

# KÖVETKEZETES TARTALMI FELTÁRÁS SZOFTVERCSERE MELLETT? KÖZTAURUSZ HASZNÁLATA A GYŐRI KISFALUDY KÁROLY MEGYEI KÖNYVTÁRBAN

*Gáncsné Nagy Erzsébet (KKMK), [e.gancs@mail.kkmk.hu](mailto:e.gancs@mail.kkmk.hu)  
Lengyel Monika (MTA SZTAKI), [lmoni@sztaki.hu](mailto:lmoni@sztaki.hu)  
Pethő Árpádné (KKMK), [plili@mail.kkmk.hu](mailto:plili@mail.kkmk.hu)*

Könyvtárunk 1997. októberétől tárja fel dokumentumait számítógépes adatbázisban. Ettől az időponttól kezdve a tartalmi feltárás ETO szakszámok és tárgyszavak együttes használatával történt. Tudásunkat az eltelt idő és a tapasztalat, lehetőségeinket pedig a rendelkezésre álló szoftver befolyásolta, alakította. Tanulságos lehet ezt a folyamatot egy rövid kronológiában áttekinteni. A helyismereti cikkek tartalmi feltárása eltér a többi dokumentumétól, ennek taglalása nem célja jelenlegi ismertetőnknek.

## **1. Tárgyszavazás kronológiája a győri Kisfaludy Károly Megyei Könyvtárban**

*1997-1999*

1997-től az SRLIB könyvtári rendszerben kezdtük meg a dokumentumok természetes nyelvű tartalmi feltárását. Célunk az volt, hogy szakkatalógus rendszerünkkel közel azonos szintű visszakeresési lehetőséget biztosítsunk használóinknak. Ebből a logikából adódik, hogy a szépirodalmi dokumentumokat is tárgyszavaztuk és tárgyszavazzuk ma is.

A tárgyszavak listája a közösen kialakított elvek alapján, a tartalmi feltárás folyamatában bővült. Nem volt tehát előre elkészített listánk. A következetességet biztosította, hogy a tárgyszavazás 1999. márciusáig egy könyvtáros feladata volt. Ez alatt az időszak alatt 6467 dokumentum tárgyszavazása történt meg. A tárgyszavak "saját fejlesztésűek" voltak, az Új Könyvek tárgyszavait követték illetve az ETO szakszámok természetes nyelvű feloldásán alapultak.

Az általunk használt szoftver az alá-főlérendeltséget és a szinonima kapcsolatot tudta kezelni. A szoftvernek ezt a lehetőségét ki is használtuk. Ugyancsak a szoftver befolyásolta azt a döntésünket is, hogy négyféle tárgyszó rendszert építettünk (Szabad, Kiról-miről, Mikor, Földrajzi). A tárgyszóra való keresés ugyanis könnyebb és gyorsabb volt, mint ha az ún. „leíró” adatmezőkben kerestünk volna, továbbá itt nyílt lehetőség összetett keresésre is.

*1999-2002*

- Tartalmi feltárásunkban fordulat 1999. júniusában következett be. Az MKE országos munkabizottságot alakított az Ungváry Rudolf által szerkesztett KÖZTAURUSZ tesztelésére, alkalmazási lehetőségének vizsgálatára. Könyvtárunk munkatársa is tagja volt ennek a munkabizottságnak, így természetes volt, hogy az 1999. novemberétől életbe léptetett szabályozás alapján, az új deskriptor létrehozásának követelményeit az alábbiakban rögzítettük:

- illeszkedjen a KKMK meglévő tárgyszókészletéhez;
- szerepeljen a KÖZTAURUSZ-ban;

Ekkor már négy könyvtáros foglalkozott tárgyszavazással. A könyvtárosok munkáját a RELEX teauruszkezelő megkönnyítette, hisz segítségével a teljes struktúra ismeretében választhatták ki a megfelelő deszkriptort, amit kézzel rendeltek a dokumentumhoz. A használók tájékozódásának megkönnyítésére a KÖZTAURUSZ-ból beírták az adott deszkriptor SRLIB által kezelni tudott kapcsolatrendszerét is (alá-fölrendelt, szinoníma).

A gond azonban az volt, hogy a KÖZTAURUSZ és az SRLIB adatbázis között nem volt meg a kapcsolat, megfosztva a használókat attól a lehetőségtől, tudásbázistól, amit a KÖZTAURUSZ nyújtott.

2002-2005

Érezve a kuszaságot 2002-ben elkezdtük saját tárgyszavaink megfeleltetését a KÖZTAURUSZ deszkriptorainak. Ungváry Rudolfnak pedig elküldtük a KÖZTAURUSZ-ból hiányzó, de beépítésre javasolt kifejezések listáját, melyek nagy részét Ungváry Rudolf be is építette a KÖZTAURUSZ-ba.

Körvonalazódott annak a lehetősége is, hogy megteremtődik a kapcsolat az SRLIB adatbázis és a RELEX teauruszkezelő között. Ettől a pillanattól kezdve a KÖZTAURUSZ-ból csak a deszkriptorokat rendeltük a dokumentumokhoz. Abban bízunk, hogy a két program közötti kapcsolat belátható időn belül létrejön, de sajnos nem így történt. Tárgyszó rendszerünk így még áttekinthetlenebb lett.

Az általunk ismert következetlenségre Prókai Margit is felhívta fel a figyelmet a Tudományos és műszaki tájékoztatásban megjelent cikkében.<sup>1</sup>

2005-

2004-ben sikeres pályázatunk nyomán megteremtődött az anyagi háttér új integrált könyvtári szoftver beszerzésére. A közbeszerzési eljárás pályázati kiírásának egyik fontos kitétele volt a KÖZTAURUSZ kezelésére való alkalmasság, mely magas súlyponttal is szerepelt a pályázat értékelésénél. A pályázatot a HUNTEKA könyvtári rendszer nyerte el.

*Mit vártunk / mit várunk el az új szoftvertől? Miért ragaszkodtunk a KÖZTAURUSZ-hoz?*

Ragaszkodásunknak csak egyik oka az, hogy 1999 óta használjuk ezt az átfogó teauruszt. Számunkra fontos érv az is, hogy egy karbantartó felületen<sup>2</sup> lehetőségünk van módosítás, illetve kiegészítés kérésére, arra tehát, hogy a szerkesztett szótár az igényeink figyelembe vételével fejlődjön, alakuljon.

Azt kértük tehát a szoftver fejlesztőtől, hogy a könyvtárosok számára a HUNTEKA beviteli felületén álljon rendelkezésre a KÖZTAURUSZ teljes kínálata, teljes kapcsolatrendszere, amelyből a használni kívánt deszkriptort átemelhetik az adott dokumentumhoz. A használóknak pedig biztosítsanak olyan élő, bibliográfiai rekordhoz kapcsolt tárgyszó indexet, ahol KÖZTAURUSZ kifejezés esetén látszik a hozzá tartozó struktúra is.

Tehát nem kevesebbet kértünk, mint hogy egyfelől segítsenek a kuszaságból rendet teremteni, másrészt a HUNTEKA OPAC felületén biztosítsanak egyértelmű lehetőséget a tartalmi visszakeresésre.

<sup>1</sup> Fogalmi választás és nyelvi vonzás a tárgyszavas információkereső nyelvi modellekben / Prókai Margit. In: Tudományos és műszaki tájékoztatás, 2005. 5. sz. p. 218-234.

<sup>2</sup> [www.mvkkvar.hu/koztaurusz/kozt/index.phtml](http://www.mvkkvar.hu/koztaurusz/kozt/index.phtml)

## 2. A meglévő tárgyszó rendszer „köztauruszosítása”

Mint az előzőekből is kiderült, a könyvtár tárgyszórendszerének nagyobb része ugyan a KÖZTAURUSZ-ból származott, lefedettsége azonban két okból sem lehetett teljes:

- nem tartalmazta a KÖZTAURUSZ kapcsolatrendszerét
- nem tartalmazta az összes KÖZTAURUSZ szócikket

Az SR-LIB – HUNTEKA konverzió során ez a struktúrával rendelkező saját tárgyszórendszer a HUNMARC szerinti meghatározott helyére került – azaz a 650-es tárgyszóba illetve a mögötte felépülő 150-es besorolási állományba. Az SR-LIB - HUNTEKA áttöltés pillanatában ez a rendszer 6672 db besorolási rekordot tartalmazott, melyek közel 100 lásd kifejezéssel és több, mint 7000 lásd még kapcsolattal rendelkeztek. Ezt a képet tovább tarkította a már említett SR-LIB-es „hagyaték”, mely a korábban külön adatbázisban tárolt helytörténeti cikkgyűjtemény tárgyszórendszerének különállóságát kívánta megőrizni. A saját tárgyszórendszer összességében több mint 115 ezer bibliográfiai rekordra hivatkozott (64576 cikkek, 51409 egyéb dokumentumtípus).

Célunk nem a KÖZTAURUSZ tárgyszó rekordjaink mögötti értéklistaként (besorolási rekordok háttér állományaként) való alkalmazása volt, betöltésével azt kívántuk elérni, hogy az állományunk a KÖZTAURUSZ rendszerébe közvetlenül bekapcsolódjon. Mindeközben azonban még ideiglenesen sem veszíthettük el az állomány nagyobb részének tartalmi feltártságából eredő előnyöket, így a saját tárgyszavak kikoptatásának, és az állomány KÖZTAURUSZ kifejezéseivel való indexelésének a lehető legrövidebb idő alatt kell végbemennie. Ez – a számokat figyelembe véve (115 ezer tárgyszavazott bibliográfiai rekord, közel 40 ezer KÖZTAURUSZ szócikk) – nem történhetett tisztán manuálisan. Az állomány „köztauruszosításhoz” SQL eszközöket kellett igénybe vennünk.

### *Az integráció lépései*

A két tárgyszórendszer (a saját és a KÖZTAURUSZ) az összehasonlíthatóság érdekében nem keveredhetett, vagyis a KÖZTAURUSZ nem kerülhetett a szabvány szerinti helyére – a 150-es besorolási rekordok közé. A KÖZTAURUSZ így ideiglenesen 154-es besorolási állományként került a rendszerbe. Állományunk KÖZTAURUSZ-hoz rendelése 5 (6) lépésben írható le, jelenleg az időigényes 4. lépésnél tartunk.

1. Meg kellett vizsgálni, hogy vannak-e olyan tárgyszavak, melyek pontosan megfeleltethetők valamely KÖZTAURUSZ-os kifejezésnek.
2. Egyezés esetén az eddigi saját tárgyszórendszerre mutató bibliográfiai rekordok tárgyszó hivatkozását át kell irányítani a KÖZTAURUSZ-t tartalmazó besorolási állományra.
3. A 650-ből törölni kell a beazonosított tárgyszavakat, hogy a rekordok ne legyenek duplán tárgyszavazva.
4. El kell végezni a nem algoritmizálható egyezések globális kézi javítását (karbantartó eszközzel).
5. A régi tárgyszórendszer teljes „kikoptatását” követően a KÖZTAURUSZ-t szabványos helyére (150-650) kell átmozgatni
6. A további tárgyszavazásnak a KÖZTAURUSZ deskriptor elemeivel kell történnie

A két tárgyszórendszer összevetését megnehezítette, hogy az OSZK-tól kapott KÖZTAURUSZ első változataiban a deskriptorok többnyire (de nem mindig) nagybetűs írásmódban szerepeltek, ezért többször is újra kellett tölteni. A kifejezések összehasonlítása

először a normalizált értékek alapján történt, „megölve ezzel az ékezeteket”, így érdekes párok illetve rekord kapcsolatok jöttek létre (ÁGY – AGY: „A rejtőzködő elme : egy fiziológus széljegyzetei / Ádám György”, vagy BOR – BÖR: „Bor és mítosz : a bor szellemi képének forrásai”). A hiba hamar kiderült, javításához meg kellett ismételni a betöltést.

Jelsorozat szerinti összehasonlítással 4340 tárgyszó volt megfeleltethető valamely KÖZTAURUSZ-os kifejezésnek, ami közel 50 ezer (48234) bibliográfiai rekordot érintett. 2332 olyan régi tárgyszó volt, melyet nem lehetett automatikusan hozzárendelni egyetlen KÖZTAURUSZ tárgyszóhoz sem. Ebben benne voltak a már fentebb említett cikk-tárgyszavak is (1391 db rendszó, 64576 érintett cikkek) – melyek elkülönültségét a könyvtár továbbra is fenn kívánta tartani. Az összevetés hatékonyságáról tehát el lehet mondani, hogy több mint 90 %-os<sup>3</sup> eredményességgel dolgozott. A maradék közel 1000 tárgyszó KÖZTAURUSZ-szal történő megfeleltetéséhez fejlesztés alatt van egy olyan funkció, ami lehetővé teszi, hogy különböző besorolási állományhoz kapcsolódó bibliográfiai rekordok is gyorsan egységesíthetők legyenek, azaz egy-egy saját tárgyszóhoz tartozó összes bibliográfiai rekord globálisan kerülhessen át az egyezőnek ítélt KÖZTAURUSZ szócikk alá.

Feltételezhetően lesznek olyan saját tárgyszavak, melyeket a KÖZTAURUSZ mellett is szeretnénk megőrizni, ezeknek a szabad tárgyszavak közé kell kerülniük.

#### *A „köztauruszosítás” problémái*

A fent ismertetett lépések természetesen nem voltak problémamentesek.

1. Még az összehasonlítás során merült fel a rendszó szerint egyezőnek ítélt, ám a saját tárgyszórendszer nem megfeleltethető tagjára mutató kapcsolattal rendelkező rekordok problémája. Konkrétan 338 db olyan saját tárgyszó volt, ami olyan régi tárgyszó rekordhoz kapcsolódott, ami nem volt megtalálható a KÖZTAURUSZ-ban. A kapcsolataik miatt ezeket nem töröltük, hanem rejtetté tettük. Ennek eredményeként a tárgyszó böngészésben nem látszanak, kapcsoltuk azonban még látszik a beazonosítatlan tárgyszó rekordoknál.
2. A HUNTEKA OPAC felülete több szempont szerinti keresést, illetve böngészést tesz lehetővé. A fenti számokat figyelembe véve a közel 40 ezer KÖZTAURUSZ szócikk közül nagyjából 33 ezerhez nem kapcsolódik bibliográfiai rekord. Ez a keresésben nem okoz gondot, mivel a különféle tárgyszavak közös index alapján kereshetők, és az olvasó nem szembesül a „készlettel”, vagyis a lehetséges tárgyszavak listájával. A böngészésben azonban meglehetősen zavaró 30 ezer „üres” kifejezés, nem beszélve a hozzájuk kapcsolódó hasonló mennyiségű „lásd” kifejezésről, melyek ugyancsak böngészhetők. A használt tárgyszavak kiszűrése és böngészésben való megjelenítése tehát mindenképpen problémás. Nem tekinthető kielégítőnek az a megoldás, miszerint a böngésző indexben minden kifejezés mellett szerepeltetjük az adott tárgyszó előfordulásait, vagyis a hivatkozó bibliográfiai tételek számát, mert a jelenlegi kihasználtságot figyelembe véve oldalakat kellene böngésznie az olvasóknak, míg hivatkozással rendelkező kifejezésekre találnak<sup>4</sup>. Ha egy adott KÖZTAURUSZ tárgyszóhoz csak a hozzá tartozó struktúrát emelnénk be a böngészésbe, még mindig szép számmal lennének olyan tárgyszavaink, amik automatikusan átemelődtek, de bibliográfiai rekord nem tartozik hozzájuk. Ha azonban meg tudjuk oldani, hogy a tárgyszó indexben csak azok a tárgyszavak jelenjenek meg, melyek rendelkeznek

<sup>3</sup> A százalékos érték az eredeti és a jelenlegi adatbázisban szereplő tárgyszó / KÖZTAURUSZ által hivatkozott rekordokból következtethető vissza.

<sup>4</sup> Gräff Zoltán-Téglás György-Thék György: A magyarországi könyvtárügyi fejlesztések és a Textlib = TMT 52. évf. 2005. 9.sz. 388

bibliográfiai rekordkapcsolattal, vagyis „élő” tárgyszavak, akkor már csak az „élő” tárgyszó alatti kifejezések között lesznek „üres” tárgyszavak (azaz egy élő KÖZTAURUSZ tárgyszóra kattintva láthatóak lesznek a hozzá kapcsolódó további, bibliográfiai kapcsolat nélküli KÖZTAURUSZ kifejezések). Ezek megléte viszont szükséges, különben kiszakítjuk a tárgyszavakat az eredeti hierarchiájukból. Az „élő” rekord böngészés hátterében egy az összes besorolási rekordot tartalmazó fiktív táblára tett üres böngésző index áll, amit egy SQL-script kétóránként feltölt a hivatkozott besorolási rekordokkal.

3. Végül, de nem utolsó sorban, gondoskodnunk kellett arról, hogy az export / import során a tárgyszavak a megfelelő helyre kerüljenek. Egyfelől tehát biztosítani kellett, hogy az ideiglenesen szabálytalan HUNMARC mezőben helyet foglaló (654) KÖZTAURUSZ kifejezések a HUNMARC /USMARC export során (pl. MOKKA feltöltés) a helyükre (650) kerüljenek. Másfelől viszont felül kellett vizsgálni az import (HUNMARC, Z39.50) folyamatát. A KÖZTAURUSZ integritásának megőrzése érdekében az importtal bekerülő tárgyszavak nem kerülnek be a KÖZTAURUSZ struktúrájába, hanem egyelőre továbbra is a 150-es saját tárgyszó rendszer számára használt besorolási állományba kerülnek. Az egységesítés végeztével, a saját tárgyszavak teljes kikoptatását követően, ezzel kapcsolatosan további döntésre lesz szükség. (653? 654?).

### **3. Kézzelfogható eredmény – példák alapján**

Az előadás során néhány konkrét példával is szeretnénk illusztrálni a KÖZTAURUSZ – Huntékába történő integrálásának eredményességét – mind a feldolgozó felületen – mind pedig katalógusunk OPAC felületén.<sup>5</sup>

### **4. Együttműködés a könyvtárosok és a szoftverfejlesztők között**

Az együttműködés nem volt zökkenőmentes. A HUNTEKA fejlesztőinek jól kimunkált, kommunikációs-, és hibabejelentő rendszere (BugReport) van a napi kapcsolattartásra, a konzultációra, a feladategyeztetésre. Az írásos kommunikáció azonban nem könnyű. Minden jó szándék ellenére is sokszor éreztük úgy, mintha nem egy nyelvet beszélénk. Nekünk, könyvtárosoknak is meg kell tanulnunk az informatikusok fejével gondolkodni, igényeinket egyértelműen megfogalmazni.

Abban bízom, hogy amikor ez az előadás elhangzik, azt tudjuk majd bemutatni, hogy létezik Magyarországon egy könyvtári rendszer, a HUNTEKA, amely a könyvtárosok és a szoftverfejlesztők szoros együttműködésének köszönhetően a szoftverrel dolgozó könyvtárosok és a használók számára is biztosítja az információ visszakeresés gyors és megbízható módját. Igennel válaszolhatunk a címben feltett kérdésre. Lehet következetes tartalmi feltárás szoftvercsere mellett is, de ehhez olyan együttműködés kell, ami a HUNTEKA fejlesztők és a Kisfaludy Károly Megyei Könyvtár munkatársai között kialakult.<sup>6</sup>

---

<sup>5</sup> <http://infoserver.kkmk.hu/katalogus/index.jsp>

<sup>6</sup> Ezúton fejezzük ki külön köszönetünket az Icron Kft. munkatársainak, Szabó Mihálynak és Sóti Róbertnek, akik adatbázisunk „köztauruszosításában” közreműködtek.

***Felhasznált irodalom:***

1. Tezauruszkezelő programok és a RELEX / Ungváry Rudolf. In: Tudományos és műszaki tájékoztatás, 2001. 1. sz. p. 3-15.
2. Fogalmi választás és nyelvi vonzás a tárgyszavas információkereső nyelvi modellekben / Prókai Margit. In: Tudományos és műszaki tájékoztatás, 2005. 5. sz. p. 218-234.
3. Tezaurusz a felhasználói felületen : az optimális tezauruszmegjelenítés problémái. / Ungváry Rudolf. In: Tudományos és műszaki tájékoztatás, 2001. 3. sz. p. 99-108.
4. A magyarországi könyvtárügyi fejlesztések és a Textlib / Gräff Zoltán-Téglás György-Thék György. In: Tudományos és műszaki tájékoztatás, 2005. 9.sz. p. 392-398.