

Az ETO adatbázis magyar kiadása és konkordanciája

Minden bizonnyal az egymást követő ETO-kiadások legnehezebben összeállítható, egyben a legterjedelmesebb egybevetése, konkordanciája van készülóban. Az idő sürgeti, hiszen a nagy szolgáltató könyvtárak 2008 elejére ütemezték átállásukat az új kiadású jelzetrendszerre, ez pedig konkordancia nélkül sokkal nehezebb.

Barátné Hajdu Ágnes a Könyv, Könyvtár, Könyvtáros 2007. évi februári számában beszámolt az ETO-fejlesztés mostani rendszeréről, példákat is sorolt a fontosabb változásokból. Teljes valójában azonban a tételes egybevetés-kor derül ki, mi minden történt az utolsó előtti, 1990-es magyar ETO-kiadás és a 2000. évi változtatásokkal záródott 2006. évi között. Bizony, nagyon sok minden!

Említsük meg legelőször is azt a tényt, hogy míg a mi 1990-esünk (a továbbiakban: 1990) középkiadás volt, addig a 2006-os (a továbbiakban: 2000, mert a 2000-es UDC MRF adatbázis jelzeteivel zárt) pedig teljes. A kettő közti jelzetmennyiség aránya nagyjából 1:2, aminek igen nagy a jelentősége.

Az ETO magyar számítógépes adatbázisában lévő, WINISIS-szel kezelt jelzetek megnevezései és egyéb szöveges részei kétnyelvűek. Úgy készültek, hogy a feldolgozott 1990-es jelzetekhez hozzáfűztük a UDC Consortium 2000-es évjáratú, akkor aktuális rekordjait, és a rekordon belül elkészítettük a mezők fordítását, illetve többszörösen ellenőriztük azokat.

Adatszerkezet

A jelzetrekordok kétnyelvűek. (A rekordok mezőleírását lásd a következő oldalon!) Az 1–99 közötti mezőtartomány magyar megfelelői az 500–599, a 100–199 közöttiek a 600–699, a 900–999 közötti adminisztratív mezők a 700–799 mezőtartományba kerültek (a táblázatban dőlt betűvel jelölve). Nem kettőztük meg azokat a mező-

ket, amelyek csak számokat tartalmaznak, vagyis nem volt szükség a fordításukra.

Az ETO következő kiadásai várhatóan cédén jelennek meg Magyarországon, a fenti rekord-szerkezetből a technikai jellegű adatokat elhagyó exporttal. Jelenleg a 2001 és 2004 közötti változásokkal rendelkezünk, az elképzelések szerint ezt 2006-ig kiegészítve, vagyis naprakészen szeretnénk a cédé-adatbázist közreadni. Ehhez szükség lesz a kétévnyi hiány előfizetési díjának előteremtésére és a kiadási jog megszerzésére, vagyis közel 2700 euróra, ami nem könnyű dolog a mai finanszírozási viszonyok között. És akkor a honosítás (fordítás, lektorálás, ellenőrzés) költségeit még nem is említettük. Talán egy pályázat segíthet rajtunk...

A formátumról a felhasználók körében végzendő felmérés nyomán születik majd döntés. Kérjük az integrált könyvtári rendszerek használóit, fejlesztőit, gondolják át, milyen formában tudja rendszerük a jelzetrekordokat fogadni, és a sunt@oszk.hu címre írják is meg, hozzátéve, ha esetleg speciális igény szerinti kimenetet kérnek majd, amikor a 2006-os frissítés hozzáférhető lesz. A WINISIS kimenetei: ISO 2709, text és XML.

UDCC-segédanyagok, a konkordancia forrásai

A rekordszerkezet jól mutatja, hogy a 900-as mezőtartományban kapunk információt a jelzet sorsáról: új, törölt, módosult, a revízió típusa (melyik mezője változott), a revízió forrása, története, megjegyzés ezekhez.

Úgy tűnik, minden együtt van ahhoz, hogy az adatbázisból kinyerjük a konkordanciát, a helyzet azonban egészen más. Sajnálattal kellett ismét megállapítanunk, hogy az MRF-adatbázis alapja nem a mi 1990-es kiadásunk, hanem több nemzeti kiadásból szemezgetve állították össze. Büszkén hozzátehetjük persze, hogy mi is a kiindulópontok között voltunk és később is vettek át új jelzetként a miénkből, illetve módosítottak a miénk szerint már meglévőket, értelemszerűen újként, ill. módosítottként megjelölve azokat.

Rekordszerkezet

Mező-azonosító (tag)	Mezőnév	Almezők	Max. hossz	Adattípus*	Ismételhetőség**
1	UDC number / ETO-szám		1000	0	0
2	Table / Táblajel		1000	1	0
3	Special auxiliary type / Speciális alosztás típusa		10	0	0
4	Combination type		6	0	1
10	Derived from par.instr. at UDC		50	0	0
11	Parallel divided as UDC	abtg	50	0	1
12	Par.div. provides spec.aux.		10	0	1
13	Applic.note provides spec.aux.		10	0	1
100	Description (With subfields for language versions)	egsl	250	0	1
105	Verbal examples (With subfields for language versions)	egsl	250	0	1
110	Scope note (With subfields for language versions)	eg	250	0	1
111	Application note (With subfields for language versions)	eg	250	0	1
115	Combination examples	abcdgnry	250	0	1
120	Parallel division examples	adg	250	0	1
125	References	atg	250	0	1
504	Jelzet-összekapcsolási típusok		1000	0	1
510	Párhuzamos jelzetszerkesztés		1000	0	0
511	Párhuzamos jelzetszerkezet	abtg	1000	0	1
512	Speciális alosztás, párhuzamos jelzetszerkezet		1000	0	1
513	Speciális alosztás alkalmazási magyarázata		1000	0	1
600	Megnevezés		1000	0	0
605	Szóbeli példák	eg ^{***}	250	0	1
610	Érvényességi terület magyarázata	eg ^{***}	1000	0	1
611	Alkalmazási terület magyarázata	eg ^{***}	1000	0	1
615	Jelzet-összekapcsolási példák	abcdgnry	1000	0	1
620	Párhuzamos jelzetszerkesztés: példák	adg	250	0	1
625	Utalások	atg	250	0	1
704	Bevezető megjegyzések		250	0	1
714	Megjegyzés a törléshez		250	0	1
724	Megjegyzés a revízióhoz		250	0	1
725	A revízió története	fds	1000	0	0
752	Speciális jelek használata		250	0	1
755	Szerkesztői megjegyzés		250	0	1
757	Admin. megjegyzés (belső)		1000	0	0
758	A következő Extensionből (belső)		1000	0	0
760	Mutató		1000	0	1
799	Csak ideiglenes (belső)		1000	0	1
901	Introduction date / Bevez.		1000	2	0
903	Introduction source / B.forr.		1000	0	1
904	Introduction comments		250	0	1
911	Cancellation date / Törl. idő		1000	0	0
912	Replaced by UDC# / Módosítás		1000	0	1
913	Cancellation source / T. forr.		1000	0	1
914	Cancellation comments		250	0	1
921	Last revision / Utolsó rev.		1000	0	1
922	Revision type / Rev.típus		1000	0	1
923	Revision source / Rev.forrás		1000	0	1
924	Revision comments		250	0	1
925	Revision history	fds	100	0	1
951	Terms for index only / Index		1000	0	1
952	Use special characters		250	0	1
955	Editorial annotation		250	0	1
957	Admin notes (internal)		250	0	0
958	For next E&C (internal)		1	0	0
999	Temp work only (internal)		4	0	1

* 0 – betű és szám is lehet, 1 – csak betű, 2 – csak szám

** 0 – nem ismételhető, 1 – ismételhető

*** Az almezőszerkezet itt felesleges, megszűnik

Az új kiadás előkészítésekor tapasztaltak nyomán és megsejtve, hogy a konkordancia összeállítás nem lesz könnyű, a táblázatot kiegészítettük két, eddig soha nem közölt adattal: az új jelzet bevezetési dátumával, ill. a revízió (módo-

sítás) dátumával, ezeket egyszerűen (és automatikusan, vagyis egyenkénti ellenőrzés nélkül) megjelenítettük a jelzet rekordjából. Amiért az iménti mondat alapján büszkélkedésre lehetett okunk, a tényleges egybevetés sok esetben még-

is azt eredményezte, hogy a jelzet a mi utolsó kiadásunkhoz képest nem új és nem is módosult. (A tetemes mennyiség és a gyakorlati használat szempontjából nem különösebben lényegesség okán most letettünk arról, hogy e kiigazításokkal megterheljük a konkordanciajegyzéket.)

Szerződésünk alapján 2005-ben a konzorcium átadta a törölt jelzetek adatbázisát, hiszen azok nincsenek már benne a teljes, csak a különálló, éves törlések adatbázisában. (Pedig jó lenne, ha benne lennének, sokkal könnyebb volna átlátni a revíziókat.) A kapott adatbázis neve: General ref 1994–2003.

Gondok

Indexelés után derült ki, a törölt jelzetek adatbázisa 5046 jelzetet tartalmaz. Az utód nélkül, elavulás okán törölt jelzetek száma 1096, elég magas szám! Dátumuk 1993 és 1998 közötti, tehát nemhogy 2003-ig nem tartalmaz adatokat, de a 2000-esek sincsenek benne.

Sebaj, az évenként érkező összeállítás külön adatbázisban tartalmazza a tárgyévi új, módosult és törölt jelzeteket, sőt szövegfájlban is megkaptuk ezeket – szerencsére, mert kiderült, a törölt adatbázisa egészen mást tartalmaz: nem tudni, milyen célú gyűjtés volt. Félévnyi „mail-ostrom” után végre sikerült hozzájutni a tényleges 2000. évi adatokhoz. Ezt és a 2000-ig a Megnevezés mezőben módosult jelzeteket „összesöpörve” összeállt az alap, amit a konkordanciakészítés reális fundamentumának hittünk. A törölt/áthelyezett jelzetek száma 6814-re emelkedett, az utód nélkül töröltek száma mindössze hárommal, 1099-re növekedett.

Milyen egyszerű: készítünk egy szendvicsadatbázist, az elején lévő UDC-segédanyag, valamint a 2000-es jelzetek közé beékeljük az 1990-est, és ennek a változásait keressük a rekordban lévő tájékoztató információk alapján...

A konkordancia alapelve

A konkordancia az 1990-es és a 2000-es állapot közti átjárást oldja meg. Mindenekelőtt tisztázni kellett a konkordanciakészítés alapvető elvét. Ez pedig az volt, hogy „jelzet nem vész el, legfeljebb átalakul”, mint ahogy a fogalmak sem vesznek el, legfeljebb (de teljesen sosem) elavulnak. Márpedig ha adott jelzet valamely dokumentumot jellemez egy élő katalógusban s a dokumentum létezik, megvan,

jelzetének is lennie kell. Mindezt azért volt szükségszerű kijelenteni, mert 2884 darab 1990-es jelzet tűnt el nyomtalanul a 2000-es táblázatból. Ez nagyon magas szám, és annak ellenére magas, hogy 748 olyan jelzet is akad, ami szerepel ugyan a töröltek adatbázisában, de később (új jelzetként jelölve!) visszakerült.

A „szendvics” említett három elemének egybevetésekor, eljutva az I. i segéd táblához, tapasztalhattuk, hogy a szempont szerint nem önálló általánosan közös alosztások (.00) csoportja kiüresedett! Ami különben nem volt váratlan, hisz már Babiczky tanár úr óráján azt hallottuk, hogy használatát lehetőleg kerüljük. Csak elvétve akadt néhány elszánt, a túlbonyolításra hajlamos kolléga, aki használta ezeket az alosztásokat.

A források száma további anyaggal bővült, az eddig is rendelkezésre álló E&C (Extension and Correction) évenként kiadott füzetével. Csakhogy lehetőleg olyan forrást akartunk használni, amiből a vágólapon keresztül vagy még egyszerűbben átemelhetőek a jelzetek, minimálisan csökkentendő az elütés hibalehetőségét.

A szerencse megint mellénk állt, a konzorcium weboldalán némi keresgéléssel megtaláltuk a törölt jelzetek évenként felsorolását szövegfájlban. Átvizsgálva ezeket kiderült, az I. i segéd táblát 1997-ben és 1999-ben számolták fel, vagyis a kapott összesített adatbázisnak ezt is tartalmaznia kellett volna! A kérdést, miért nem, már nem is volt érdemes feltenni, csak elgondolkodni azon, vajon a rendelkezésre álló forrásainkat ennyi kellemetlen tapasztalat mellett tekinthetjük-e biztosnak és teljesnek. Az utóbbi kételyre egyértelmű „nem” a válasz, a másiktól – gondoltuk –, menet közben kiderül, mennyire pontos. Előre jelzem: nem az... Az I. i segéd tábla esetében a néhány átirányított jelzet mellett a legtöbb megszűnről itt is csak ennyit találunk: „A szempont szerinti speciális alosztások I. i segéd táblázata teljes egészében megszűnt. Az így jelölt fogalmakat viszonyítással vagy az I. k táblázat újjólág létrehozott -02 jelzetsorával osztályozzuk, szükség szerint kiegészítve a -056.15... alosztással.”

Végezetül: az előzőekkel azonos fontosságúnak tekintettük, hogy a gépi rendszerek számára is használható számítógépes állományt hozzunk létre a változások minél nagyobb arányban automatikusan végezhető átvezetéséhez. Az egyértelmű volt, ebben nem tudunk 100%-os eredményt elérni.

Új munkamódszer

Természetesen nem mondtunk le a jelzetátírányítás konzorciumi adatbázisáról, azzal, hogy tételeit bizony ellenőrizni kell. A kétféle mélységű táblázat egybevetésekor számos eset fordul elő.

1) Az 1990-es jelzet tovább finomodik, ekkor szöveges feloldása is változhat, és részletező továbbosztásokra kell utalni.

2) Bármilyen ellentmondásos is, előfordul, hogy az 1990-es táblázat tartalmaz részletező alosztásokat, a 2000-es nem, ebben az esetben „vissza kell irányítani” az alosztásokat a hierarchia magasabb szintjére. A változások automatikus átvezethetősége szempontjából ez a kedvezőbb megoldás, mert az előző példa emberi döntést igényel.

3) A jelzetfeloldás automatikus szövegelemzésével kiszűrtük a 100%-os biztonsággal változás nélküli jelzeteket. Az elemző algoritmus *Tar Katalin* (OSZK Informatikai Igazgatóság) kiváló munkáját dicséri.

4) A képződött maradék tartalmaz a szövegelemzésből eredő, ám esetleg nem releváns továbbutalásokat, ezeket törölni kell.

5) Különleges és figyelembe veendő eset a módosulásoké, mert megesik, hogy a jelzetfeloldás fő része (az adatbázisban a megnevezés) több részre (a 600-as, 605-ös és a 615-ös mezőre) oszlik. Ezt úgy tekintjük, hogy a jelzet nem módosult.

6) Utólag látható, szigorúbban kellett volna megkövetelni a jelzetek szöveges részének fordítóitól, hogy a világért se nyúljanak hozzá a meglévő magyar szöveghez, ha az értelmezésében megegyezik az angollal. Ne „csinosítsanak”. Sajnos, sokhelyütt ez történt, amitől a szövegelemzés két ilyen jelzetet nem tekint azonosnak. Megint csak emberi döntés kell a minősítéshez.

Néhány példa a konkordanciaesetekre

003.326 (Földközi-tengeri logografikus írások. Ókrétai (minoszi), ciprusi szótagírás, hettita képírás)

Mód.: Földközi-tengeri logografikus írások

—>003.326.3 Ciprusi szótagírás

—>003.326.5 Hettita képírás

A jelzetet továbbosztályozták, részekre bontották.

005 (A szervezés (organizáció) tudománya. Szervezéstudomány. A szervezés módszertana, analízis, szintézis, osztályozás és taxonómia elmélete és alapelvei, rendszerezés, szisztematizálás általában)

--> 001.82 Szervezetek tanulmányozása

--> 57.06 Az élő szervezetek nevezéktana és rendszertani osztályozása. Taxonómia

A jelzet megszűnt. A fogalom két részre, két szempontú megközelítést megjelenítő fogalomra osztódik. Az egyik egyben egy új jelzet is: 001.82, míg a másik egy már meglévő jelzetre utal, tehát itt nincs módosulás, csak a párhuzamosság megszüntetése.

(234.373.25) (Pilis)

H. h. (234.3739Pilis)

Sok földrajzi alosztás a korábbinál kevésbé részletező, ezért az egyszerű földrajzi alosztás helyett kénytelenek vagyunk névalosztással pontosítani.

006.071 (A gazdaságra, az életszínvonalra gyakorolt hatás)

—>006.07 *A szabványosítás és a szabványok hatásai*

A jelzet megszűnt, új helyét az E&C nem (vagy hibásan) jelöli, a magyar konkordancia szerkesztői által javasolt visszaléptetés a megoldás.

061.236.62 (Pithias-lovagok)

—>061.236.6 *Érdekvédelmi célkitűzésű titkos és félig titkos társaságok, mozgalmak, szövetségek, rendek*

A jelzet megszűnt, új helyét az E&C nem (vagy hibásan) jelöli, a magyar konkordancia szerkesztői által javasolt visszaléptetés a megoldás. (A szerkesztők javaslatai minden esetben dőlt betűsek.)

*

Összegzésül: kialakult a konkordancia nyersanyaga, amit egybevetésekkel „csak” ritkítani kell. Induló állománya 17 333 jelzetbokr volt. Kis makrókkal a felesleges elemek törlését lehetett gyorsítani, a többi pedig csak szem és friss szellemű megítélés dolga.

A „maradék” is tetemes lesz, vagyis tényleg minden idők legnagyobb terjedelmű konkordanciája készül...

Fejős László