

Ungváry Rudolf

A névterek és az adatok tulajdonságai

A történet, avagy a fogalmi és a nyelvi szint

A nevek (tulajdon- és köznevek) névtereket alkothatnak, melyek többek között gyűjteményi rendszerekben, vagy a szemantikus web céljaira használhatók. Megjelenítésükre és kezelésükre már vannak jelentős példák. Az egyes szakterületeken jelenleg kisebb vagy nagyobb mértékben egymástól eltérő szabványok és adatcsere-formátumok szabályozzák a nevek formáit, relációit és logikai szerkezetét. A névterek tulajdonságai az adatok tulajdonságaiból vezethetők le, az adatok tulajdonságai pedig a tipológiájukból adódnak. A tanulmány tárgyalja a névterek történetét, és az adatok terminológiáját.

Kezdetben voltak a nevek. Ezek tulajdonnevek és köznevek (általános nevek) voltak. A tulajdonnevek a térben és időben konkrét dolgok (instanciák) nevei voltak, a köznevek a tudatban létező fogalmak nevei. Mindezek a nevek a tudatban léteztek.

A tudatban létező neveket ki is mondták, ezek – feltehetően – a beszéddel egy időben keletkeztek és vele változtak (módosultak, kiegészültek, fejlődtek) a nyelvben és a tudatban. A nevek a tudatban kezdettől fogva egymással szoros tartalmi kapcsolatban álltak, azaz strukturáltak voltak és rendszert alkottak. E tudatban létező rendszer (egyelőre talán) tudományos értelemben még nem igazán ismert.

A neveket többnyire más nevekkel összekapcsolva mondták ki, mindig az adott helyzet (eset) függvényében. Ami neveket egymással az esetek függvényében kimondtak, azok a tudatban gondolatok voltak, a beszédben mondatok. A gondolatoknak tehát a tulajdonnevek és a fogalmak, a mondatnak pedig a nevek a részei.

A mondatokban összekapcsolt nevek kapcsolatainak (az eseteknek) is fogalmak az alapjai, de ezek a mondatok felszínén csak jelzésszerűen (ragok, képzők, sorrend, hangsúly) jelennek meg, ugyancsak a mondatok részeként. A fogalmakat, melyek eme jelzésszerű elemek alapjai, a legutóbbi időben mélyeseteknek nevezték el. Egyelőre nincs igazán megegyezés abban, melyek ezek a fogalmak.¹ Eldöntetlen, hogy a nevek és fogalmak létezése megelőzi-e a mélyeseteket képviselő fogalmak megjelenését vagy sem. Ahhoz mindenesetre, hogy a nevek és fogalmak, továbbá a fogalmak és a fogalmak közötti, szemantikainak hívott kap-

csolatok tárgyalhatók legyenek, lényegében nincs szükség se a mélyesetekre, se a velük összefüggő mondattani (szintaktikai) kérdésekre figyelembe vételére.

A természetes módon kialakult, a tudat által „kezelt” nevek nem egyértelműek. Egy-egy dolognak (entitásnak)², illetve egy-egy fogalomnak³ sokszor egynél több neve is lehet, egy-egy név sokszor egynél több konkrét dolgot, illetve fogalmat nevez meg.⁴ A természetes (beszélt) nyelven kifejezett nevek és dolgok, illetve a nevek és fogalmak közötti kapcsolatok számszerű jellemzője (kardinalitása) $M : N$.

Idővel a neveket le is írták, azaz a tudattól független jelhordozókon rögzítették. Ezáltal az elfelejtett nevek is megőrződtek, és a nevek áthagyományozódása biztonságosabbá vált. A rögzített neveket később adatoknak nevezték, közöttük kapcsolatokat is felismertek, és az utóbbiakat relációknak nevezték⁵. Kialakult a nevek különféle rögzített struktúrája, melyeket igyekeztek rendszerré szervezni, és gyakorlati célokra (pl. tárolt információk keresésére) felhasználni. Ehhez azonban a nevek őserdejében rendet kellett teremteni, hiszen – mint mondtuk volt – a beszélt nyelv nevei nem egyértelműek: Vannak nevek, amelyek egyszerre több dolgot, illetve fogalmat neveznek meg (ezek a nevek a homonimák), és vannak dolgok, illetve fogalmak, melyeknek egyszerre több neve is létezik (ezek a nevek a szinonimák).

Idővel a neveket szabványok, szabályok, megállapodások alapján egységesítették. Ez azt jelentette, hogy az egyes rendszerekben minden dolognak és minden fogalomnak egységes szabályok szerint

kialakított neve keletkezett. A név csak egyetlen dolgot, illetve csak egyetlen fogalmat azonosíthatott, és egy dolognak, ill. egy fogalomnak csak ez az egyetlen neve lehetett. Ha az így egységesített neveket kezdetben a dokumentumok tartalmának, témájának, tárgyának a kereshető tárolásához is használták, akkor tárgyszavaknak kezdték nevezni őket.

Idővel pontosabban felismerték az összefüggést a név és a dolgok, illetve a fogalmak között, és megalkották a deszkriptor (descriptor) és a nemdeszkriptor (nondescriptor) fogalmát. Egy dolognak, illetve fogalomnak csak egy deszkriptora lehetett, és egy deszkriptor csak egy dolgot, illetve fogalmat nevezhetett meg. Azaz a deszkriptor és a dolog, illetve fogalom között a reláció számszerűsége (a továbbiakban kardinalitása) 1 : 1. A deszkriptorhoz kapcsolódnak a dolog, illetve a fogalom összes jellemzői. A nemdeszkriptorok a szinonimákat vagy a homonimákat képviselték, és helyettük deszkriptort, vagy deszkriptorokat kellett használni.

Idővel azt is felismerték, hogy adott dolognak, illetve fogalomnak a különböző felhasználók számára más és más név volt fontos, ezért több nevet kellett deszkriptorként egységesíteni. De hogy az adott dolgot, illetve fogalmat egyértelműen (1 : 1 kardinalitással) azonosítani lehessen, erre a célra a deszkriptorok közül ki kellett választani egy kitüntetett nevet (preferred term). A többi deszkriptor csak ezen a kitüntetett néven keresztül kapcsolódott a dologhoz, illetve fogalomhoz, de a kitüntetett névvel egyenrangúan, azzal párhuzamosan volt használható például tárgyszóként (e deszkriptorokat ezért nevezik párhuzamos neveknek [parallel form of name] is). Valójában a kitüntetett név is deszkriptor, csak éppen olyan, amelynek a feladata, hogy a dolgot, illetve a fogalmat 1 : 1 kardinalitású relációban azonosítsa, a (többi) deszkriptor viszont ezt csak a kitüntetett néven keresztül azonosította. Ugyancsak a kitüntetett néven keresztül kapcsolódhattak a deszkriptorok a kitüntetett név összes további jellemzőjéhez, többek között a szemantikai meg egyéb relációihoz is. A kitüntetett név és a deszkriptorok közötti reláció kardinalitása 1 : N.

A kitüntetett név és a deszkriptorok összessége az egységesített nevek összességét alkotják (authorized form of name). Ami ezen felül maradt, a nem egységesített nevek, azok a nemdeszkriptorok (nondescriptor)⁶. Ezek az adott fogalom egyéb, természetes nyelvű, az adott rendszerben kevésbé

elfogadottnak, mellékesnek, hibásnak tekintett nevei. Belépő szótárt (entry vocabulary) alkot(hat)nak.⁷ A nemdeszkriptorok és a kitüntetett név között a reláció kardinalitása M : N. A nemdeszkriptorok mindig egy vagy több egységesített névre utalnak.

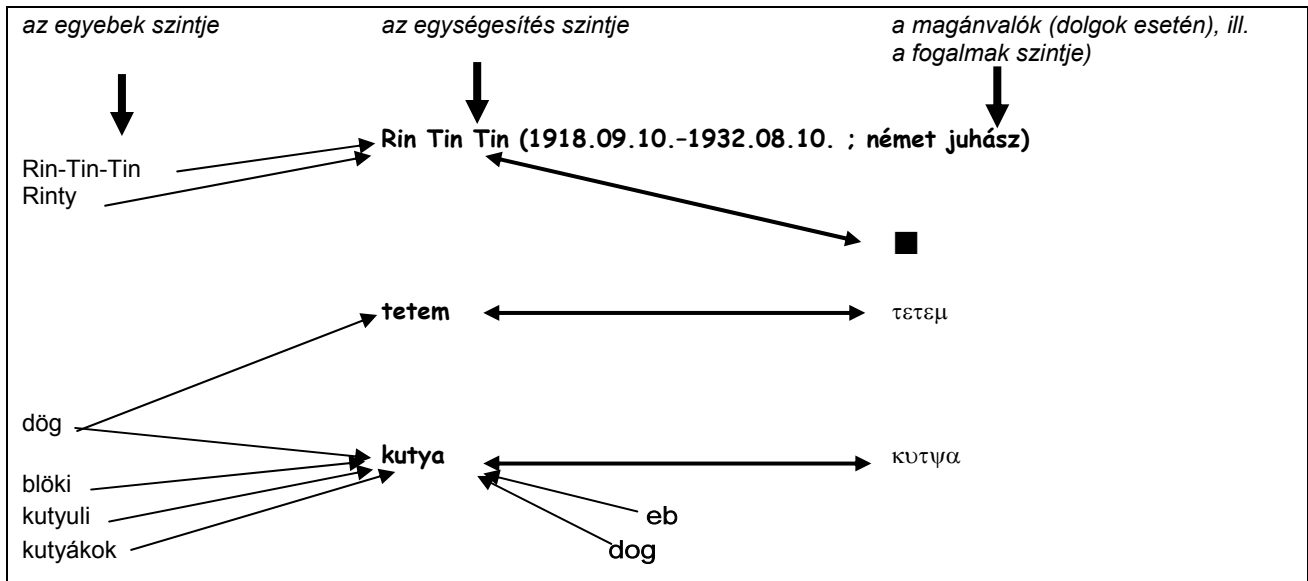
Kialakult tehát a névterek formálisabb rendje. Ez ugyan bizonyára nem azonos azzal a renddel, amely a tudat neveinek és fogalmainak világát jellemzi; az is lehet, hogy a tudatban formális értelemben egyáltalán nincsen rend. A tudatban uralkodó „rend” részben az örökléssel, részben pedig a neveléssel és egyéb kulturális hatásokkal hagyományozódik át. A formális rend olyan, hogy le lehet írni, meg lehet érteni, közvetlenül a cselekvéssel át lehet adni, és alkalmazni lehet mesterségesen létrehozott dolgokra.

A fenti folyamat eredményeként megvalósult a fogalmi és a nyelvi szint egyértelmű szétválasztása, a nyelvi szinten belül pedig megvalósult az elismert és a kevésbé elismert – az egyéb – szint megkülönböztetése. Ez azért bizonyult sikeresnek, mert a legrugalmasabbak azok a rendező rendszerek, amelyekben van „egyéb” kategória. Ily módon mindennek helye lehet a nap alatt. Az „egyéb” bevezetésével a rendszer formális volta ugyan gyengült, de a tudatihoz való hasonlósága egészen kis mértékben talán megerősödött.

Az egységesített nevek és a nemdeszkriptorok összessége a besorolási nevek vagy adatok állományát alkotják, azaz segítségükkel rendszerbe lehet szervezni a dolgok, illetve fogalmak neveit (és nagyon kis mértékben, áttételesen a dolgokat és fogalmakat magukat is, de itt célszerű, ha nincsenek illúzióink). Mindezzel megszületett a dolgok és fogalmak neveinek mesterséges világa, a névtér, most már csak a továbbfejlődéséről lehet szó.

Az 1. ábrán kitüntetett nevek, deszkriptorok és nemdeszkriptorok példáit mutatjuk meg.

A névterekben tehát vannak nevek, melyek konkrét dolgokat vagy fogalmakat azonosítanak (pl. Rin Tin Tin [1918.09.10.–1932.08.10. ; német juhász], kutya), és vannak önmagukban álló nevek, melyek a dolgokhoz, illetve fogalmakhoz meghatározott relációkban kapcsolódnak (pl. Rin-Tin-Tin, eb, dög). A fogalmak maguk csak a tudatban léteznek, ezeket nem lehet „rögzíteni” (csak a nevüket), nem lehet közvetlenül sem átadni valaki másnak. Ezért is vannak, akik vitatják, hogy egyenrangú létezők a dolgokkal. A dolgok maguk pedig, ahogy az leg-



1. ábra Névterstrukturúra: a dolgok és $\phi\omicron\gamma\alpha\lambda\mu\alpha\kappa$, a fogalmak nevei, illetve konkrét dolgok nevei (kitüntetett nevek), a velük ekvivalens nevek (természetes, illetve mesterséges nyelven kifejezett deskriptorok), a nem egységesített nevek (nemdeskriptorok) és az azonosítók (lásd a 4. pontot)
 → = 1 : N kardinalitás; ↔ = 1 : 1 kardinalitás

alább *Kant* óta tudható, egyáltalán nem ismerhetők meg közvetlenül, csak a fogalmak, a gondolkodás, vagy alacsonyabb szinten az érzékek, az ösztönök közvetítésével. Ezért mondják, hogy magánvalók. Ha valaki netán egész életét Rin Tin Tin kutyával, a pénztárcájával vagy a feleségével töltötte, akkor sem fogja igazán átélni, érteni, és ismerni ezt a három valamit... Az ellenkezőjét persze hiheti. Ezen a problémán csak a fogalmak, és a belőlük összeállított gondolatok segítenek valamennyire.

Az önmagukban álló ekvivalens nevek a fogalmakhoz és egymáshoz ekvivalencia relációban, a nem egységesített nevek lásd típusú relációkban kapcsolódnak a fogalmakhoz, a fogalmak között pedig szemantikai relációk állnak fenn.

Az ekvivalencia és a lásd típusú relációk nyelvi relációk, a szemantikai relációk fogalmi relációk.

Ha egy (vagy több) név, illetve nem egységesített név egyetlen fogalomhoz kapcsolódik, akkor *szinonim* (vagy annak tekintett) *nevekről* (szinonimáról), ha egy nem egységesített név több fogalomhoz kapcsolódik, akkor *homonim nevekről* (homonimáról) beszélünk.

A nevek rendszerré szervezett, minden közvetlen gyakorlati alkalmazástól független rendszerei a névterek. A névtereket részei a tulajdonnevek és

köznevek, a közöttük fennálló szemantikai relációk és a kezelésükhöz szükséges egyéb adatok. A névterek a gyakorlatban adatbázisba szervezett és számítástechnikai eszközökkel kezelt strukturált szótárak (műszaki⁸ rendszerek).

Az adatbázisokban a névterek többnyire kiegészülnek formális azonosítókkal (és számos további elemmel, melyeket itt nem tárgyalunk). Elvileg a formális azonosítókra nem volna szükség, legfeljebb a nem egységesített nevek, a nemdeskriptorok esetén, mert maguk az egységesített nevek teljesen egyértelműen azonosítanak. Műszaki emberek feltehetően azért kedvelik mégis az alkalmazásukat, mert rövidek, és nem zavarják a programok készítőit azzal, hogy az egységesített neveknek mindig van független jelentése is. Akárhogy is legyen, a formális azonosítók és a nevek közötti reláció kardinalitása mindig 1 : 1.

A névterek mind a gyűjteményi, mind az ismeret-szervezési rendszerekben, mind pedig az ontológiákban hasznosíthatók.

Korábban, és különösen a 19. és 20. században megpróbálták a dolgokat osztályok rendszerébe foglalni. Azaz nem a fogalmak rendjéből indultak ki, hanem feltételezték, hogy a dolgok maguk szerveződnek osztályok rendjébe, és megpróbálták mintegy kitalálni ezt a dolgokban rejlő rendet.

Így születtek az osztályozási rendszerek. Ezek kialakítását természetesen befolyásolta a tudat által ismert fogalmak ismeretlen rendje, de jóformán alig, sokszor pedig nem is volt benne köszönet. Fogalmak dolgában ugyanis nagyon el lehet szállni. Ez a dolgokkal sokkal nehezebb, mert így vagy úgy, de az ember előbb-utóbb megbotlik bennük, vagy az ember fejére esnek, sőt olykik vissza is tud ütni.

A gyakorlatban nemcsak a dolgokat, hanem a fogalmakat is megerősítették ezekkel a rendszerekkel. E heroikus erőfeszítések eredményeként rendkívül figyelemre méltó rendező rendszerek születtek, melyek közül nem egyet világrendszeré válva ma is használnak. Ezen kívül pedig – főleg a közigazgatásban – számos nagyon speciális nomenklatúra született, melyek ugyancsak hasznosak a maguk területén, mivel akármilyen rossz is „genetikailag” egy rendező rendszer, ha már töviről-hegyire ismerik a használói, remekül beválhatnak. Előbb-utóbb ugyanis bármi bármire ráhúzható.

Az ebbeli törekvések eredményei a természetes nyelven alapuló taxonómiák, illetve a mesterséges nyelven alapuló, hierarchikus enumeratív (felsoroló) vagy egyéb elvek szerint szervezett osztályozási rendszerek és nomenklatúrák. A névterekhez ezek is kapcsolódhatnak. Az osztályozási rendszerek osztályai és a fogalmak rendszere között nincs pontos megfeleltetés, adott névtér fogalmi és adott, a névtérhez kapcsolt osztályozási rendszer osztályai (taxonjai) közötti reláció kardinalitása elvileg mindig $M : N$.⁹

A fenti, az osztályozási rendszerekre vonatkozó kétely azonban ne tévesszen meg senkit. Számos

osztályozási rendszer, taxonómia és nomenklatúra remekül bevált, jól használható, és főleg elterjedt rendező rendszer. Kár volna névtereket úgy létrehozni, hogy a fontos osztályozási rendszerek létezéséről a névterek készítői ne vegyenek tudomást. Noha azzal, hogy a névterekben sikerült valamennyire visszakecmeregni a gondolkodás ősföréséhez, készítőik is csak a sötétben tapogatóznak (egyelőre), ami az összhangot illeti a tudatban létező fogalmi világ és egy-egy névtér között.

Az osztályozási rendszerek jelzeteit végképp értelmetlen kiegészíteni formális azonosítókkal, de ami lehetséges, az akkor is megvalósul, ha nincs értelme.

A névterek nevei más nyelven kifejezett nevekkel is összekapcsolhatók. Az egyes nyelveken kialakított névterek nem pontos megfelelői egymásnak. Egy adott névtérben tehát az idegen nyelvű ekvivalens neveket megkülönböztetett minősítéssel kell összekapcsolni az adott névtérben érvényes nyelv neveivel. A kapcsolatok azonban mindig csak a kitüntetett nevek között állhatnak fenn, továbbá akkor, ha a kitüntetett névhez kapcsolódó deskriptor nem szinonimája a kitüntetett névnek, és ezért lehet más nyelvű ekvivalens neve. Adott esetben kapcsolódhat a más nyelvű név nemdeskriptorhoz is, ha az nem szinonimája a kitüntetett névnek. Lévén, hogy sok minden van, ami megfeleltethető mással, de az egyes nyelvek szinonimái és homonimái nem ezek.

Névtér és osztályozási rendszer összekapcsolását (bevonva a formális azonosítókat is) a 2. ábra szemlélteti.










<i>formális azonosítók (kódszavak)</i>	<i>Nevek</i>	<i>más nyelvű ekvivalensek</i>	<i>Osztályozási rendszer jelzetei (kódszavak)</i>
0000020	blöki		
0000301	dög	beast	
0005321	eb		
0051052	kutya	dog	599.742.13
0051056	kutyák		
0051126	kutyuli		
0001329	Rin Tin Tin (1918.09.10.-1932.08.10. ; német juhász)	Rin-Tin-Tin (1918.09.10.-1932.08.10. ; sheep-dog)	
0001330	Rin-Tin-Tin		
0001402	Rinty		
0901209	retem	cadaver	616-091.1

2. ábra Névtérstruktúra formális azonosítókkal, más nyelvű ekvivalensekkel és osztályozási rendszer (példánkban az ETO) jelzeteivel (kódszavaival)

A 3. és a 4. ábrán látható **kutya** kitüntetett név rekordjának egy lehetséges felhasználóbarát megjelenítését a Getty tezauruszainak grafikus felületét alapul¹⁰ véve a 3. ábra szemlélteti. A rekordban számos más, a névre vonatkozó adat is megjelenik (szöveges magyarázatok, források, továbbá kódszavakkal az, hogy az élő beszéd szava (É), bizalmas (B), szleng (S), használata szokásos, kur-

rens (K), főnév (F), adott esetben többes számú alak (T) stb.

Az **eb** deskriptor megjelenítése esetén minden olyan adat, amely a kutya esetén is megjelenik és a fogalom jellemzője, értelemszerűen szintén megjelenhet (azaz opcionálisan kérhető). Az **eb** deskriptor saját adatai természetesen csak az **eb** rekordjának megjelenítéséhez tartoznak.

AZ: 005321	
 kutya	
<i>Meghatározás:</i> Négylábú, a ragadozók nemébe tartozó háziállat	
<i>Névváltozatok:</i>	
kutya	(kitüntetett, É, K, F)
eb	(deskriptor, É, K, F)
blöki	(nemdeskriptor, É, B, F)
dög	(nemdeskriptor, É, K, F)
kutyák	(nemdeskriptor, É, S, F, T)
kutyuli	(nemdeskriptor, É, B, F)
<i>Hierarchikus helye:</i>	
 élőlény	
 állat	
 gerinces	
 ragadozó	
 kutyaféle	
 kutya	
 háziállat	
 kutya	
<i>általánosabban:</i>	háziállat ragadozó
<i>fajtája</i>	(élettanilag) agár juhászkutya puli uszkár (rendeltetés szerint) mentőkutya rendőrkutya vakvezető kutya
<i>átfogóbb része</i>	falka tépőfog
<i>rendeltetése</i>	házórzés
<i>alanya</i>	ebtenyésztés
<i>egyéb összefüggése</i>	ugatás veszettség
<i>terjedelmének eleme angolul</i>	Rin Tin Tin (1918.09.10.–1932.08.10. ; német juhász) dog
<i>ETO-jelzete</i>	599.742.13
Források:	
Élettani lexikon A kutya (időszaki kiadvány)	
Fölvéve: 2012. 01. 23.	
Állapota: elfogadott	

3. ábra Névtértétel (rekord) megjelenítése a Getty-tezauruszok megjelenítésének mintájára [3]

<p>AZ: 0051056</p> <p>☞ kutyák (nemdeszkriptor, É, S, F, T)</p> <p><i>Meghatározás:</i> A kutya említése városias szlengben, karikírozó célzattal.</p> <p><i>lásd</i> kutya</p> <p>Források: Bartos Tibor: Magyar szótár.</p> <p>Fölvéve: 2012. 01. 23.</p> <p>Állapota: elfogadott</p>
<p>AZ: 000030</p> <p>☞ dög (nemdeszkriptor, É, B, K, F)</p> <p><i>Meghatározás:</i> Több jelentésű szó. Elhullott állat teteme, vagy becsmerlő értelemben élőlény vagy ember, vagy ritkán becéző értelemben állat.</p> <p><i>lásd vagy</i> kutya tetem</p> <p>Források: Tóth Béla: Szájrul szájra</p> <p>Fölvéve: 2012. 01. 23.</p> <p>Állapota: elfogadott</p>

4. ábra Nemdeszkriptor-névtértétel (rekord) megjelenítése a Getty-tezauruszok megjelenítésének mintájára [3]

A fentiek értelemszerűen vonatkozhatnak a nemdeszkriptorok rekordjainak megjelenítésére is, noha ezek esetében alapértelmezésként csak a „lásd típusú” utalás jelenhet meg. A 4. ábrán két példa látható.

Ahogy a fogalmakat a tudat rögzíti (felfogja és „tárolja”), úgy a nevet hordozón rögzítik. Ez a tárolásuk módja. A hordozón rögzített jelek, különösen, ha nyilván is tartják őket, a tárolás és a kezelés szempontjából adatok. Következésképp a rögzített nevek is adatok.

A fogalmak és a köznevek természete kettős. A fogalmaknak tartalma és terjedelme van, a közneveknek jelentése és jelölete. Az előbbi azonos azzal, amit a fogalom tartalmában felfogunk, az utóbbi meg az a tárgy, amelyre a név vonatkozik, és ami a fogalom terjedelme.

A tulajdonnevek nem fogalmak, hanem konkrét dolgok neve, elvileg nincsen bennük semmi kettősség, hiszen egy tulajdonnév csupán megnevezése egy konkrét dolognak, afféle címke. De azért abból súlyos logikai probléma keletkezett, hogy a tulajdonnevek (a logikában: individuumnevek) nem fogalmakat neveznek meg, hanem konkrét dolgokat, és ezért elvileg nem lehet jelentésük. Hiszen a jelentés szemantikai–fogalmi jelenség, egy tárgy puszta megnevezése viszont nem annyira szemantikai, mint jelölési, technikai kérdésnek látszik.

Még egy állat is tud egy az egyben azonosítani, mert ehhez nincs szükség fogalmakra.

A névterekben ezt a problémát a gyakorlat áthidalja. A tulajdonnevekhez ugyanúgy lehet szemantikai relációkban akár más tulajdonnevet, akár bármilyen köznevet kapcsolni, mint a köznevekhez. Mintha csak fogalmak lennének. Nem véletlenül: a rögzített nevek adatok is, az adatokkal könnyebb manipulálni, mint akár a konkrét dolgokkal, akár a fogalmakkal. Ettől függetlenül is az a helyzet, hogy a neveknek is van némi „fogalmisága”: amikor azt mondjuk egy miniszterelnökre, hogy egy „Mussolini”, akkor a nevet fogalomként használjuk.

Ideje tehát, hogy sorra vegyük azokat az adattípusokat, amelyek a nevek vonatkozásában relevánsak. Ez elég fárasztó lesz, mert az adattípusok rendszere meglehetősen strukturált, noha első látásra ezt nem sokan gondolnák.

Terminológiai rész, avagy az adatok szintje

Az adatok az adatok szintjei szempontjából (lényegében ontológiai szempontból) lehetnek elsődleges (primer), másodlagos (meta-) és harmadlagos (tercier) adatok.

(1) Elsődleges az adat, ha konkrét dologra, illetve fogalomra vonatkozik (tehát nem adatra). Ez az adatok elsődleges szintje. (Az elsődleges adatokat nevezik primer adatoknak is.)

(1.1) A *nevek* a névterekben elsődleges adatok. A nevekhez számos további adat tartozhat, ezek összessége a nevekkel együtt összetettebb információs egységet (névtételt, névkortot) alkot.

(1.2) Elsődleges adatok továbbá a gyűjteményi rendszerekben¹¹ (mint amilyenek az irat- és levéltári, a múzeumi, valamint a könyvtári rendszerek) tárolt eredeti anyagokban és dokumentumokban (a továbbiakban: dokumentumokban) szereplő adatok (a dokumentumok rögzített tartalma). (A gyűjteményi rendszerek elsődleges adatait nevezik még *tárgyi adatoknak* is, mivel a keresés célját, tárgyat alkotják.)

A gyűjteményi rendszerek elektronikusan tárolt elsődleges adatain belül biztosítható a szabadon választott szavak szerinti keresés.¹²

(2) Másodlagos vagy *metaadatok*¹³ az adat, ha adatokra vonatkozik.¹⁴

A gyűjteményi rendszerekben a dokumentumokat a leírásaik azonosítják, melyek ugyancsak információs egységeket (dokumentumtétteleket, dokumentumrekordokat) alkotnak. A leírások és azok elemei metaadatok. Ez az adatok másodlagos vagy metaszintje. A dokumentumok tárolását úgy kell megoldani, hogy e rekordok a metaszinten kereshetők legyenek.

A metaadatokhoz is számos további adat tartozhat, ezek összessége metaadat-rekordokat alkot. A gyűjteményi rendszereknek éppen az a legfontosabb jellegzetessége, hogy a tárolt elsődleges, tárgyi adatok metaadatok segítségével leírhatók, kezelhetők és kereshetők.

Ha a névterek neveit gyűjteményi rendszerekben a dokumentumok vagy információk leírására alkalmazzák, akkor e nevek nem csak elsődleges adatok, hanem metaadatok is. A névterek neveinek ez az alkalmazása jellegzetes.

(3) A *harmadlagos (tercier) adatok* a metaadatok azonosítását, leírását és kezelését, valamint a tárolt adatok keresését biztosítják.

Gyűjteményi rendszerekben harmadlagos adat például a metaadat típusa (főcím, alcím, kiadási év, transliteráció típusa stb.), a felvétel/módosítás/törlés dátuma, rekordállapot (végleges, ideiglenes, törölt stb.), a név típusa (főnév, köznyelvi, történelmi stb.), transliteráció típusa, vonatkozó szabvány, tárgyszótípus (fő-, al- és melléktárgyszó), kitüntetett név, deskriptor, nemdeskriptor, elágazásjelölő stb.

Ha azonban ezek az adatok a névtéri rekord elemei, akkor elsődleges adatra vonatkozik, és ezért nem harmadlagos, hanem metaadat.

Valójában a harmadlagos adatok is metaadatok, és ezért, ha következetesek vagyunk, a (2) szerinti metaadatok esetében másodlagos metaadatokról, a (3) szerinti metaadatok esetében pedig harmadlagos metaadatokról kellene beszélni. A múlt század hetvenes éveitől nem is használták a metaadat kifejezést, hanem elsődleges, másodlagos és harmadlagos adatokról szolt a szakirodalom. Ezt követően azonban az informatika mindinkább meghatározta az adatterminológiát, ugyanakkor az informatikusok nem ismerték a manuális gyakorlatban kialakult szóhasználatot; nagyobb baj, hogy nem is nagyon törődtek a terminológiai részletekkel. Gyors eredményekre, könnyű, rövid, áttekinthető fogalmi világra törekedtek. Az addig használt adatformátumokról azt vallották, hogy túl bonyolultak, kereskedelmi szempontból egyszerűbb formátumok szükségesek. A Dublin Core formátum megjelenésével, mely elsőként lett metaadat-formátumnak nevezve, lassan feledésbe merült a másodlagos és harmadlagos adat fogalma, és csak a metaadat nevet használták. Ennek azonban az lett a következménye, hogy nem volt világos a harmadlagos adatok speciális jelentősége. Gyűjteményi rendszerekben például a földrajzi koordináták harmadlagos adatok, hiszen metaadatokhoz (a helynevekhez) kapcsolódnak, a rendszeren belül azoknak az adatai. A harmadlagos adatok egy része (a névosztályozási rendszerek, tipológiai stb.) kötött szótárakat igényel, de ez is csak a gyakorlatban volt felismerhető, nem tartozik

máig sem a gyűjteményi szakterületek szokásos ismeretanyagához. Mindez éppen a gyűjteményi rendszerek világában fékezte a kezelőrendszerek fejlesztésében az innovációt.

A továbbiakban az egyszerűség kedvéért a másodlagos metaadatok esetén metaadatokról, a harmadlagos metaadatok esetében pedig harmadlagos adatokról beszélünk.

Az adatok strukturáltságuk szempontjából (lényegében szerkezeti szempontból) lehetnek azonosító és leíró adatok.

- (4) Az *azonosító adatok* szabványokon, vagy egyéb normatív előírásokon, illetve megállapodásokon alapuló, természetes nyelven vagy formalizált, mesterséges nyelven kifejezett („kódolt”), egységesített, sok esetben összetett (strukturált) adatok.

Az azonosító adatok lehetnek elsődlegesen és másodlagosan azonosító adatok.¹⁵

- (4.1) Azok az azonosító adatok, amelyek csak egyetlen rekordhoz tartozhatnak, és ez fordítva is igaz, azaz egy rekordhoz csak egyetlen ilyen adat tartozhat, az *elsődlegesen azonosító adatok*. Lehetnek elsődleges, meta-, vagy harmadlagos adatok.

Ennek az összefüggésnek a kardinalitása 1 : 1. A közbeszédben ezeket az adattípusokat nevezik röviden azonosítóknak. Szerepük a tökéletes azonosítás, miáltal az általuk azonosított egység egészében válik kezelhetővé. Ilyen adat például a névterekben a kitüntetett névalak (és rajta keresztül a deskriptorok is), a gyűjteményi rendszerekben a dokumentumrekord azonosítója. Mindezekkel az elsődlegesen azonosító adatokkal a névrekordok egyértelmű azonosító leírása, kezelése és keresése valósítható meg. (Pontosan fogalmazva, ezek elsődlegesen azonosító elsődleges adatok.)

Például a *Babits Mihály* (1889–1941) személynév egy névtérben egyértelműen, azaz elsődlegesen azonosítja Babits Mihály költőt. Ez mindaddig így van, amíg nem kerül a névtérbe ugyanilyen nevű és kronologikus jellemzőjű újabb személy neve, aki azonban nem költő, hanem favágó. Ebben az esetben

a költő névtéri neve a következőre változhat: Babits Mihály (1889–1941 ; költő).

Elsődlegesen azonosító adat névterekben az 1. ábrán a kutya kitüntetett névhez kapcsolódó 0051052.

A gyűjteményi rendszerekben a dokumentumrekordok vonatkozásában a kitüntetett névalak és a deskriptorok nem elsődlegesen azonosítanak, hiszen ugyanaz a kitüntetett név egynél több dokumentumrekordhoz kapcsolódhat!

- (4.2) Azok az azonosító adatok, amelyek egynél több rekordhoz tartozhatnak, és egy rekord is több ilyen adatot tartalmazhat, a *másodlagosan azonosító adatok*. Lehetnek elsődleges, meta-, vagy harmadlagos adatok.

Az összefüggés kardinalitása M : N. (A kezelőrendszerekben ezt 1 : N kardinalitássá kell átalakítani.)

A természetes nyelvben a nevek másodlagosan azonosítanak, mivel lehetnek homonimák és szinonimák (a névterek kitüntetett nevei viszont elsődlegesen azonosítanak).

A névterek gyűjteményi rendszerekben alkalmazott nevei másodlagosan azonosító metaadatok. A kutya fogalomnévhez gyűjteményi rendszerben például több dokumentumrekord, és egy dokumentumrekordhoz nem csak a kutya kapcsolódhat.

Az *azonosító metaadatokkal* (legyenek elsődlegesen vagy másodlagosan azonosítók) a dokumentumrekordok egyértelmű azonosító leírása, kezelése és keresése valósítható meg. A tárolt dokumentumok bizonyos tulajdonságait azonosítják (cím, szerző, megjelenési hely, megjelenési év, dokumentumtípus, státusz, tartalom/tárgy/téma stb.).

Az elsődlegesen azonosító metaadatok biztosítják a dokumentumrekordok (és egyéb rekordok) egészében való kezelését.

A másodlagosan azonosító metaadatok egy része gyűjteményi rendszerekben a tárolt rekordok egyértelmű kereshetőségét lehetővé tévő *besorolási adat* (authority data). Ezek az adatok a rekordokon belül a keresés számá-

ra hozzáférési pontokat (access point) alkotnak.¹⁶

A névterek gyűjteményi rendszerekben alkalmazott nevei mindig besorolási adatok.

Besorolási adatok lehetnek *tulajdonnevek* (az állat-, személy-, család-, rendezvény- és területi nevek, műcímek) és *köznevek*.

A fogalomnevek, s ezzel besorolási adatok szerepét játszhatják a *kódszavak* (fogalmak mesterséges nyelven kifejezett neve, melyeket többnyire osztályozási rendszerek [taxonómiák, nomenklatúrák] tartalmaznak). Az osztályozási rendszerekben elsődlegesen azonosítják az osztályt, gyűjteményi rendszerekben viszont a nevekhez hasonlóan másodlagosan azonosító metaadatok. A 2. ábrán ilyen kódszó az 599.742.13. Könyvtári gyakorlatban az osztályokat azonosító kódszavakat *jelzetnek* nevezik. Ugyancsak kódszavak a nevek azonosító mesterséges nyelven kifejezett karakterláncok (az 1. ábrán ilyen a 000132).

A besorolási adatok egyrészt a *kitüntetett nevek*¹⁷ és a *deszkriptorok*¹⁸. Kialakításukra vagy szabványok – pl. ISBD, ISAAR, MSZ 3424 – vonatkoznak, a köznevek deszkriptorai és nemdeszkriptorai pedig kötött szótárból – tárgyszójegyzék, teaurusz – származnak. A könyvtári szakterületen ezeket a metaadatokat *egységesített besorolási adatoknak* (authority data) (is) nevezik. Gyűjteményi rendszerben nincs olyan deszkriptor, amely ne lenne egységesített besorolási adat, és ez fordítva is igaz.

Az 1. ábrán látható Rin Tin Tin (1918.09.10.–1932.08.10. ; német juhász), a kutya és az eb gyűjteményi rendszerekben deszkriptorok (az első kettő kitüntetett név), és egyben egységesített besorolási adatok.

Elvileg minden, a fogalmat jellemző adat a kitüntetett névalakhoz kapcsolódik, de a kezelőrendszereknek gondoskodniuk kell arról, hogy az egyes deszkriptoroknál is megjelenjenek adott esetben ezek a kitüntetett névalakhoz kapcsolódó jellemző adatok (l. a 3. ábrán pl. a típusadatokat, a relációkat).

A keresés a deszkriptorok (beleértve a kitüntetett névalakot) szerint végezhető, vagy úgy,

hogy találatként az összes deszkriptorral jellemzett dokumentumrekord, vagy csak a kereséskor megadott deszkriptorral jellemzett rekord jelenjék meg.

A besorolási adatok másrészt a *nemdeszkriptorok* szerepét játszó nem egységesített nevek (egyéb nevek). A nemdeszkriptorok mindig a kitüntetett névalakhoz, és azon keresztül adott esetekben az egyes ekvivalens deszkriptorokhoz kapcsolódnak (rájuk „utalnak”). A keresés a nemdeszkriptorok szerint is végezhető. A könyvtári szakterületen ezeket a metaadatokat *utalói adatoknak* („*utalóknak*”, „*nem egységesített besorolási adatoknak*”) nevezik,¹⁹ mivel velük keresve mindig azok a találatok jelennek meg, amelyek a nemdeszkriptorok helyett álló kitüntetett névhez (és rajta keresztül a deszkriptoraihoz) tartoznak. Például az 1. ábrán nemdeszkriptor a blöki, a kutyus, a kutyákok.

A nemdeszkriptorok a keresőrendszerekben is belépő szótárak (entry vocabulary) elemei lehetnek, melyekbe a közbeszédben, egyben a keresésre feltételezhetően ugyancsak alkalmazott esetleges nevek és leíró megnevezések (akár a szófelhők szavainak jelentős része) is felvehetők. Például „megőrülök”, vagy „éhes vagyok”, amelyek adott rendszerben a „pszichológus”, a „lelki segély”, illetve a „vendéglő”, „étterem”, „kifőzde” stb. deszkriptorokra utalhatnak. Ezek a belépőszótári adatok akár ki is zárhatók a böngésző keresésből, azaz a felhasználónak nem kell ezeket semmiféle mutatóban látnia (vagy csak opcionálisan jeleníthető meg a mutatójuk). Arra való, hogy a keresők szabadon megadott keresőszavai esetén nagyobb valószínűséggel biztosítsák a helyes találatokat.

A dokumentum szerzőségét (alkotóját), címét, adott esetben akár a kiadóját is stb. azonosító besorolási adatok *formális* adatok²⁰ közé tartoznak. Formális adatok ezeken kívül is léteznek (pl. kiadásjelzés, terjedelem). A formális adatok egy része nem azonosító, hanem leíró adatnak minősül, mivel nem kereshetők (legfeljebb szövegszóként lehet bennük keresni).

A névterek gyűjteményi rendszerekben alkalmazott tulajdonnevei lehetnek formális adatok.

A dokumentum tartalmát, tárgyát azonosító besorolási adatok a *tárgyszavak* (tárgyi kifejezések = topical term)²¹.

A névterek gyűjteményi rendszerekben használt tulajdonnevei lehetnek tárgyszavak is, hiszen dokumentumok tárgya lehet személy, család, testület, rendezvény, mű, földrajzi hely, cím.

A névterek gyűjteményi rendszerekben alkalmazott köznevei csak tárgyszavak lehetnek.

Gyűjteményi rendszerekben a névterek nevei (a tulajdonnevek és a köznevek) egyszerre besorolási adatok (tehát másodlagosan azonosító metaadatok), ugyanakkor önmagukban, tehát a tárolt dokumentumrekordoktól függetlenül azonosító elsődleges adatok.

A nevek tehát attól függően lehetnek elsődlegesen azonosító elsődleges adatok vagy másodlagosan azonosító metaadatok, hogy milyen szerepben használatosak: a megnevezett konkrét dolgok, illetve fogalmak azonosításának szerepében elsődlegesen azonosítanak (azaz elsődleges adatok); az elsődleges adatok azonosításának szerepében másodlagosan azonosítanak (azaz metaadatok).

A személynév például az adott személy vonatkozásában kitüntetett névalakként elsődlegesen azonosító elsődleges adat. Egy személynek csak egy kitüntetett névalakja lehet. Ugyanez a személynév az elsődleges adatokat tartalmazó dokumentumrekordok (s rajtuk keresztül a dokumentumok) vonatkozásában másodlagosan azonosító metaadat: egy kitüntetett névalakhoz több dokumentum tartozhat.

Például a földrajzi hely, képződmény tulajdonneve (a földrajzi név) e hely vonatkozásában kitüntetett névalakként elsődlegesen azonosító elsődleges adat, az elsődleges adatokat tartalmazó dokumentumrekordok (s rajtuk keresztül a dokumentumok) vonatkozásában másodlagosan azonosító metaadat.

A köznevek kitüntetett névalakjai elsődlegesen azonosító elsődleges adatként a fogalmat azonosítják: egy kitüntetett névalak szerepét játszó köznévként csak egyetlen fogalmat

azonosíthat. Másodlagosan azonosító metaadatként a dokumentum tartalmát azonosítják: egy szaktárgyszóhoz több dokumentum tartozhat.

A nevekhez és a dokumentumrekordokhoz kapcsolódó dátumok (pl. felvétel dátuma) és egyéb adminisztratív adatok attól függően elsődleges adatok, hogy névtérben, metaadatok, hogy dokumentumrekordokra, és harmadlagos adatok, hogy metaadatként szereplő nevekre alkalmazzák ezeket. Ha kereshetők, akkor másodlagosan azonosító adatok, ha nem kereshetők, akkor leíró adatok.

- (5) A másodlagosan azonosító metaadatok másik, többnyire kódszavak képviselte fajtája²² vagy a rekordkezeléshez szükséges, vagy olyan adatok, melyek az adott kezelőrendszerben nem kereshetők, azaz nem hozzáférési pontok az adatbázisban. Ezek a *vezérlő metaadatok* (control metadata).²³

A vezérlő metaadatok, ha nem kódszavak, a névterek közneveiből kerülnek ki. Többnyire előre meghatározott kötött szótárakat alkotnak.

Néhány példa.

A rekordtípus (könyv, időszaki kiadvány, aprónyomtatvány, kézirat, okirat stb.) meghatározza az adott dokumentum leírásához használandó többi metaadatot.

A státus vagy a dokumentum valamilyen állapotát (ép, töredékes, kölcsönözhető stb.) vagy a besorolási rekord, illetve a névtérrekord állapotát (végleges, ideiglenes stb.) határozza meg.

A jogosultság a használók vonatkozásában a hozzáférést határozza meg.

Mindezek a példákban felsorolt köznevek játszhatják a deskriptor szerepét is, és használhatók kereséshez is. Ha a vezérlő adatok a kereséshez is használhatók, akkor ebben a szerepükben valójában besorolási adatok is.

Harmadlagos adat is lehet vezérlőadat (pl. a tárgyszó típusát megadó név: személynév, testületi név; tárgyszó, altárgyszó, mellék-tárgyszó stb.).

- (6) A *leíró adatok* természetes nyelven rögzített (a köznyelvhasználat követelményei szerint megfogalmazott), formálisan nem strukturált metaadatok. Lehetnek *leíró elsődleges adatok*, *leíró metaadatok* és *leíró harmadlagos metaadatok*.

A leíró adatokkal a dokumentum- és névkordok azonosító tulajdonságainak, illetve tartalmának kötetlen leírása valósítható meg. A leíró adatok alapján szöveges formában a tárolt nevek, illetve a tárolt dokumentumok tulajdonságairól lehet mintegy „elbeszélő módon” tájékozódni, elsődleges rendelteté-

sük nem a keresés. Leíró adatok például a dokumentumrekordokhoz kapcsolódó szöveges magyarázatok, a nevekhez kapcsolódó meghatározások, forrásadatok (lásd a 3. ábrát), a tárgyak szöveges leírásai, továbbá a dokumentumokban szereplő címek, kiadónevek stb.

Az elsődleges adatokon belüli kereséshez hasonlóan a leíró adatok szövegszavai alapján (is) lehet keresni.²⁴ Ezek a szabadon választott keresőszavak (helytelenül „szabad tárgyszavak”).²⁵

Adathierarchia

adat

elsődleges adat

azonosító elsődleges adat
 elsődlegesen azonosító elsődleges adat
 másodlagosan azonosító elsődleges adat

 vezérlő elsődleges adat
 leíró elsődleges adat

kitüntetett név/deszriptor [névtérben], VIAF, URI név [beszélt nyelvben], utalóként szereplő név [névtérben], felvétel dátuma [névtérben] névtípus [névtérben], névstátus [névtérben] név meghatározása [névtérben], dokumentumszöveg

metaadat

azonosító elsődleges metaadat
 elsődlegesen azonosító metaadat
 másodlagosan azonosító metaadat

 besorolási adat
 egységesített adat
 utalói adat
 vezérlő metaadat
 leíró metaadat

dokumentumrekord-azonosító, ISBN, URI a kereshető, nem névnek számító adatok, ha nem vezérlők

kitüntetett név, deszkriptor nemdeszkriptor dokumentumtípus, dokumentumstátus, jogosultság cím és szerzőségi közlés bibliográfiai megjegyzések, dátumok (ha nem kereshetők)

harmadlagos adat

azonosító harmadlagos adat
 elsődlegesen azonosító harmadlagos adat

 másodlagosan azonosító harmadlagos adat
 vezérlő harmadlagos adat
 leíró harmadlagos adat

névazonosító (besorolási rekordazonosító) [ha a név metaadat]
 név besorolása (kategóriája, osztálya) [ha a név metaadat]
 névtípus, névstátus, jogosultság [ha a név metaadat],
 név meghatározása [ha a név metaadat]

Összegzés

Az adat típusát a szerepe is meghatározza. Név-adat két különböző – névtéri, illetve gyűjteményi – szerepben két különböző típusú adat. A tulajdonnevek és a köznevek vagy névterek azonosító elsődleges adatai, vagy gyűjteményi rendszerek azonosító metaadatai. A nevek egy-

máshoz bináris nyelvi és szemantikai relációkban kapcsolódnak [22].

Gyűjteményi rendszerekben a nevek a kereshető tárolást biztosítják, ezért nevezik az ilyen neveket információkereső nyelvek [21] szavainak [19] is, melyek között ugyancsak nyelvi és szemantikai relációkat adnak meg. Ezeket az állományokat

strukturált szótárakba szervezik, melyeket tezauszoknak, egyszerűbb válfajukat tárgyszójegyzéknek nevezik.

Információkereső nyelvi szavak lehetnek:

- az egységesített személy-, testületi és rendezvénynevek, valamint (az irattári, de főleg a múzeumi rendszereket illetően megszorításokkal), a címek állományai;
- a tárgyszavas (tárgyszójegyzékbe foglalt), a deskriptoros (tezauszba foglalt) nyelvek lexikai egységei;
- természetes nyelven alapuló hierarchikus osztályozási rendszerek („ontológiai”) szavai;
- az osztályozási rendszerek/nomenklaturák kód-szavai/jelzetei.²⁶

Az adatok tipizálása mindig valamilyen szempontból végezhető el. A fenti tipizálás szempontja az, hogy mi a viszonyuk az adatoknak az általuk megnevezett egyedi dolgokhoz, illetve egymáshoz képest. Nevezik ezt ontológiai szempontú tipizálásnak is, mert benne az adatok szintjeiről van szó. A fentiekben tárgyalt ontológiai tipizálás tekinthető át a 2. ábrán példákkal.

A névterek felépítése

A névterek a következő szintek alapján épülnek föl:

- a)** A (nemzetközi) szabványok szintje (pl. a levéltári ISAAR és ISAAD, a könyvtári ISBD, a tezauszokra vonatkozó ISO 2888).
- b)** A kommunikációs formátumok szintje (pl. a levéltári EAC, a könyvtári MARC).
- c)** Az adatmodell szintje, a fogalmi, nyelvi és előfordulási szintek korrekt szétválasztásával.²⁷
- d)** A (szótár-)kezelőrendszer szintjét (pl. a Getty kezelőrendszere [VCS = Vocabulary Coordination System, amely nem azonos az előbbi táblázatokkal], a könyvtári szakterületen a Relex).

Az **a)** (nemzetközi) szabványokhoz való alkalmazkodás alapkövetelmény.

A **b)** kommunikációs formátumok az adatcseréhez valók. Bármilyen **d)** kezelőrendszerről is legyen szó, minden olyan kommunikációs formátumban meg kell oldania az adatcserét, amely formátumok a gyűjteményi szakterületen szokásosak.

A **c)** (alapvetően a Codd-féle értelemben vett) adatmodellezés összetett feladat: van olyan

szintje, amelyben szükséges a rendszerprogramozói mellett a felhasználói közreműködés (hogy a rendszerprogramozó megfelelő információkat kapjon), és van olyan része, amely már informatikai feladat.

A **d)** kezelőrendszer az **a)**, **b)** és **c)** szintek értelemszerű figyelembevételével készül.

Mindez egyébként a nemzetközi téren ismert Getty-cég tezauszai²⁸ esetén is érvényesült.

Egyszerű lenne, ha a Getty, vagy valamelyik jelentős szoftverház a kezelőrendszerét is rendelkezésre bocsátaná, de ez nem így van. A Getty-cég például csak adatszótárát adja közre, továbbá állományainak kommunikációs formátumát, nem a kezelőrendszerét. Az utóbbi saját megfogalmazásuk szerint házon belüli rendszer („in-house editorial system”). Maguk a táblázatok (az adatszótár) valójában az adatmodell egyik vetületét alkotják, de egyáltalán nem a teljeset (pl. alig van nyoma az eljárásmodell adatainak). Ha valaki ennek nyomán saját kezelőrendszert akar kialakítani, keményen megfizeti az iskola-pénzt. (Ezen kívül a Getty külön szerződésben adatállományokat is átvesz, és ilyenkor saját állományát is rendelkezésre bocsátja. Ehhez előírja a saját kommunikációs formátumát.)

A Getty-féle adatszótár alapja egyrészt az ISO 2888-as és 5964-es nemzetközi tezauszszabvány (magyar megfelelője az MSZ 3418), másrészt a MARC formátum.

Célszerű különbséget tenni a kezelőrendszer (mint amilyen a Getty – nem hozzáférhető – VCS kezelőrendszere, vagy a – hozzáférhető – Relex), és a végfelhasználó számára szükséges szép grafikus felület (mint amilyen a Getty – mindenki által hozzáférhető – grafikus felülete) között (a Relexhez ez még sajnos hiányzik). Az utóbbi szép grafikus felület²⁹ csak a végfelhasználói keresésre használható, az előbbi a professzionális kereséshez, új rekordok létrehozásához, meg-lévők módosításához és törléséhez, tehát a teljes adminisztrációhoz.

A MARC-ban (egyúttal a Relexben) szereplő adatelemeket minden további nélkül el lehet rendezni adott esetben ugyanabban a táblázatos formában, ahogy az adatelemeit a Getty a szótárában közreadta. A Getty adatainak több, mint 90%-a megfeleltethető MARC adatokkal (a többi az eljárásmodell adata). Az EAC adatelem-

választéka ennél jelenleg jóval kisebb, de ugyan-
csak elrendezhető a Getty-féle táblázatok szerint.

Az egyes gyűjteményi rendszerekben kisebb-
nagyobb mértékben eltérnek az igények, de az
adatszerkezetet ez érdemben nem befolyásolja.
Például a levéltári anyagok jelentős része közirat,
a könyvtárakban viszont – a kéziratok anyagok
kivételével – túlnyomórészt nem köziratokat gyűj-
tenek; ezért kétségtelen, hogy a levéltári anyagok
személy- és testületi neveinek vonatkozásában
minden relációnak (tehát nemcsak az előzmé-
nye–folytatása, hanem az átfogóbb testülete-
részttestülete, a létrehozási és megszűnési ideje,
személy születési és halálozási ideje stb.) na-
gyobb a szemantikai súlya, mert hitelességi ösz-
szefüggései is vannak. De a felsorolt relációk (és
még több) kezelése a könyvtári és múzeumi
szakterületen is triviális.

Hasonló a helyzet a megjegyzések (scope note)
terén: levéltári szakterületen (pl. az ISAAR tanú-
sága szerint) sokkal hosszabb definíciós és tör-
téneti megjegyzésszövegeket alkalmaznak a
testületek és a személyek esetében, mint a
könyvtári szakterületen, ennek ellenére az adat-
modell szintjén abban semmi különbség nincsen,
hogy ezeket a megjegyzéseket mindegyik szakte-
rületen szükségesnek tartják. Strukturálisan (s
ezért a programrendszerek szintjén) tehát **lé-
nyegtelen különbség**, hogy az egyik szakterüle-
ten hosszabb, a másikon rövidebb megjegyzés-
szövegek szokásosak.

Mindezzel csak arra szeretnék rámutatni, hogy
számítástechnikai–adatmodellezési szempontból
mind a három gyűjteményi rendszer esetében
azonosak a szemantikai, így a relációkezelési,
tágabban pedig az adatmodellezési igények (az-
az a **modellezési/programozási munka lényegé-
ben azonos problémákat vet fel**).

Jegyzetek

¹ A Kr. e. 4.–5. században a Pandzsábi Taxilában élt
és tanított Panini, a szanszkrit nyelv s egyben a vi-
lág első nyelvtankönyvében felismerte ezeknek a
mélyeseteknek a létét. Felfedezése lényegében ha-
tástalan maradt a nyelvtudomány fejlődésére, és
csak a múlt század második felétől, amikor ezek-
nek az eseteknek a létét Paninitől függetlenül
megint felismerték, épült be a modern tudományba.
A programok elemi szerkezetét formalizáló
Backus–Naur formulát, melyben a mélyesetek al-
kalmazása gyakorlati szerepet kapott, Backus–
Naur–Panini formulának is nevezik.

² A dolog (szinonimája: entitás) fogalmának megér-
téséhez először a valami fogalmát kell meghatá-
rozni, jobb megértéséhez pedig célszerű a tárgy
fogalmát is bevonni a meghatározásba. Valami
minden, ami lehetséges vagy nem lehetséges, volt,
van, lesz. Ellentéte a semmi. Dolog ezen belül min-
den, amiről szó lehet, noha nem szükséges, hogy a
figyelem rá irányuljon, mert ha ez megtörténik, a
dologból tárgy lesz. Részletesebben I.: [13].

³ A gondolkodás egyik alapkérdése, hogy a fogalmak
önálló entitások (azaz dolgok), vagy sem. A nomi-
nalista felfogás szerint nem, a realista felfogás sze-
rint igen. A formális logikai mai állásából valójában
az következik, hogy a fogalmak is dolgok, hiszen
lehetnek a gondolkodás tárgyai. Márpedig olyan
tárgy nem léteik, mely ne volna dolog (amikor a
gondolkodás éppen nem irányul e dologra). Részle-
tesebben I.: [13]

⁴ A továbbiakban úgy járunk el, hogy a tulajdonneve-
ket a konkrét, a köznevek pedig az elvont dolgok
neveinek tekintjük abban az értelemben, hogy az
utóbbiak fogalomnevek. Tehát például a kutya köz-
névvel megnevezett fogalom terjedelmébe eső dol-
gok ugyan konkrétak, de a „kutya” név nem ezeket
a konkrét dolgokat nevezi meg, hanem a kutya fo-
galmát. Részletesebben I.: [13]

⁵ A szemantikai relációkról és a kardinalitásukról
részletesebben I.: [17].

⁶ A deskriptor és nemdeskriptor fogalmát eredeti-
leg a teaurusz-szabványokban [7] [8] [9] [11] ve-
zették be, a bibliográfiai szabványok [6] máig nem
vettek róluk tudomást, a MARC formátumba azon-
ban az elmúlt évtizedekben gyökeret vert. Nem
mondható ez el több más adat-csereformátumról,
és egyéb normatív dokumentumról például az EAC-
ról sem [1]. A mizériáról részletesebben I.: [14] és
[15].

⁷ A kitüntetett névalak és a deskriptor fogalmának
korrekt megkülönböztetése a gyakorlatban még
nem igazán érvényesül. A két fogalmat többnyire
azonosnak tekintik, mivel a hagyományos, az onto-
lógiai megjelenését megelőző informatikai gondol-
kodásban afféle jó nominalista szellemben nem lát-
szott értelmes dolognak a fogalmakat önálló dol-
gokként tételezni. Ez rányomja a bélyegét azokra a
szabványokra is, amelyekben a dolog fogalma
megjelenik ugyan (pl. az ISAAR [5], ahol entitásról
beszélnek), de úgy, hogy nincs tisztázva, jelenthet-
e fogalmat is? A névtérmodell kialakítását megcél-
zó FRANAR-dokumentumban [4] azonban már sze-
repl a fogalmi és a nyelvi szint szétválasztása.

⁸ A „műszaki” jelző meglepő lehet, noha magától
értetődik (mondhattuk volna, hogy „technikai”): hi-
szen gépek által kezelt rendszerekről van szó, azok
pedig műszakiak.

- ⁹ Adatbázisokban ezen a normalizálással lehet segíteni, miáltal a kardinalitások részben 1 : 1, többnyire pedig 1 N lesznek.
- ¹⁰ Az 1995-ben alakult Getty Images képügynökség az elmúlt két évtizedben négy jelentős tezauszot tett közzé a weben. Ezek valójában hatalmas földrajzi, képzőművészeti és építészeti, valamint személynév-névterek. A tezauszokat [3] a Getty ma kb. 70 millió vizuális dokumentumának osztályozására használják, de a weben rendelkezésre bocsátották a tezauszok pontos leírásait (segédleteit), az adatbázisukban használt adatszótárt és az adatszere-formátumot annak érdekében, hogy szerződéses ügyfeleik saját névállományait is betölthessék ezekbe a tezauszokba.
- ¹¹ Gyűjteményi rendszerek átfogó tárgyalását l.: [16].
- ¹² Ez a keresési lehetőség és alkalmazása impliciten következik minden jobb minőségű metaadat-kezelő rendszerből.
- ¹³ A terminológia információkereső nyelvekre vonatkozó részét részletesen tárgyalja [12].
- ¹⁴ A másodlagos vagy metaadatoknak semmi közük a nyílt bibliográfiai adatokon belül megkülönböztetett alapadatokhoz (cím, szerző, oldalszám stb.) és másodlagos bibliográfiai adatokhoz (ISBN, hordozó, felvétel dátuma stb.). L.: <http://openbiblio.net/principles/hu/>.
- ¹⁵ A két fő típus közötti különbségekről l.: [14][16].
- ¹⁶ A besorolási adatokat illetően egyrészt l. a nemzetközi bibliográfiai [6] és levéltári szabványt [5], továbbá a MARC21 adatszere-formátumot [10] és a tulajdonnevekre vonatkozó levéltári adatszere-formátumot [1]. Az utóbbiban a „besorolási adat” elnevezést nem használják, de angol változatában az ennek megfelelő „authority data” kifejezés szerepel.
- ¹⁷ A kitüntetett névalakot minden besorolási szabvány [4], [5] tárgyalja, de a fogalmi szintű azonosítással való kapcsolatában a lényeges dokumentum a [2].
- ¹⁸ A deszkriptorokat kiemelten tárgyalják a tezausz-szabványok [6], [7], [8], [10], továbbá a MARC21 besorolási adatszere-formátum [9].
- ¹⁹ A nemdeszkriptorok szerepét, önálló névrekordként való alkalmazásuk szükségességét részletesen tárgyalja [14].
- ²⁰ A „formális” vagy „formai” minősítés félreérthetősége miatt vitatott, mivel azt a benyomást kelti, hogy formalizált adatokról van szó, holott csak a „tartalmi” (a tartalmat azonosító) adatok ellentétére lenne hivatott utalni. Ettől függetlenül, a dokumentumot nem tartalmi értelemben, hanem a dokumentum bibliográfiai leírásában (levéltári terminológiával a levéltári anyagok leírásában) azonosító adatok kérdése a vonatkozó bibliográfiai, levéltári és múzeumi szabványokban triviális.
- ²¹ A tárgyszavakat a könyvtári szakirodalom hagyományosan tárgyalja, a „topical term” kifejezés a MARC21 terminológiájához tartozik [9].
- ²² Köznapi nyelven nemes egyszerűséggel csak „kódnak” nevezik ezeket a mesterséges nyelven megfogalmazott metaadatokat, holott a „kód” magát az egész nyelvet jelenti (l.: [8]).
- ²³ A vezérlőadatok kérdésénél tudomásom szerint módszeresen nem tárgyalja a dokumentum, de létük nemcsak minden adatszere-formátumból, hanem minden rendszer megvalósításból triviálisan következik.
- ²⁴ Ez a keresési lehetőség és alkalmazása impliciten következik minden jobb minőségű metaadat-kezelő rendszerből.
- ²⁵ A „szabad tárgyszó” fogalmának ellentmondásosságát részletesen tárgyalja [11]. Lényege, hogy a tárgyszó mindig szabványosított szó, ezért a „szabad tárgyszó” elvileg azt jelenti, hogy „nem szabványosított szabványosított szó”.
- ²⁶ A névtér adatokat módszeresen, legalábbis az itt tárgyalt ontológikus szemlélet alapján, tudomásom szerint más dokumentumban még nem tárgyalták. Minden helyreigazítást háláson fogadok!
- ²⁷ E modell az alapja például a Getty-féle, a weben rendelkezésre bocsátott táblázatoknak is (5.0 Addendum Z: Vocabulary coordination system data dictionary: draft November 2003) http://www.getty.edu/research/tools/vocabularies/guidelines/tgn_5_0_addendum_z_data_dict.pdf [3].
- ²⁸ Négy tezauszról van szó: Művészeti és építészeti, Kulturális tárgyak (műalkotások), Földrajzi és Művészek életrajzi tezauszja [3].
- ²⁹ L. például a Győr névcikket a Földrajzi tezauszban: http://www.getty.edu/vow/TGNFullDisplay?find=Gy%C5%91r&place=&nation=&prev_page=1&english=Y&subjectid=7010542

Irodalom

- [1] EAC-CPF Encoded Archival Context. Corporate Bodies, Persons, and Families: Tag library. Elements. Society of American Archivist. Staatsbibliothek zu Berlin. [Kezelő: National Archives and Records Administration] 2010. <http://www3.iath.virginia.edu/eac/cpf/tagLibrary/cpfTagLibrary.html#d1e822>
- [2] Definition of the CIDOC Conceptual Reference Model. – Produced by the ICOM/CIDOC ; Documentation Standards Group, Continued by the CIDOC CRM Special Interest Group. Version 5.0.4. November 2011. – CRM Special Interest

- Group, 2011. <http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf>
- [3] Getty Vocabularies. Search tools & databases. The Getty Research Institute, Los Angeles, 2010–<<http://www.getty.edu/research/tools/vocabularies/>>
- [4] IFLA. Functional Requirements for Authority Data. A Conceptual Model. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Draft. 2007. 04. 01. <<http://www.ifla.org.sg/VII/d4/FRANAR-Conceptual-Model-2ndReview.pdf>> <http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm>
- [5] ISAAR (CPF) : International Standard Archival Authority Record for Corporate Bodies, Persons and Families. Second ed., International Council on Archives. ICA Committee on Descriptive Standards, Canberra, 2004. <http://bfi.archivportal.hu/id-1037-isaar_cpf_szervezetek_testuletek.html#521>
- [6] ISBD : International standard bibliographic description / recommended by the ISBD Review Group ; approved by the Standing Committee of the IFLA Cataloguing Section. – Consolidated ed. – Berlin ; München : De Gruyter Saur, 2011. – xvii, 284 p. ; 25 cm. – (IFLA series on bibliographic control ; vol. 44). – ISBN 978-3-11-026379-4
- [7] ISO 2788–1986 ; Documentation – Guidelines for the establishment and development of monolingual thesauri
- [8] ISO 5964–1985 ; Documentation – Guidelines for the establishment and development of multilingual thesauri
- [9] ISO 25964–2011 ; Information and documentation – Thesauri and interoperability with other vocabularies
- [10] MARC21 Format for authority data. [Washington], Library of Congress. Network Development and MARC Standards Office; Cataloging Distribution Service. 1999 Ed., Update No. 1 (October 2001) through Update No. 9 (October 2008). Washinton, Library of Congress, Network Development and MARC Standards Office. <<http://www.loc.gov/marc/authority/ecadhome.html>>
- [11] MSZ 3418–87 ; Magyar nyelvű információkereső tezauszok. Szerkezete, részei és formái
- [12] UNGVÁRY Rudolf: Az információkeresés szavai. = TMT, 50. köt. 12. sz. 2003. p. 1–27. <http://tmt.omikk.bme.hu/show_news.html?id=3451&issue_id=446>
- [13] UNGVÁRY Rudolf: A kategóriák rendszere. – A szavak hálójában projekt tanulmánya. 2003. [kézirat] <<http://ontologia.hu/document/paper/>>
- [14] UNGVÁRY Rudolf: MARC21/HUNMARC. A besorolási adatok adatsere-formátuma. Főbb jellemzők, fejlődés és problémák. = Könyvtári Figyelő, 20. (56.) köt. 1. sz. 2010. p. 4–70. <<http://ki.oszk.hu/kfi/2010/10/marc21hunmarc-a-besorolasi-adatok-metaadat-formatuma-fobb-jellemzok-fejlodes-es-problemak/>>
- [15] UNGVÁRY Rudolf: A MARC formátum és a nemdeszkriptorok. = TMT, 57. köt. 3. sz. 2010. p. 4–19. <http://tmt.omikk.bme.hu/show_news.html?id=5301&issue_id=513>
- [16] UNGVÁRY Rudolf: Metaadatok összehasonlító vizsgálata gyűjteményi rendszerekben. = TMT, 58. köt. 6. sz. 2011. <http://tmt.omikk.bme.hu/login.html?initreq=%2Fshow_news.html%3Fissue_id%3D527%26id%3D5505>
- [17] UNGVÁRY Rudolf: A szemantikai relációk, 2011 [kézirat, előkészületben].

Beérkezett: 2012. I. 26-án.



Ungváry Rudolf
informatikus mérnök.
E-mail: ungvary@hungary.com

