

# Az általános tárgyszójegyzék munkálatairól

Horváth Tibor

Valószínű, hogy nagyobb szakmai érdeklődésre tarthat számot az általános tárgyszójegyzék ügye, amelynek munkálatai már elindultak. Alább közreadjuk azt az előterjesztést, amely a Művelődési és Közoktatási Minisztérium (MKM) számára készült, majd három szakmai részproblémát tárgyaló fejezetet. Kérem a kollégákat, az előterjesztés fejezetét úgy olvassák, hogy nem az indexelési kérdések specialistáinak készült, így evidenciákat is tartalmaz. Az MKM számára készült javaslatból csak bizonyos technikai részleteket hagytunk el, a szöveget változatlan formában közöljük.

---

## Az előterjesztés

---

I.

Amint a nagyobb méretű adatbázisoknak, pl. a nemzeti dokumentumok adatbázisának építése napirendre került, kiviláglott a magyar könyvtárügy egyik alapvető hiányossága: tematikai keresést nem képes biztosítani, mivel nincs magyar nyelvű általános tárgyszójegyzék. A fontosabb nyelvterületek mindegyike elkészítette idejében ezt az intellektuális eszközt, nem beszélve az angolszász világról, ahol eleve a múlt századtól az ún. mellérendelő osztályozás/indexelés vált uralkodóvá. Hazai könyvtáraink a tematikai/tartalmi feltárássra az Egyetemes Tizedes Osztályozást alkalmazzák általában. Az ETO kezelésére (speciális rendezés, visszakeresés, táblázat építés) születtek szoftverek, alapvető probléma azonban, hogy a) igen nehézkesen idomul a számítástechnikai környezethez, b) az örökös fejlesztés ellenére, semmilyen szoftver nem segít a 19. századi tudományszemléletén, c) feloldhatatlan az a komplementaritása, hogy minél finomabb és összetettebb a jelzet, annál veszteségesebb és zajosabb a visszakeresés és viszont. Főleg a nemzetközi együttműködés miatt, számos szakterületen ágazati tézauruszokat munkáltak ki – ennek hazai szabványa is létezik –, ezek azonban minden esetben csupán egy szűkebb tudományt vagy szakterületet fednek le.

Az általános tárgyszórendszer hiánya immár az adatbázisok építésének, ezáltal az elektronizáció további térhódításának legfőbb akadálya. Sürgetően léptek fel ennek kialakítására: az MTA Központi Könyvtára, a tudományegyetemi könyvtárak és az egyetemi könyvtárak igazgatói, a Fővárosi Szabó Ervin Könyvtár, a megyei könyvtárak igazgatói tanácsa,

részben az OSZK, a Magyar Könyvtárosok Egyesülete.

A munkálatok elindulásának legfőbb akadálya az volt, hogy e hatalmas feladat – lényegében az ETO-nál nagyobb lélegzetű munkáról van szó – mindegyik könyvtár kapacitását, teljesítőképességét, felkészültségét magasan meghaladta, így mindenki csak sürgetését és óhaját tudta megfogalmazni.

Tipikus összkönyvtári feladatról van szó. Mivel a munka elkezdése nem tűr halasztást, olyan szervezeti megoldást kell találni, amely lehetővé teszi a különböző könyvtárakban dolgozó szakértők bevonását. A feladat irányítása pedig olyan szakemberekre bízandó, akik értékelt produktumokat tettek már le ezen a területen és a tárgyszókészítési know how biztos ismerői. A probléma az, hogy összkönyvtári munkaként egy helyről kell a finanszírozást és vezetést megoldani.

### II.

Maga a tárgyszórendszer három (négy), egymásra épülő eszközként képzelhető el, hasonlatosan egy piramiselvű építkezéshez. A csúcstól egy néhány ezer tárgyszóból álló, generikus kifejezéseket tartalmazó, de minden részében kiépített szint képezheti. Kisebbségi könyvtárak alkalmazhatják, egyben ez a rész töltheti be az „összekötő nyelv” funkcióját is, kb. BSO-funkciókkal (Broad System of Ordering).

A középső szint a fenti részletezése, de osztályozó fogalmaiban teljes összhangban a fentivel. A fenti szint lényegében ennek generikus, „csúcs” tárgyszavait tartalmazó része. Középméretű könyvtárak és nagyobb könyvtárak a nem fegyűjtőkori dokumentumok indexelésére alkalmazhatják.

A harmadik szint részletes, specifikus tárgyszavakkal is rendelkező, mély indexelést tesz lehetővé, a fenti két szinttel hasonló viszonyban, ahogyan az előzőekben jellemeztük. Technikailag ezt kell elkészíteni, majd ebből lehet a fenti szinteket kivonni, perfektt átjárást biztosítva a szintek között.

A negyedik szint az ágazati tézauruszok képezik, amelyek már elkészültek vagy készülni fognak, de nem ennek a munkálatnak a keretében. A „közlekedés” ezekben úgy biztosítható, hogy a tárgyszómunkák során a tézauruszok generikusabb kifejezései beépülnek a felettes szintekbe.

Minden szint lexikai egységei ugyanúgy épülnek fel; lexikai egység lehet tárgyszó és utaló. A tárgyszavak ne unitermesített, de ne is szélsőségesen összetett alakban jelenjenek meg. (Pl.: jonatán almák betegségei – jonatán alma és betegség. De nem jonatán és alma és betegség.)

A tárgyszó kiegészül értelmezővel, szinoníma kapcsolatok feltüntetésével (szimmetrikus relációként, ezért kívánatosak lexikai egységként az utalások), szakcsoportot (tudomány, szakterület) feltüntető kódokkal, a „szint” utalásaival (ld. fentebb), legyen lehetőség idegen nyelvű ekvivalensek megjelenítésére, és maradjon üres hely helyi célokra. Egyetlen további utalás képzelhető el: LM (lásd még), amely az összes lehetséges (nem szinoním) kapcsolatot kifejezi. Ezt helyileg finomítani lehet. Az LM szintén szimmetrikusan lenne kiépítve.

A földrajzi (hely) és az időt (időpont, tartam stb.) kifejező tárgyszavak relatíve ritkák, de szabály vonatkozik ezek használatára és forrásaira.

A munkálatok fázisában szakcsoportos részek munkaeszközként feltétlenül kell hogy készüljenek. Átmenetileg erre a feladatra az ETO vagy a BSO használható, akár mindkettő.

E tárgyszórendszer rendelkezésre kell álljon géppel olvashatóan, adatbázisba szervezve és nyomtatott formában is. Aktualizálása évente, később ritkább periódusokban.

Fontos tartozékok lennének az egymásra épülő használati szabályzatok, kb. három, fokozatosan finomodó indexelési utasítás. Az egyszerűbbtől egy kb. PRECIS bonyolultságú és magas kvalifikációt igénylő eljárásig. Valamennyi szabályzat minden szinten alkalmazható. E használati szabályok következtében az általános tárgyszójegyzék gyökeresen különbözne minden eddigitől, hiszen ezeknek a lényeges részeknek a megadásával ágazati tárgyszórendszerek és tézauruszok is adósak.

### III.

Néhány technikai részletet szükséges még megjegyezni.

1. A források összegyűjtése elkezdődött: az LC tárgyszójegyzék, a BSO rendelkezésre áll, német nyelvterületen is tájékozódunk. Az ETO index, számos ágazati tárgyszójegyzék, a tudományos nomenklatúrák elérhetők. (Ez természetesen nem azt jelenti, hogy valamelyik külföldi forrást vesszük át!) A hazai források értelemszerűen hozzáférhetők.

2. Rendelkezésre áll egy szoftver, amely szótárak adatbázisba szervezését teszi lehetővé és támogatja tárgyszójegyzék építését, pl. az utalások oda-vissza ellenőrzését.

3. A szakértői stábhoz a következő elképzelések alakultak ki. A fentebb felsorolt könyvtárak munkatársaiból szükséges a csoportot megalkotni, vigyázva arra, hogy nem elvi ötletekre van szükség, hanem főként

operatív munkára. Szeretnénk a munkához pályakezdeket, fiatalokat is bevonni az MKE-n keresztül, főleg gyakoroltatási szempontból.

4. A munka költségei nehezen becsülhetők, hozzávetőlegesen nettó 7-8 millió forintból a tárgyszójegyzék kiépíthető. Ez két-három részletben esedékes, ugyanis a munkák két-három évig tartanak. A pénz telepítésére az az irányadó, hogy egy helyre kerüljön és kizárólag e munkálatokra lehessen igénybe venni. Ez a hely lehet akár egy társadalmi szervezet, akár valamelyik olyan könyvtár, amelynek munkatársai intenzíven részt vesznek a munkában.

#### IV.

A jelen előterjesztést szakmai fórumokon szükséges megvitatni. Jóváhagyása után a munkálatok elindíthatók. A feladat finanszírozását az MKM-től kérjük. Ennek oka, hogy a központi feladatok támogatása minisztériumi feladat. Mivel a munka több évre elhúzódik, a pályázati rendszerbe való indítását kockázatosnak ítéljük.

Eddig tart az előterjesztés. Ebben tárgyszórendszeréről esik szó, nem tárgyszójegyzékről. A tárgyszórendszer ennél több. Ennek, továbbá más kérdéseknek kifejtését ezúttal mellőzzük. A három – talán legizgalmasabb – résztémát alább ismertetjük.

### A logikai egységek

#### V.

A lexikai egységek kiépítettségének megítéléséhez érdemes elméleti megfontolások alapján tisztázni, hogy az indexelés folyamatában valójában *mit tekintünk szótári feladatnak, azaz magának a tárgyszórendszernek mit kell teljesítenie, s mit az indexelési folyamat eredményeként született indextételeknek.* A kérdés történetileg is érdekes, valójában egy fejlődésmenet meghatározott érettségi fokán született kielégítő megoldás. Nézzük a problémát ebben a vetületben.

A hierarchikus rendszerek abban voltak lenyűgözőek, hogy egy világnézetet tükröztek, sőt világtrendet. A választott felosztási alapok gyökerei filozófiai rendszerekbe nyúltak; minden valódi filozófiai rendszer megteremtette a maga tudományrendszerét. Még inkább: átfogó gondolko-

dók kisebb nagyobb részletességgel tudományrendszereket is alkottak. Az osztályozási rendszereknek ezt a nagyhozamú áradását most nem érdemes felidézni részleteiben, de *Jaspers* magyarázatát mégis érdemes feleleveníteni. Arról van szó, hogy a tudomány vagy filozófia mindig tisztének tartotta egy szintézis vagy világkép megalkotását, persze tudományos igénnyel. Ám a huszadik századi tudomány felgyorsult fejlődése nyomán az empirián túli, az érzékelhetőség számára hozzáférhetetlen valóságszférák kerültek a kutatás vizsgálódási körébe; számos, a komplementaritás elvére épülő törvény született (csak a heisenbergi határozatlansági relációra kell gondolnunk), de egyéb okokból is egyre inkább reménytelenné vált a szintézis, az egységes tudományos igényű világkép. A huszadik század filozófiája lemondott erről, s ezért sóhajt fel *Jaspers*: *„Nincs világkép, csak a tudományok rendszere van.”*<sup>1</sup>

A tudományrendszerezésnek ezt az (egyébként teljesületlen) funkcióját mutatja az a fejlemény is, hogy a marxista filozófiának – mert a világmagyarázat ellentmondás-mentes szintézisét ígérte, így a tudományok haladtával egyre primitívebb világképe egységéért egyre inkább lemondott a tudományosságról – utolsó szakaszában virágzásnak indult a tudományrendszertan élvonalbeli szovjet filozófusok vezetésével.

Az osztályozási rendszerekkel szemben volt tehát egy néha ki is mondott igény: a tudományok rendszerezésén keresztül az egész valóság reprezentációja legyen. (Itt nem térnék ki *Ludvig von Bertalanffy* más közelítésében hasonló szintézisalkotó igényére.) Nos, amikor elhagytuk a hierarchikus osztályozási rendszerek piramisait, ez a néha ki is mondott igény megmaradt, kísértett a tézauruszok elméletében. Mellesleg: a tézauruszokat inkább felszíni közelítésben lehet az angolszász mellérendelő indexelési eszközök és eljárások örököseinek tartani; mélyebb alapjaiban, szellemében a tudományrendszerek utódai. De ez ismét egy más kérdés, amelyre itt nem jut hely kitérni.

Dehát, hogyan lehetett pl. egy tézauruszban mindezt megoldani? A deszkriptorok relációinak egyre finomabb kidolgozásával. Ám egy tézauruszban vagy szótárban csak olyan reláció jelezhető meg, amelyik kontextus nélkül, eleve, a priori igaz. Olyasmi, minthogy a szék az ülőbútorok fajtája, speciese, hogy az algebra a matematika része stb. Más szóval: a tézaurusz az univer-

zum reprezentációja legyen, csak másként, mint a tudományrendszerek; feltételezve, hogy az univerzum ilyen fogalmi hálóban leképezhető.

Ne firtassuk azt a kérdést, hogy leképezhető-e; inkább azzal foglalkozzunk, hogy erre van-e szükség?

Aligha. Az indexelésnek ugyanis van néhány alapszabálya. Ezek egyike az ún. relációmegőrző transzformáció elve. Sokan *Landry*-posztulátumként emlegetik.<sup>2</sup> Ez kimondja, hogy egy eredeti szöveget (forrásszöveget) úgy kell átvinni a reprezentációjába, azaz az indextételébe, hogy a tárgyszavak-deszkriptorok stb. között azok a relációk maradjanak érvényesek, amelyek az eredeti szövegben érvényesek. Azaz: az indextétel köteles megőrizni az eredeti szöveg fogalmai között létesített relációt. Mármint, minél eredetibb egy közlemény, a fogalmak közötti asszociációk annál szokatlanabbak, váratlanabbak, annál kevésbé kezdenek hasonlítani azokhoz a relációkhoz, amelyeket szótárilag meg lehet határozni. E relációk a posteriori jellegűek. Példával élve: szótárilag mondjuk egy tézauruszban a „busa” a pontyalakúak speciese, relációba hozható még a növényevő halakkal, az édesvízi halakkal stb. De ha egy közlemény a busa étrend vérnyomáscsökkentő hatásáról szól, akkor a reláció a busa és a vérnyomáscsökkenés között áll fenn, s ez tükröződik az indextételben, bár nincs a szótárban. Márcsak azért sincs, mert minél több a közlemény, annál többféle az asszociáció, az új, az értékes asszociáció még nem is kerülhet a tézauruszba.

A relációmegőrző transzformáció követelménye a szótárban, a tézauruszban nem valósítható meg. Kivitelezhető viszont az indextételben oly módon, hogy az indextétel meghatározott szabályok alapján készül. Ez kötelező kell legyen: mert ha csak egyszerűen hozzárendeljük a tárgyszavak egy részhalmazát egy közleményhez, minden szabály nélkül, akkor az olyan, mintha egy épület alkotóelemeit összekupacolnánk, téglát, betont, gerendát, vezetékét, szerelvényeket stb., s ebből kellene kitalálni, milyen az épület, lakóház, óvoda, iroda, könyvtár, étterem vagy más.

Hogyan lehetséges a végtelen számú kapcsolatot, fogalmi asszociációt – a jövőbelieket is – belecsempészni az indextételbe? Ugye, milyen meghökkentő követelmény! Pedig csak olyan egyszerűen kell gondolkodni, mint *Noam Chomsky*, aki egyszer nagyon eltöprengett azon, vajon miként lehetséges az a csoda, hogy az em-

beri nyelv lexikailag véges számú szótári egységből áll, a grammatikája véges számú szabályból, mégis végtelen számú gondolat fejezhető ki vele! Így született a generatív grammatika. A megoldás tehát Kolumbusz tojása, csak figyelni kell természetes nyelvünkre, pl. hogyan teszi lehetővé végtelen fajtájú kapcsolat kifejezését néhány nyelvtani esettel. A megoldás: egyfajta szintaxis bevezetése az indexelési szabályokba.

Biztos tehát, hogy ami a „mondattan” feladata, azt ne vigyük a szótárba. Ha ezt tesszük, a tárgyszórendszer alig lesz alkalmas tárgyszavazásra. De önmagában állóan egyre tökéletesebb lesz egy nemlétező valóság üresedő leképezésére.

Visszatérve, melyek azok a szótári funkciók, vagy másként, mi várható magától a tárgyszórendszertől? Mindössze két dolog.

Az első, hogy a jelentések egyértelműek legyenek, de nagyon egyértelműek. Szinonímia, homonímia területén nincs pardon (ezek is relációk), itt a legcsekélyebb lazaság sem tűrhető el. A második: a konstrukció segítsen rálelni a legmegfelelőbb tárgyszóra vagy tárgyszavakra. Ha lennének további relációk, akkor ezért kellenének és kizárólag ezért. A szótári relációknak nincs jelentősége az indextételben – bár ezt csak sarkítva, az érthetőség kedvéért lehet kijelenteni. Ám a legmegfelelőbb tárgyszó kiválasztását más is segíti: szakrend, szakcsoportos részek, hierarchikus részek, esetünkben klaszterek, amelyek térképszerűen tárják elének a tárgyszavak nagyobb, összefüggő csoportjait.

Hasonlóan kell gondolkodnunk a facettákról vagy fazettákról. A facetták egyfajta kezdetleges szintaxis kifejezésére valók. De szótárilag mondani meg előre, hogy egy tárgyszó milyen facettába kerülhet, ugyanolyan, mint bármilyen előre szóló, eleve, mondhatni előltételes besorolás. A szótárban nem kérnek kenyeret, nem ártnak, de nincs értelmük. Az általános tárgyszójegyzékben különösen nincs értelmük. A facetták is az indextételben élnek igazán.

---

### A szakcsoportokról

---

#### VI.

A szakcsoportos/szakrendi rész valójában sem nem szakcsoportos, sem nem szakrendi; matematikai függőségben lévő tárgyszavak klaszterei-

ből áll majd. A *klaszter* terminus technicust még kevesen értik, különösen hivatalos előterjesztések bírálói, ezért az előterjesztésben szerencsésebbnek tűnt meghagyni a hagyományos szakcsoport és szakrend elnevezést, amely csupán egy feladatot jelez. A klaszterekből álló résznek ugyanis valóban azokéhoz hasonló szerepe van, csak a funkciót jobban tudja teljesíteni. Ez a szerep az, hogy a nagyobb lexikai egységek áttekintését nagyobb összefüggésben teszi lehetővé, így támogatja a legjobb tárgyszavak kiválasztását. Emellett egy-egy klaszter képes feltárni a kutatási/fejlesztési témák mélyszerkezetét, képes követni is a változásokat, méghozzá interdiszciplináris vetületben.

A klaszteranalízis arra képes, hogy objektumok, elemek stb. csoportosítását végezze el valamennyi tulajdonságuk alapján. A logikában az osztályalkotás, majd az osztálybesorolás preferált tartalmi ismérvek szerint lehetséges. Az ismérv(ek) preferálása az összes többi tulajdonság figyelmen kívül hagyásával jár. Márpedig minden dolognak kimeríthetetlen sok tulajdonsága van. A klasszikus logikai eljárások nagy korlátainak egyikéről van szó; ugyanis hiába hangsúlyozzák, hogy „lényeges” ismérv legyen az osztályképzés alapja, hiszen nagyfokú szubjektivitás és önkény volt abban, hogy ki mit tart lényegesnek. Az osztályalkotás (és besorolás) inverz műveleténél, a felosztásánál, a felosztási alapok megválasztásában ugyanígy jelent meg az önkény. Egy példa talán beszédesebb. Ha megalkottuk – mondjuk az órák osztályát, akkor ebbe egy konkrét órát be lehet sorolni annak az egyetlen tulajdonságának alapján, hogy méri az időt, azaz rendelkezik az „óraság” osztályalkotó ismérv tulajdonságával. De el kellett tekintenünk más jellemzőktől, pl. hogy iparművészeti remek, hogy műszaki újdonságokat hordoz, vagy – hogy szélsőig élezzem – az a bizonyos óra a Nemzetinél, ahol a hatos megáll, éppenséggel nem időmérési alkalmassága miatt rögzült a pesti szerelmesek tudatában.

A klaszteranalízis valamennyi tulajdonságot figyelembe vesz a klaszterek formálásában. Ezért mondják, hogy a klaszterálás „objektív”. Ennek ellenére rendkívül rugalmas, ahogy alább látható lesz.

Ismeretes, hogy a modellalkotásban egy könyvtár absztrakt modellje egy dokumentum-ismérv mátrix, amelynek celláit numerikusan töltik ki, a sorvektorként megjelenő dokumentumok és az oszlopvektorként értelmezett tárgyszavak köl-

csönös leképezése szerint. Binárisan kitöltve 0 és 1 kerül a cellákba. Egy oszlopvektor mutatja, hogy hányszor osztották ki a tárgyszót az indexelés során és melyik dokumentumok kapták meg. A sorvektor egy dokumentum képe, mutatván az összes tárgyszót, amelyet a dokumentum kapott. A mátrixból minden tárgyszóra leolvasható, hogy milyen, mekkora az „asszociativitása”, melyik más tárgyszavakkal társul gyakrabban, melyikkel ritkábban, melyekkel soha, melyekkel majdnem mindig, s.i.t. Ebből származtatható egy ismérv-ismérv mátrix, az együttes előfordulás számát, eseteit tartalmazva. Lesznek szorosan összekapcsolódó tárgyszavak, lesznek egymástól idegenek. Ezt a szorosságot, idegenséget egy skálán (egy intervallumban) lehet mérni. A klasztereknek kiszámítására az ún. hasonlósági függvények szolgálnak, amelyek oszlop-, illetve sorvektorok között mutatják a hasonlóság mértékét, vagy a már kész (vagy előre definiált) klaszterek „közepét” reprezentáló centroid vektor és a sor- oszlopvektorok között. Tucatnyi hasonlósági függvény ismeretes. A hasonlóság kifejezésére bevezethetők küszöbértékek, amelyek „elvágnak” a kapcsolatokat, úgy hogy a küszöbértéket meghaladó szorosságot mutató tárgyszavak alkotnak egy klasztert.

Klaszterbe nemcsak az ugyanazon tudományághoz tartozó tárgyszavak kerülnek, hanem amelyek valóban vonzzák egymást, így a klaszter mindig tudományközi, ahogyan az élő kutatási témák is interdiszciplinárisak.

A küszöbértékek magasabban-alacsonyabban szabhatók meg. Előző esetben több klaszter keletkezik, klaszteren belül magas kohézióval és kisebb elemszámmal (tárgyszóval), míg alacsony küszöbérték népesebb klasztert produkál alacsonyabb kohézióval. A küszöbértékek változtatásával mindkét klaszter típus megalkotható.

A klaszterek megformálásának feltétele azonban az – láttuk előbb –, hogy három-négy évig indexeljenek, utána néhány könyvtár indexelt tételeit bekérve alkothatók meg a klaszterek. Ezek „objektív” szakcsoportokat, illetve – a hierarchikus klaszterek választása következtében „szakrendet” nyújtanak. Addig azonban, amíg erre lehetőség nyílik, átmeneti szakcsoportokkal, mintegy munkaeszközként bevezetett szakrendfélével kell dolgozni. Ezt nagyjából a BSO alapján kialakított, csak generikus fogalmakat tartalmazó szerkezet biztosítja. Mivel a szógyűjtés során a tárgyszavak ETO ekvivalensét is igyek-

szünk feltüntetni – sok esetben nem lehet –, más eljárás is elképzelhető. Valószínű, hogy ellenőrzési célokból egyik-másik munkaszakaszban ETO szerinti rendezésre (csupán átfogó csoportokba!) is szükség lesz.

Fentebb talányos volt az a mondat, hogy a klaszterek tükrözik a kutatási/fejlesztési témák mélyszerkezetét. Talán most már világosabb, hogy miért és miként. Az indexelés gyakorlatában az élő tudomány közleményei kerülnek köllégáink asztalára. Ha egy dokumentumhoz hozzárendeli a tárgyszavak egy csoportját, akkor egy élő tudományos eredmény alapján e tárgyszavak közötti egymáshoz rendelést is elvégzi. Tehát, nemcsak a dokumentumokhoz osztja ki az őket megilletőket, hanem rögzít egy tudományos tény, az egymás közötti asszociációt is megállapítja spontán módon. Igaz, egyszerre csak egy szerző által létesített fogalomkapcsolatok alapján; de a módszer statisztikus, ne feledjük. Az asszociáció pedig azért érvényes, mert maga az élő tudomány így produkálta az eredeti közleményekben. A fogalmak együttes előfordulásának gyakorisága tükrözi a tudomány mélyebb szerkezetét. Nagyjából úgy, ahogyan *Thomas Kuhn* modellezi a tudományt,<sup>3</sup> illetve, ahogyan *Kunszt György* javasolta feltérképezni a hazai kutatás szerkezetét, méltatlanul elfelejtett nagydoktori disszertációjában.<sup>4</sup>

Ha az asszociativitás fogalmi hálói ily módon, mintegy a könyvtári indexelő tevékenység melléktermékeként rendelkezésre állnak, akkor különböző elemzések sokasága is lehetővé válik. Hasonlóan *Garfield* indexművének sorsához. Amikor ötletével előállt, s kiadta az első science citation-eket, még az volt a szándéka, hogy a „hólabda” módszer visszakeresési stratégiájának találja meg automatizálható algoritmusát. Ezt is megtalálta, de inkább tudománytani vizsgálatok kimeríthetetlen kincsesbányáját állította elő.

Mivel a klaszteranalízis a hazai könyvtártechnológiában csak elméletben ismert – így se nagyon –, az igazi nagy munkát előkészítendő, *Csabay Károly* megkísérel tárgyszavak közötti távolságot, illetve közelséget meghatározni. A „távolság” és „közelség” a metrikus tér fogalmait, egyben a klaszteranalízis alapfogalmait. Kísérletét már indexelt néhány ezer műszaki szakirodalmi tétel segítségével végzi. (A *Csabay*-kíséret konkrét célja valójában egy indextétel és használói kérést reprezentáló keresőprofil közti távolság

definiálása. Az eredmény azonban sokrétűen használható.)

Ezt követően az első „éles”, szótárépítési célú klaszterálást a már évek óta a gyakorlatba bevezetett pedagógiai tárgyszójegyzékkel fogjuk végezni. Ehhez több tízezer indexelt tétel áll rendelkezésre.\*

Mielőtt bárki elrettenne, megnyugtatóan közölni kell azt is, hogy a klasztereket tartalmazó rész fizikailag kisebb-nagyobb tárgyszócsoporthoz áll majd, egyszerű és áttekinthető lesz. Használatához nem kell ismerni az előállítás elveit és technikáját.

---

### Alkalmazási szabályok

---

#### VII.

Alkalmazási, indexelési szabályok nélkül egy tárgyszókészlet olyan, mint egy étterem lenne akkor, ha mindenféle nyersanyagot megszerezne az ingyen fogásokhoz, csak éppen a konyha és a szakács hiányozna az ételek elkészítéséhez; jobb híján az ételek nyersanyagait vinnék ki a vendégnek, mondván, rágja össze az ízeket maga. Szükséges tehát olyan eljárás, amely nemcsak alakilag szabályozza, hanem biztosítja a tárgyszavaknak indextétellel szintetizálását. Nem csupán analógiaként lehet a természetes nyelvre hivatkozni, mert ott a szavakat a grammatikának egyik fontos része, a szintaxis generálja mondatokká. A közlés pedig mindig is mondatokban testesült meg.

Az alkalmazási szabályokat három – esetleg négy – egymásra épülő, egymást kibontó szinten lehet megadni. Egy igen egyszerű, akár önképzéssel is elsajátítható szinttől egy magas szintaxist felmutató nívoig kell kidolgozni. Utóbbi-ként igényes alkalmazásra a PRECIS-t javasoljuk. Mivel a PRECIS ismertetése elolvasható<sup>5</sup>, ezúttal csak az egyszerűbbek hozzáférhetően, inkább koncepciószerű elképzeléseit érdemes taglalni. Annyit kell még előre bocsátani, hogy az alkalmazás szintjei függetlenek a tárgyszójegyzék há-

---

\* Az eredményekről az első közlemény rövidesen várható a Tudományos és Műszaki Tájékoztatásban. Az előmunkálatok tapasztalatai alapján lehet az általános tárgyszójegyzék, valóban nagyléptékű munkáit elkezdni.

rom szintjétől, bármelyik részletezésű tárgyszórendszer dolgozhat kimunkáltabb vagy elnagyoltabb alkalmazási szabályokkal.

Valamennyi alkalmazási szinten megjelenő indextétel a visszakeresés oldaláról kétféle tárgyszóból áll: akarjuk-e a tárgyszót elérési ismérvként is szerepeltetni, vagy sem. Utóbbi esetben egy tárgyszónak csupán az indextétel pontos jelentésében van feladata, az üzenet hitelességét elősegítő értelmi funkciója van, de maga a tárgyszó nem alkalmas arra, hogy keresését „vezéreljen”. Pl. a kezdő orvosok fizetése tételben a „kezdő” fontos jelentésű minősítő, de senki nem keres „kezdő” alatt. Más szóval, az indextételben megjelennek a tárgyszavak, amelyekre a tételt nem kell invertálni, ide nem kell besorolni. Az indextételben jelölni kell, hogy a tárgyszó egyben elérési ismérv is.

Minden szakirodalmi közleményben megragadható valamilyen cselekvés, folyamat, történés, változás vagy ennek felfogható jellegzetesség. Pl. gyártás/, háború/, javítás/, kísérlet/, lepárlás/, növekedés/, szervezés/, verés/ stb. Legtöbbször megtalálható a cselekvés tárgya, akire vagy amire a cselekvés irányul, aki/ami elszenved. Autót lehet gyártani, árakat növelni, gyereket verni, konyakot lepárolni stb. A cselekvés mellett megjelenik tehát egy „tárgy”. Elemezhető továbbá cselekvő, végrehajtó személy vagy közeg. A Ford Művek gyárt, az ár növekszik, a tanító ver stb. A végrehajtó, cselekvő vagy ágens (a nyelvtanban agentivus) olyasminek tekinthető, mint a nyelvtani alany. Negyedikként kapunk egy olyan csoportot, amelyet legszerencsésebb – *Dahlberg* nyomán – komplementumoknak nevezni;<sup>6</sup> ezek a tárgyszavak az előző három komponenst egészítik ki. Hely, idő említhető komplementumként.

Egyszerű sorrendiség előírásával mondatféle hozható létre. Javasolni lehet – sokan megfontolták – a tárgy-cselekvés-cselekvő-komplementum sorrendet. Ha pl. a sorrend a mérnök-képzés, illetve a képzés-mérnök, egyértelművé válik, hogy a téma a mérnökképzés vagy a mérnök részvétele az oktatásban.

Ha a témát egyetlen tárgyszó írja le, akkor „tárgy” funkcióban áll. Pl. a „Franciaország-útikönyv” tételben Franciaország a tárgy, az útikönyv a komplementum. A „tárgy” tehát olyasmi is, ami *Ranganathannál* a „personality”.

A sorrendet, a sorrenden belüli helyet, pozíciót valamilyen egyezményes jel vagy a funkciót ki-

fejező szimbólum mutatja. Szükség van erre, mert nem minden pozíciót kell minden tételnél kitölteni, másfelől ugyanabban a pozícióban több tárgyszó állhat. Pl. a „Dánia-Finnország-Norvégia-útikönyv” tételben mindhárom országnév tárgy. Hasonlóképpen ismétlődhet a cselekvés, végrehajtó, komplementum.

További szabályok vonatkoznak arra, hogy mi legyen az olyan tárgyszavakkal, amelyeknek értelmi funkciójuk van, de nem lényegesek elérési ismérvként. Mi legyen az összetett kifejezésekkel stb. A részletek sorolhatók, felírók azonban ennyi is elegendő.

Ha a fentiek alapján valaki néhány tételt indexelni kezd, azonnal észreveszi, hogy e szabályok éppen csak összetákolják az indextételt, a közlést magát csak főbb vonalaiban szintetizálják, hasonlattal élve, csak a főtisztek helyzetét szabtuk meg, holott a játszmában tisztek, gyalogok is részt vesznek. A finomabb problémákat a szőnyeg alá söpörtük a komplementum bevezetésével. A további szintek ezen javítanak.

A közbülső szint vagy szintek újabb szintaktikai funkciók jelölését teszik lehetővé. Nyelvtani analógiával élve nemcsak alany, igei állítmány és tárgy szerepelhet a mondatban, hanem más bővítmény is. Legalább további három szintaktikai feladat jelölésére van szükség: „minősítő”, „része” és „eszköz”.

A minősítő szakaszoló szerepet kap. Nem csupán diákokról lehet szó, hanem középiskolás diákokról, erdélyi diákokról, nagyothalló diákokról, nem csak autóról, hanem használt autóról, nem csak futásról, hanem hosszútávú futásról, nem csak tündérről, hanem gonosz tündérről stb. A minősítőként használható tárgyszavak szintén fellelhetők a szótárban, de szabály írja elő összekapcsolhatóságukat, továbbá azt is, hogy szerepeljenek akkor, ha nem akarjuk elérési ismérvként is feltüntetni.

A „része” funkció méltán kapott idézőjelet. Lehetne mondani, hogy birtoklás, birtokos szerkezet kifejezésére való; de hát a kérdés bonyolultabb, mert logilailag a fenti példa tündérének a gonoszsága is birtoka. Inkább néhány példát. Erdélyi diákok/nak az/ egyetemjárása, a Duna mellékfolyói, orvosnak a magatartása (lám orvosi magatartás), mozdonynak a kazánja, karmesterek versenye, valamint a kezdete, fénykora, a szakszervezetek pénze. Valódi rész is, mint a mozdony kazánja, a cserebogár potroha, a csarnok tetőszerkezete.

A harmadik, az „eszköz” (instrumentalis) azt az eljárást, feltételt, dolgot, szerszámot jelöli, amelynek (akinek) segítségével vagy feltételeivel a cselekvést végrehajtják vagy a történés végbe megy. Bottal ütök a lábnyomot, hipnózissal gyógytanak, meleggel erjesztenek stb.

Ezen a szinten természetes a tér és idő jelölése is.

Hol helyezkednek el ezek az első szint komplementumából kiemelt tárgyszavak? Annak a tárgy, cselekvés, ágens funkcióban álló tárgyszónak a társaságában, amelyikre vonatkoznak. Ha pl. a minősítő tárgyszó a folyamatot pontosítja, akkor vele együtt lép föl, ha a tárgyat, akkor a tárgy társaságában van, ha a végrehajtót, akkor az ágenssel együtt szerepel.

Most már megfogalmazható az első és a második szint lényegi különbsége. Az első szinten az indextétel tárgy, cselekvés, ágens pozíciójú tárgyszavakból áll, míg a másodikon tárgy, cselekvés, ágens pozíciójú *tárgyszó blokkokból*. A komplementum is e szerint módosul.

Legyen a magyarázó példa egy ismert idézet. „Náadorispán látja Toldit a nagy fával.” Minden diák tudja, hogy a „nagy fa” Toldi kezében van. Pedig ez csak a kontextusból következik, a mondat szintaktikai szerkezetéből nem. A „nagy fa” lehetne a náadorispáné is, ahogyan egy ugyanolyan szerkezetű mondatból látható: Náadorispán látja Toldit a távcsővel. Itt teljesen bizonytalan, hogy Toldié-e a távcső vagy a náadorispán használja. Hajlamosak vagyunk a távcsövet a náadorispán kezében látni. A hasonló bizonytalanságok elkerüléséért az összetartozó tárgyszavakat közös blokkba kell foglalni; a blokk egészének funkció-

ját továbbra is az első szintnél tárgyalt szerkezet mutatja. A jelölés oldaláról pedig meg kell különböztetni a blokkok határát és a blokkon belüli szerkezeti egységek elválasztó jeleit.

A különböző jelöléseket az alábbiakban lehet összefoglalni. Fel kell tüntetni, hogy a tárgyszó elérési ismérv, a tárgyszó szintaktikai szerepét, a blokkok határait, végül a blokkon belül a tárgyszavak elválasztó jeleit. Az utóbbi két jelölés az egyszerűbb, alsó szinten nem él.

Az indexelési szabályok legmagasabb szintjén egy teljességében kibontott szintaxis áll, amelyet a PRECIS szabályai alkotnak. Az olvasót ennek megismerésére a már hivatkozott tanulmányhoz utaljuk.

## Irodalom

1. JASPERS, K.: Bevezetés a filozófiába. Bp. Európa, 1987. 87.p.
2. LANDRY, B.C.: A theory of indexing. Indexing theory as a model for information storage and retrieval. Ph. D. dissertation. Ohio State University, Columbus (Ohio), 1971.
3. KUHN, Th. S.: A tudományos forradalmak szerkezete. Bp. Gondolat, 1984.
4. KUNSZT Gy.: A tudományos kutatás logikai modellezése és tematikai irányítása. Általános elmélet és építéstudományi alkalmazások. Doktori értekezés. Bp. 1973. Sokszt.
5. HORVÁTH T. - ORBÁN É.: PRECIS. = Tudományos és Műszaki Tájékoztatás. 1987. 7.sz. 311-323.p.
6. DAHLBERG, I.: Ontical structures and universal classification. Bangalore, Sarade Ranganathan Endowment, 1978.

---

AZ ELEKTRONIKUS DOKUMENTUMOK SZERZŐI JOGAINAK VÉDELME még nem kielégítő. A digitális formában tárolt és továbbított dokumentumokat különösen könnyű illegálisan felhasználni, ezért a szerzői jogok tulajdonosai nem szívesen engedélyezik a dokumentumok ilyen formába való áttételét.

Mint egy világméretű felmérés kimutatta, e dokumentumok szerzői jogvédelmére sehol sem találtak megnyugtató megoldást, az esetleges szabványok is voltaképpen csak vállalatközi egyezmények. - A felmérés negatív tanulságai alapján dolgozta ki a Bull cég franciaországi vállalata a CITED (Copyright in Transmitted Electronic Documents) modellt. Alkalmazásának lehetőségeit egyelőre két kísérleti rendszerben próbálják ki: egyrészt a digitálisan (az ADONIS rendszerben) tárolt vizuális dokumentumokra, másrészt a digitális hangfelvételekre vonatkoztatva.

(XIII Magazine: news review, 1992. 2.no.)

# Academic Bookshop

1052 Budapest Gerlóczy u. 7  
1364 Budapest, Pf. 52  
Nyitvatartás: H-P: 9-17<sup>h</sup>  
MNB számlaszám: 216 01283

Tisztelt Ügyfelünk!

Ezúton tájékoztatjuk arról, hogy boltunknak számos, tudományos körben elismert kiadóval van kapcsolata, mint például:

*Oxford University Press, Baurverlag, Penguin Books Ltd., Schwer Verlag, Harcourt Brace Jovanovich Ltd., Kluwer Academic Publisher Group, McGraw-Hill Book Company (U.K.) Ltd., Chapman & Hall Ltd., Carl Hanser Verlag, K.G. Saur Verlag GmbH & Co. KG., Walter de Gruyter Berlin-New York., DLV., Urban & Schwarzenberg, Vieweg, Jungjohann, Eugen Ulmer Verlag, Bouvier, Encyclopedia Britannica, Wm. Collins Sons & Co. Ltd., BBC English.*

Ügyfeleink igényének megfelelően természetesen tovább bővítjük a kiadók számát.

Boltunk a kiadványokat katalógus áron kínálja. Egyéni és nagybani megrendelések kielégítésével is foglalkozunk. Ha a fent említett kiadók katalógusaiból rendel, úgy 4-6 hét alatt be tudjuk szerezni a kiadványt.

Amennyiben új kiadóval kell fölvennünk a kapcsolatot, vagy a tengerentúrról kell behozatnunk a kiadványt, az hosszabb időt vesz igénybe, de tájékozódás után értesítjük ügyfelünket, s akkor módosíthatja megrendelését.

Extra kívánságokat is teljesítünk, ha ügyfelünk vállalja a többletköltségeket.

Postai utánvétellel is értékesítünk könyveket.

Bordás Gitta  
könyvesbolt vezető

**„Minket minden ügyfél érdeklődése kötelez.”**