránya, hogy legfeljebb 4 400 lexikai egységet képes kezelni. A TSPR-nél ez az érték elég nagy (142 ezer lexikai egység). Előnye a TSPR-nek, hogy segítségével a lexikai egységek permutált mutatója egyszerű módon kinyomtatható.

A relációk kezelése szempontjából az ASTUTE annyiban előnyös, hogy vele három különböző nyelv egymással párhuzamos, háromhasábos kinyomtatása lehetséges. A TSPR segítségével definiálható a legtöbb reláció, és a hibellenőrzési lehetőségek gyakorlatilag minden előfor-

duló ellentmondásra kiterjednek. Nemcsak az ellenőrizhető automatikusan, hogy egy nemdeszkriptorhoz deszkriptorok közötti relációban más deszkriptort kapcsolnak, hanem az is, ha ugyanazt a deszkriptorpárt két különböző relációjellel kapcsolták össze, ha az irányított relációláncban végtelen ismétlődés – ciklus – fordul elő stb. A nemdeszkriptorok és deszkriptorok, valamint a különböző nyelvű kifejezések közötti ÉS, VAGY és DE NEM relációkat csak a TSPR kezeli.

Különösen jelentős a tezauszkészítés szempontjából az összegyűjtött lexikai egységek besorolási rendszerének – például a szakcsoportosításnak – a megváltoztathatósága, melyre egyrészt a szerkesztési munka tapasztalatai nyomán van szükség, másrészt olyankor, amikor a tezausz adatállományát különböző felhasználói szempontok szerint rendezve kell kinyomtatni. Ez is csak a TSPR segítségével valósítható meg, ugyanolyan egyszerű módon, mint az, hogy egy-egy tezausból meghatározott részeket válogassanak ki, hogy belőlük speciálisabb résztezauszokat állítsanak elő. Ennek különösen a nagy volumenű tezauszok és tezauszrendszerek esetén van jelentősége.

3. A gépesítés hatása a tezauszok szerkezetére

A gépesítés két oldalról befolyásolja a szerkezetet: egyrészt a gépi tezauszt befogadó információkereső rendszerek, másrészt a számítógéppel támogatott tezauszkészítés részéről. A legjelentősebb hatások a relációk automatikus kezelhetőségéből és a szintaktikai relációk használatának lehetőségéből adódnak. Automatikus indexelés esetén ezekhez járulnak még a tőszótár használatának következményei.

3.1 A relációk automatikus kezelésének hatása

A legtöbb fejlettebb programrendszer biztosítja az automatikus nemdeszkriptor–deszkriptor utalást. Az indexelő közvetlenül a nemdeszkriptorral is osztályozhat, a programrendszer automatikusan összekapcsolja a felhasznált kifejezést a szinonima- illetve kváziszinonimacsoportot képviselő kifejezéssel. Bár ez a tény nem annyira a tezausz szerkezetét, mint inkább használatát befolyásolja, előfordul, hogy emiatt kisebb jelentőséget tulajdonítanak annak, hogy melyik kifejezés legyen a deszkriptor és melyik a nemdeszkriptor. Ez ellentétben van a tezausz-szabványok ama követelményével, hogy az inkább elfogadott kifejezést kell deszkriptornak kijelölni.

A legkorszerűbb rendszerekben – mint a STAIRS VS/TLS és a GOLEM – a nemdeszkriptor–deszkriptor utalás operatív módon ki- és beiktatható: a kérdésben előírhatjuk, hogy a rendszer vegye figyelembe a kérdésben szereplő kifejezések nemdeszkriptor–deszkriptor

kapcsolatait vagy hagyja figyelmen kívül őket, és csak azokat a dokumentumokat válogassa ki, melyeket a kérdésben szereplő kifejezésekkel indexeltek. A GOLEM-ben például az ebtényésztés magyarországi irodalmára vonatkozó, természetes nyelven megfogalmazott kérdés a következőképpen fogalmazható meg az első, illetve a második esetben:

ADES SYNOUT EB•TENYÉSZTÉS•MAGYARORSZÁG
ADES SYNIN EB•TENYÉSZTÉS•MAGYARORSZÁG

ahol a SYNOUT, illetve a SYNIN az utalás ki-, illetve beiktatásának a parancsa, s ettől függően vagy nem kapjuk meg a KUTYA lexikai egységgel indexelt dokumentumokat, vagy megkapjuk őket.

Ezekben az információkereső rendszerekben tehát nemcsak szabad nemdeszkriptorral indexelni, hanem kell is; azaz mindig azt a lexikai egységet kell használni – ha van ilyen a tezauszban –, amelynek megfelelő alakú szó a dokumentumban is szerepel, tekintet nélkül arra, hogy ez nemdeszkriptor vagy deszkriptor. Célszerű a valódi szinonímián és a kváziszinonímián alapuló nemdeszkriptor–deszkriptor kapcsolatokat eltérő módon jelölni, mivel nyilvánvalóan csak az utóbbi esetben van értelme az automatikus utalás kiiktatásának. Kváziszinonímia esetében ilyenkor a keresőkérdés pontosítható. A gépi tezauszokban tehát előfordulhat kétféle nemdeszkriptor–deszkriptor kapcsolat is. A szabványosítás ezt a lehetőséget nem tartja nyilván.

Az előbb említett programrendszerekben a deszkriptorok közötti relációk is kezelhetők automatikusan. Ez azt jelenti, hogy a keresőkérdésbe bevonhatók mindazok a deszkriptorok, melyek a tezauszban kapcsolatban állnak a keresőkérdés deszkriptorláncának valamelyik deszkriptorával. Így módon a kérdés különböző értelemben kiterjeszthető. A legjelentősebb az automatiikus hierarchiaszint emelése, illetve süllyesztése, azaz a fölérendelt, illetve az alárendelt deszkriptorok bevonása a keresőképbe. Ezáltal a találatok száma növelhető, illetve csökkenthető anélkül, hogy a kérdés értelme megváltoznék (csak általánosabb, illetve specifikusabb lesz). Az AIDOS-ban például az agártenyésztés magyarországi irodalmára vonatkozó keresőkérdés az AGÁR-ra vonatkozóan a következőképpen terjeszthető ki, ha attól tartunk, hogy túl kevés – például 10 tétel alatti – a találatok száma:

AGÁR < 10 (001)

ahol a <10 utal arra az esetre, amikor a találatok száma kevesebb tíznél, az O betű a fölérendelt (Oberbegriff) bevonását, a 01 pedig a közvetlen – első szinten kapcsolódó – deszkriptort jelenti.

E lehetőség következtében a gépi tezauszokban nagy jelentősége van a logikailag pontosan értelmezett

nem-faj reláció (az ún. „szigorú hierarchia”) betartásának. Ilyen reláció áll fenn pl. az AGÁR-KUTYA, ÚTÉPÍTÉS-ÉPÍTÉS, ASZTAL-BÚTOR, LOBBANÁSPONT-PARAMÉTER fogalmak között, s legalábbis megfontolandó, hogy milyen következményei lesznek a hétköznapi értelmezésben „tágabb” vagy „szűkebb” összefüggés (az ún. „laza hierarchia”) alkalmazásának, mint például AGÁR-VADÁSZAT, ASZTAL-BÚTOR-IPAR, LOBBANÁSPONT-HÓTAN. Az utóbbiak valójában különféle irányított relációkat képviselnek, mint a „tevékenység-tevékenység eszköze”, „rész-egész” stb. A magyar szabvány előírásai következetesen alkalmazzák az alá-fölérendeltségi viszonyt. Az ISO 2788 nemzetközi szabványban a „laza hierarchiát” is engedélyezik, de lehetőséget adnak a „szigorú hierarchia” megkülönböztetésére (BT/NT, illetve BTG/NTG relációjelek formájában).

A másik következménye az automatikus relációkezelésnek az, hogy a gépi teauruszokban olykor a szokásosnál több relációtípust különböztetnek meg. Így például a MÉM Növényvédelmi és Agrokémiái teauruszában a „növény-károsító”, „növényvédőszer-károsító” és „károsító-betegség” relációtípusokat is használják.

Gépi teauruszokban azért értelmezik és különböztetik meg pontosabban a relációkat, hogy a keresőképző automatikus módosítása valóban automatikusan történhessen, logikailag megbízhatóan definiált relációtípusok alapján. Ezek használatakor előre tudhatjuk, milyen irányban terjesztődik ki a kérdés, s nem kell minden esetben a teauruszban is ellenőrizni, hogy az adott relációval milyen fogalmakat kapcsolnak össze.

Gyakori eset, hogy fontosabb fogalmaknak sok az alárendeltje, ezek azonban különféle szempontok szerint csoportosíthatók (pl. „felhasználás”, „működés” stb.). Az ISO szabványban erre az esetre azt javasolják, hogy a hierarchialáncot fel kell osztani e szempontok szerint ún. „elágazási szintekre” („node label”), ezeket megfelelő megjegyzésekkel jelölve. Mivel a megjegyzések gépi kezelése körülményes, a relációjeleké viszont nem, a gépi teauruszokban nem megjegyzésekkel jelölik az elágazásokat, hanem hátravetett értelmezővel annyi fölrendelt deskriptort alkotnak, ahány szempont szerint az alárendeltet csoportosítani kívánják.

Pl.

REPÜLÉS	
REPÜLÉS (RAKOMÁNY SZERINT)	REPÜLÉS (HASZNÁLÓK SZERINT)
F REPÜLÉS	F REPÜLÉS
A SZEMÉLYSZÁLLÍTÓ	A KATONAI REPÜLÉS
REPÜLÉS	KERESKEDELMI
TEHERSZÁLLÍTÓ	REPÜLÉS
REPÜLÉS	KUTATÓREPÜLÉS
	SPORTREPÜLÉS

A megoldás előnye, hogy a deskriptort képviselő, mesterséges homonimák segítségével kereséskor automatikusan utalni lehet egy közvetlen gyűjtőfogalomra, ha az egyes speciális fogalmak alapján nincs elég találat.

Ha az elágazások egyéb relációkon alapulnak, akkor egyes gépi teauruszokban önálló relációval (pl. „eszköz”, „rendeltetése”, „növényvédőszer”, „betegség” stb.) jelölik az elágazást, nem pedig megjegyzéssel. Pl.

KÖNYV	és nem	KÖNYV
T KÖNYVSOROZAT		T KÖNYVSOROZAT
P CÍMLAP		P CÍMLAP
GERINC		GERINC
R TÁROLÁS		X BETŰ
KÖTÉS		KÖNYVÉSZET
NYOMÁS		<i>Művelet szerint:</i>
X BETŰ		TÁROLÁS
KÖNYVÉSZET		KÖTÉS
		NYOMÁS

A szabványban javasolt megoldás kirívó hibája — amiért a „manuális teauruszokban” sem válik be —, hogy tisztázatlan marad az összefüggés inverze.

3.2 A szintaktikai relációk hatása

Az információkereső nyelv szintaxisa lényegében a deskriptorlánc szerkesztési szabályainak összessége. A STAIRS, GOLEM, AIDOS lehetővé teszik a deskriptorlánc elemeinek összekapcsolását, biztosítva a „mondatok” egyértelműségét. A „magyar autók exportja Franciaországba” tényállást például a következőképpen különböztetik meg az AIDOS-ban a „francia autók exportja Magyarországra” tényállástól:

MAGYARORSZÁG-SZEMÉLYAUTÓ, EXPORT, FRANCIAORSZÁG;
FRANCIAORSZÁG-SZEMÉLYAUTÓ, EXPORT, MAGYARORSZÁG;

ahol a csillag egyszerű kapcsolatjelölőt — angolul: „link” — képvisel.

A GOLEM-ben arra is lehetőség van, hogy az ún. „indexek” segítségével a különböző deskriptorláncokba felvett elemek között is kapcsolatokat jelöljenek.

A kézi rendszerekben — pl. cédulakatalógusokban — szintaktikai relációkat nem használnak, s ugyanez vonatkozik az egyszerűbb információkereső programrendszerekre is, amilyen az ISIS vagy a SZIV. A fejlett programrendszerekben használt szintaktikai relációk alapvetően befolyásolják a teaurusz megnevezéseinek fajtáit és szerkezetét, különösen pedig a teauruszal indexelt állományokról készült mutatók felépítését.

Mivel a többértelműség a szintaktikai reláció alkalmazásával kiküszöbölhető, a GOLEM, az AIDOS vagy a STAIRS tezauszaiiban problémamentesen vehetők fel uniterm – egytagú – kifejezések. A szabványok szerint azokat az összetett kifejezéseket, melyek az összetevők sorrendjétől függően mást és mást jelentenek (pl. OLAJ-TRANSZFORMÁTOR és TRANSZFORMÁTOROLAJ), összetett formában – prekombináltan – kell felvenni a tezauszba. A GOLEM-ben például nincs erre szükség, a deszkriptorláncban az OLAJ és a TRANSZFORMÁTOR deszkriptorokat utólagosan – posztkombinálva – is egyértelműen kezelni lehet:

TRANSZFORMÁTOR::OLAJ;;
 illetve
 OLAJ::TRANSZFORMÁTOR;;

ahol a szintaktikai reláció jele a kettőspont, melyet azért alkalmaznak kettőzve, hogy a tárolt kifejezésekben, illetve szövegekben írásjelként egyedileg továbbra is használni lehessen. Az ilyen összetett kifejezéseket – ún. „összetett deszkriptorokat” – mindig csak az első tag, az ún. „aspektus” szerint lehet először keresni, ez az elem képviseli tehát a „fókuszot”. Attól függően, hogy melyik kifejezés kerül az aspektus helyére, képviseli az összetett deszkriptor az egyik vagy a másik jelentést.

Ugyancsak a szintaktikai relációk alkalmazhatóságából következő nagyobb biztonsággal magyarázható, hogy a gépi tezauszokban sokszor az adott szakterületen fontosnak számító összetett kifejezéseket sem veszik fel, hanem csak morfológiai összetevőiket, holott a szabványok ezt kifejezetten előírják. Erre olyankor kerül sor, amikor az adott kifejezésnek nincs a hierarchikus szerkezetben fölérendeltként összefoglaló szerepe, s ezért az automatikus hierarchiaszint emelés szempontjából nem játszik szerepet. (Az automatikus hierarchiaszint emelés kihasználhatósága érdekében viszont sokszor használnak összetett, mesterségesen alkotott, a szaknyelvben olykor nem is használt kifejezéseket, hogy velük az egyébként fölérendelt deszkriptorral nem rendelkező deszkriptorokat a tezauszban összefoghassák [XXI].)

A szintaktikai relációk használatának lehetősége fokozza a tezausszal végzett tartalmi feltárás analitikus jellegét, ami a melléknevek és egyéb szófajok deszkriptorként való használatában is megnyilvánul, bár ezt a szabványok nem ajánlják. A magyarországi gépi tezauszokban például nemcsak az olyan kifejezések szerepelnek, mint BILLENŐ, FRONTÁLIS, GÉPI stb., hanem olyan morfémák – töredékszavak – is, mint MIKRO-, MENTES, NÉLKÜLI stb. A szabványok szerint az utóbbiak szereplése kizárt az információkereső tezauszokban.

3.3 Az automatikus indexelés hatása

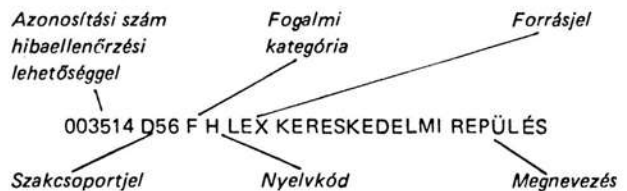
Az automatikus indexeléshez a programrendszerben a szavak töveit az ún. tőszótárban tartják nyilván. Ebben a főnevek és melléknevek tövei mellett az igetöveket is feltüntetik. A hatékony információkereséshez a tartalmilag közel álló töveket – szófajukra való tekintet nélkül – össze kell vonni és közös deszkriptorral kell képviselni. Ennek következtében a gépi tezauszban a szótövek vagy – például – az igék is megjelennek, külön relációjellel kapcsolódva a vezéredeszkriptorokhoz.

Az ilyen gépi tezauszoknak valójában része az ún. „negatív tezausz”, mely a tartalom szerinti keresésből kizárt, jelentés nélküli, üres (negatív, nulla-) deszkriptorokat tartalmazza. Ilyen lehetőségekre nyílik mód a STAIRS programrendszerrel.

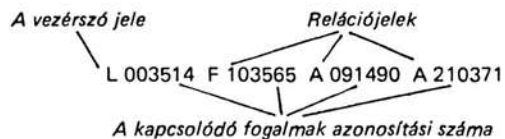
3.4 A számítógéppel támogatott tezauszkészítés hatása

A tezauszok készítéséhez használt programrendszerek vagy megnevezéseket, vagy megnevezéseket és relációkat kezelnek. A kezelés szempontjából megkülönböztetik a megnevezés- és a relációtételeket.

A megnevezéstétel felépítése a TSPR-ban például a következő:



A kapcsolattétel felépítése:



A végső kinyomtatásban a megnevezéstételben szereplő egyes adatok tetszés szerint elhagyhatók. A tétel felépítésétől függ tehát, milyen formában jelenik meg a géppel kinyomtatott tezausz. Ma az a helyzet, hogy ahány programrendszert készítenek, annyiféleképpen alakítják ki a megnevezés- és kapcsolattételt, s ettől függően változnak a gépi tezauszkinyomtatások.

Az információkereső rendszerekben a megnevezéseket általában nem kapcsolják össze belső azonosítási számmal, ezért minden adatkezelés magához a különböző hosszúságú megnevezéshez kapcsolódik. A kapcsolattételben tehát nem két számot, hanem – mint például az

AIDOS esetén – két teljes megnevezést kell kiírni, ami lényegesen körülményesebb, továbbá a módosítások alkalmával hibaellenőrzésre sincsen mód.

A gépesítés hatásával a szabványosítás – mint erre korábban rámutattam [XXIII] – *nem foglalkozik*, holott a gépi tezauszok terjedésével ez a hatás egyre erősebben érvényesül.

Megjegyzések

- [1] Az adatok értelmezhető, az információk értelmezett ismereteket képviselnek. A számítástechnikában mindig az adattétel, a dokumentációban – és a továbbiakban e tanulmányban is – többnyire az információtétel kifejezést használják.
- [2] A könyvtári és dokumentációs szakirodalomban [IV] inkább egy másik szóhasználat honosodott meg. Eszerint az adatbázist az információkereső rendszer passzív részének, a kezelőrendszert pedig az aktív részének nevezik. Ebben az értelmezésben tehát az információkereső rendszer gyűjtőfogalom. Ha nem programrendszerről van szó, akkor az információkereső rendszer aktív része azonos az információkereső nyelvvel, mivel a „kezelőrendszer” ebben az esetben az emberi tudatban van. Véleményem szerint az információkereső rendszer fogalmát nem célszerű ilyen általánosan értelmezni. Automatizált körülmények között ez a rendszer nem más, mint maga a kezelőrendszer. Alátámasztja ezt, hogy a szöveges, bibliográfiai adatokat kezelő programcsomagokat készítők általában információkereső rendszernek nevezik. Pl.: STAIRS = Storage and Information Retrieval System; SZIV = Szöveges Információvisszakereső Rendszer.
- [3] Az NDK-ban 1974-ig elkészült 23 tezausz esetében átlagosan 4300 [XII], az 1980-ig elkészült 15 magyarországi tezausz esetében átlagosan 11 000 [XXII], a többnyelvű tezauszokra vonatkozó 1980-as felmérés szerint átlagosan 5800 lexikai egység esett egy tezauszra.
- [4] Az AIDOS információkereső rendszer DOS változatának elkészítéséhez pl. 300 emberre volt szükség [I], ami kb. 60 ember 5 évi munkájának felel meg. Az OS változat további emberéveket vett igénybe.
- [5] Csak az AIDOS legújabb, OS operációs rendszerben működő változata felel meg ennek a színvonalnak. Az eredeti – nálunk adaptált – DOS változatban az előny értékét kérdésessé teszi néhány szigorú megszorítás. Például a nem-faj reláció kizárólag monohierarchikus lehet, rajta kívül csak egyetlen további deskriptorok közötti – rokonsági – reláció létezik, melyet ráadásul csak ugyanazon szakcsoportba sorolt deskriptorok között lehet meghatározni.
- [6] A GOLEM-et annak idején a müncheni olimpiára készítették; információkereső előírásait a laikus számára könnyen érthetően fogalmazták meg, hogy a rendszert egyáltalán nem ismerő újságírók is használhassák.
- [7] A SZÁMALK-ban használt ISIS-BABILON az egyetlen olyan magyarországi online adatbázisrendszer, melyhez bárki szabadon hozzáférhet. A GOLEM csak a felhasználó vállalatok – SZKI, MÉM NAK, ÉTK – illetékes szakembereinek áll rendelkezésére.

- [8] Ez a rendszer nem tévesztendő össze a közel 180 adatbázist kezelő amerikai DIALOG Information Retrieval Service nevű ismert rendszerrel, mellyel az utóbbi években már Magyarországon is – pl. az SZKI-n és az OMIKK-on keresztül – kapcsolatot teremtettek. Az NTMIK DIALOG rendszerének üzembehelyezésére az OMIKK-ban kerül sor [XX].
- [9] A BIBDOSZ nemzetközi mércével mérve is igényes vállalkozás volt, mely alkotójának, *Pethes Ivánnak* korai halálával lényegében kereskedelmi forgalmazáshoz érett dokumentáció nélkül maradt, s ma már alkalmazását is megszüntették az Országos Vezetőképző Központban.
- [10] Az egyik legjelentősebb hazai online fejlesztés [XX] keretében minden jel szerint két olyan programcsomag – a CDS ISIS és az NTMIK DIALOG – beszerzése mellett döntöttek, melyek információkereső nyelvet kezelő képessége sajnos elavult. A fejlett rendszerek beszerzéséhez feltehetően nem biztosítják az anyagi fedezetet. Kétséges azonban, hogy hosszú távon az elavult termékek beszerzése a gazdaságos.
- [11] A 70-es évek elején a KSH-ban használt TEXT-PACK információkereső rendszerben nyilvántartott kulcsszavakhoz egy másik hazai tezauszalkészítő és nyilvántartó programrendszer kidolgozásához is hozzákezdtek, de az elkészült modulok csak a megnevezések szintjén működtek, relációkat nem tudtak kezelni [III].

Bibliográfia

- [I] SOPS AIDOS Generierungsvariante DOS/ES Version. Berlin, ZIID, 1976. 98 p.
- [II] BALÁZS K.: Felzárkózás – követés – lemaradás. Egy intézettörténeti ásatásból = Valóság, 24. köt. 8. sz. 1981. p. 18–31.
- [III] BALOGH Z.: Tezausz előállítására alkalmas programcsomag = Számvitel és Ügyviteltechnika, 1973. I. melléklet. 33 p.
- [IV] Fogalomgyűjtemény az informatika válogatott területeiről. Összeállította: Várady É. [et al]. Bp. OMKDK, 1972. 191 p.
- [V] GOLEM (BS 2000) Softwareprodukt. Siemens-System 7000/4004. Datenbankdienste. Bedienungsanleitung. Teil 1. und 2. München, 1976.
- [VI] HALASSY B.: Adatbázisok kezelésének alapvető kérdései. Bp. SZÁMOK, 1978. 440 p.
- [VII] A HIR 2 rendszer alapelve, működése, jellemzői. Fejlesztési javaslat. Összeállította: Csáky Béla. Bp. SZÁMKI, 1981. 26 p. (SZÁMKI 2364/81)
- [VIII] ILJON, A.: Le logiciel ASTUTE: un outil informatique pour la création, la mise à jour l'édition de l'impression de thésauri mono- et multilingue = Documentaliste, 14. köt. 5–6. sz. 1977. p. 25–33.
- [IX] CDS/ISIS: Általános leírás. Tervezet. Párizs, Unesco CDS, 1978. március 55 p. [SZÁMALK fordítás].
- [X] Hazai szoftver kínálat 1982. Bp. KSH. 15•137•VIII p.
- [XI] KUNSZT GY.: A tudományos kutatás logikái modellezése és tematikai irányítása. Bp. Akadémiai Kiadó, 1974. 435 p. /Tudományszervezési füzetek/

- [XII] MANECKE, H. J.: Statistische Analyse von in der DDR angewendeten Thesauri = Wissenschaftliche Zeitschrift der TH Ilmenau, 20. köt. 4/5. sz. 1974. p. 213–217.
- [XIII] MIR. Általános felhasználású információ tároló és visszakereső rendszer. Rendszerterv. Készítette: Informatika Gmk. Bp. 1982. augusztus. 28 p.
- [XIV] Az NTMIR programcsomagjai: DIALOG, SZKIF. Bp. OMFB, 1981. 62 p. /NTMIR dokumentumai, 19./
- [XV] PETHES I.–TOLNAI GY.: BIBDOSZ. Egy megvalósított komplex számítógépes könyvtári és dokumentációs rendszer = Könyvtári Figyelő, 24. köt. 5. sz. 1978. p. 495–512.
- [XVI] SCHLANGER L.: Tezauszépítés számítógépes segítségével. Bp. OSZK–KMK, 1983. 280 p.
- [XVII] SLYPE, G., van: Qualitative und quantitative Merkmale ein- und mehrsprachiger Thesauri = Dritter europäischer Kongress über Dokumentationssysteme und -netze. München. Kommission der europäischen Gemeinschaften, Verlag Dokumentation, 1977. p. 355–367.
- [XVIII] IBM System/370. Storage and Information Retrieval System/Virtual Storage – Thesaurus and Linguistic Integrated System. General Information Manual GH12–5123–1. Second Edition (November 1976) 38 p.
- [XIX] SUMAROKOV, L. N.: Pakety prikladnyh programmi dla obrabotki naučno–tehnicheskoj informacii: sostojanie a perspektivy = Problemy MSNTI 1/1982. p. 3–22.
- [XX] SZABÓ I.–KERTÉSZ J.–SZARKA GY.: Számítógépes információrendszer fejlesztési elképzelései az OMKDK-ban = Információ–Elektronika, 16. köt. 5. sz. 1981. p. 292–294.
- [XXI] Számítógépes Információvisszakereső Rendszer (SZIV). Felhasználói dokumentáció. Alkalmazói leírás. Bp., SZKI, 1974. 84 p.
- [XXII] UNGVÁRY R.: Tezausz-technológia. Az információkereső tezauszok készítésének folyamata. Bp. OSZK–KMK, 1979. 277 p.
- [XXIII] UNGVÁRY R.: A magyarországi tezauszok a nemzetközi szabványosítás tükrében = Tudományos és Műszaki Tájékoztatás, 30. köt. 8–9. sz. 1983. p. 305–330.

UNGVÁRY Rudolf: A tezauszok automatizálása. A gépi tezauszok készítése és használat néhány kérdése

Az automatizált információkereső rendszerek által kezelt gépi tezauszok tulajdonságait jelentős mértékben befolyásolják a programrendszer adottságai. Az automatikus nemdeszkriptor–deszkriptor utalás következtében külön relációjellel jelölik a valódi, és külön relációjellel a kváziszinonimián alapuló kapcsolatokat. Az automatikus deszkriptorok közötti relációk kezelése következtében a szokásosnál több, és logikailag szigorúbban értelmezett relációk használatára kerül sor. A szintaktikai relációk használata következtében bizonyos szóösszetételek nem szerepelnek lexikai egységként a tezauszban, mivel a posztkombinációkban biztosítható az egyértelműség. A melléknevek és egyéb szófajok kifejezései is gyakrabban szerepelnek deszkriptorként. Az automatikus indexelésnek is hasonló a hatása. Az információkereső programrendszerek tezauszkezelő képességeit a beszerzéskor nem veszik kellő mértékben figyelembe, holott e képességek szempontjából rendkívül nagy különbség van az egyes termékek között.

UNGVÁRY, R.: Automation of thesauri: its effect on their structure

The capabilities of computer programs affect considerably the characteristics of mechanized thesauri applied in automatic information retrieval systems. Due to automatic reference from nondescriptors to descriptors, relations based on real synonyms and quasi-synonyms are designated differently. In consequence of the automatic handling of relations between descriptors, more and semantically more precise relations are used than usual. As a result of using syntactical relations certain compound words are not included in the thesauri, because the unanimity can be assured through post-coordination. Adjectives and other parts of speech are used as descriptors more often than usual. Automatic indexing results in identical problems. During acquisition, more attention should be paid to the language component of the available programs, because they differ considerably in this respect.

УНГВАРИ, Р.: Автоматизация тезаурусов и ее влияние на их структуры

Свойства машинного тезауруса, управляемого автоматизированной информационно-поисковой системой, в значительной степени подвержены влиянию особенностей системы программ. Вследствие автоматической ссылки недескриптор-дескриптор, обозначают отдельными знаками отношений действительные связи и отдельными знаками связи, базирующиеся на квазисинонимии. Вследствие управления связей между автоматическими дескрипторами используется большее количество логически более строго выраженных связей. Вследствие использования синтаксических отношений некоторые словосочетания не присутствуют в тезаурусе в качестве лексических единиц, так как в посткомбинациях обеспечивается однозначность. В качестве дескрипторов часто встречаются и прилагательные и другие части речи. Аналогичное влияние имеет и автоматическое индексирование. При приобретении информационно-поискового пакета программ не обращается должного внимания на его свойства по ведению машинного тезауруса, хотя с точки зрения этих свойств имеется очень большая разница между отдельными системами.

UNGVÁRY, R.: Die Automatisierung der Thesauri und ihre Auswirkung auf die Thesaurusstruktur

Thesauri können – als integrierter Teil eines Informationsrecherchesystems – automatisiert verwendet, bzw. – mit Hilfe spezieller Programmsysteme – „automatisiert“ erstellt werden. Die Eigenschaften der Thesauri werden in bedeutendem Masse durch das jeweilige Programmsystem beeinflusst. Infolge der automatischen benützung der Beziehungen zwischen den Nichtdeskriptoren und Deskriptoren wird die Äquivalenzrelation und die Quasiäquivalenzrelation unterschiedlich gekennzeichnet. Infolge der Möglichkeit, die Beziehungen zwischen den Deskriptoren ebenfalls automatisch zu benützen werden mehr Relationen als üblich verwendet und sie werden semantisch genauer definiert. Infolge der Möglichkeit, beim Indexieren syntaktische Relationen anzugeben werden gewisse Komposita nicht als lexikalische Einheiten im Thesaurus aufgenommen, da bei der Postkombination die Eindeutigkeit gesichert werden kann. Adjektive sowie sonstige Wortarten werden des öfteren als Deskriptoren aufgenommen. Auch das automatische Indexieren führt zu ähnlichen Ergebnissen. Bei der Beschaffung von Informationsrecherchesystemen wird der Bedeutung der Sprachlichen Verwaltungskomponente meistens ungenügend Aufmerksamkeit geschenkt, obwohl hinsichtlich der „sprachlichen“ Fähigkeiten zwischen den handelsüblichen Softwareprodukten sehr grosse Unterschiede bestehen.

A KÖZELJÖVÖBEN JELENIK MEG

A KGST-tagországok Nemzetközi Tudományos és Műszaki Információs Rendszere. Átfogó ismertetés

(Az NTMIR dokumentumai sorozat 22. füzeté)

1976 óta nem jelent meg átfogó ismertetés az NTMIR-ről. A végbement változásnak és fejlődésnek megfelelően az eltelt időszak szükségessé tette egy új átfogó ismertetés kiadását. A kiadvány részletesen bemutatja az NTMIR szervezeti felépítését, az NTMIK tevékenységét, a nemzetközi speciális információs rendszerek működését és a nemzetközi ágazati információs rendszerek működését. Kitér az információs rendszerek működésének biztosításához szükséges elméleti-tudományos kérdésekre, így a géppel olvasható információhordozókra, a kompatibilitás és szabványosítás kérdéseire, valamint a programellátás helyzetére. Külön fejezetet szentel az információs dolgozók képzése és továbbképzése terén megvalósuló együttműködésnek.

A kiadványt ábrák és táblázatok teszik szemléletessé, terjedelme kb. 10 szerzői ív, ára 90,— Ft.

Megrendelhető az OMIKK Értékesítési osztályától: (1428 Budapest, Pf. 12.)