

Gáncsné Nagy Erzsébet – Lengyel Monika

Következetes tartalmi feltárás szoftvercsere mellett?*

A KÖZTAURUSZ használata a győri Kisfaludy Károly Megyei Könyvtárban

A győri Kisfaludy Károly Megyei Könyvtárban (KKMK) 1997 októberétől a dokumentumok feldolgozását, feltárását számítógépes könyvtári rendszer segítségével végezzük. Ettől kezdve a tartalmi feltárás ETO jelzetek 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 most nem célunk.

A tárgyszavazás kronológiája a győri Kisfaludy Károly Megyei Könyvtárban

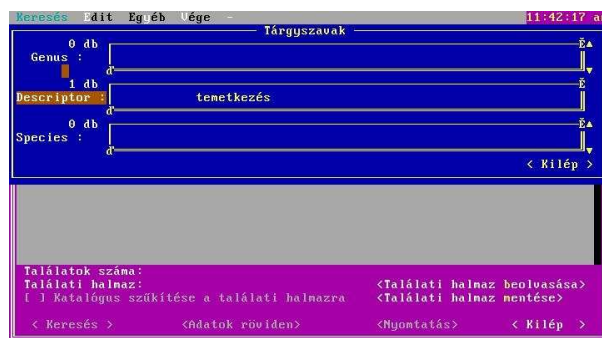
1997–1999

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

A tárgyszavak állománya a közösen kialakított elvek alapján a tartalmi feltárás közben folyamatosan bővült. Nem volt tehát előre elkészített listánk. A következetességet az biztosította, hogy a tárgyszavazás 1999 márciusáig egy könyvtáros feladata volt. Ez alatt az időszak alatt 6467 dokumentum tárgyszavazását végeztük el. A tárgyszavakat részben magunk állapítottuk meg, részben az *Új Könyvek* tárgyszavait követtük, illetve az ETO jelzetek természetes nyelvű kifejezésén alapultak.

Az általunk használt szoftver az alá/fölé rendeltséget és a szinonimakapcsolatot tudta kezelni (1. ábra). A szoftvernek ezt a lehetőségét ki is használtuk. Ugyancsak a szoftver befolyásolta azt a döntésünket, hogy négyféle tárgyszórendszert (szabad, kiról-miről, mikor, földrajzi) építettünk (2. ábra). A tárgyszó alapján végzett keresés ugyanis könnyebb és gyorsabb volt, mint ha az ún. szöve-

ges, „leíró” adatmezőkben kerestünk volna, továbbá itt nyílt lehetőség összetett keresésre is.

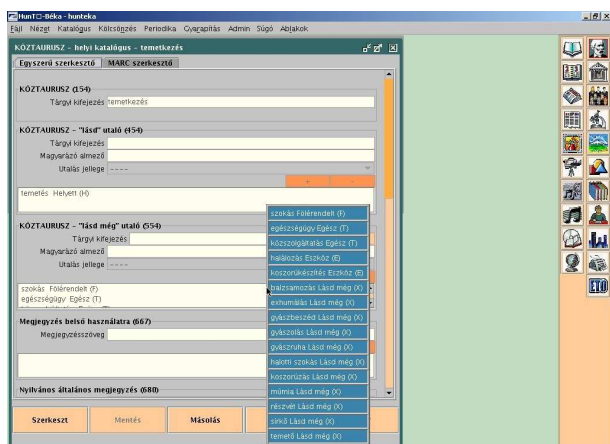


1. ábra Tárgyszóstruktúra az SRLIB-ben



2. ábra Az SRLIB négyféle tárgyszórendszere

* A 2006. április 19–21. között Miskolcon megrendezett *Networkshop* konferencián elhangzott előadás szerkesztett változata.



4. ábra Egy KÖZTAURUSZ deszkriptor kapcsolatrendszere a Huntékában

A meglévő tárgyszórendszer „köztauruszosítása”

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

- nem tartalmazta a KÖZTAURUSZ kapcsolatszerkeztét,
- nem tartalmazta a KÖZTAURUSZ összes szócikkét.

Az SRLIB – Hunteka konverzió során ez a struktúrával rendelkező saját tárgyszórendszer a HUNMARC szerinti meghatározott helyére került, azaz a 650-es tárgyszó (tárgyszórendszerből) adatmezőbe, illetve a mögötte felépülő 150-es tárgyszó-besorolásiadat adatmezőbe. Az SRLIB – Hunteka áttöltés pillanatában ez a rendszer 6672 besorolási rekordot tartalmazott, amelyekhez közel száz „lásd” utalás, és több mint 7000 „lásd még” kapcsolat tartozott. A képet tovább tarkította a már említett SRLIB-es „hagyaték”, amely a korábban külön adatbázisban tárolt helytörténeti cikkgyűjtemény tárgyszórendszerének különállóságát kívánta megőrizni. Saját tárgyszavaink összességükben több mint 115 ezer bibliográfiai rekordhoz kapcsolódtak (64 576 cikkekord, 51 409 egyéb dokumentumtípus).

Célunk most már nem az volt, hogy a KÖZTAU-RUSZ a könyvtári rendszerünkben kezelt tárgyszó-rekordjainktól elkülönült, csak manuálisan kapcsolódó háttérrendszerként álljon rendelkezésre (besorolási rekordok háttérállományaként); betöltésével azt kívántuk elérni, hogy tárgyszórekordjaink a

KÖZTAURUSZ rendszeréhez közvetlenül kapcsolódjanak. Mindeközben még ideiglenesen sem vehetjük el az állomány nagyobb részének tartalmi feltártságából eredő előnyöket, ezért a saját tárgyszavak felváltásának, és az állománynak a 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” SOL eszközöket kellett igénybe vennünk.

Az integráció lépései

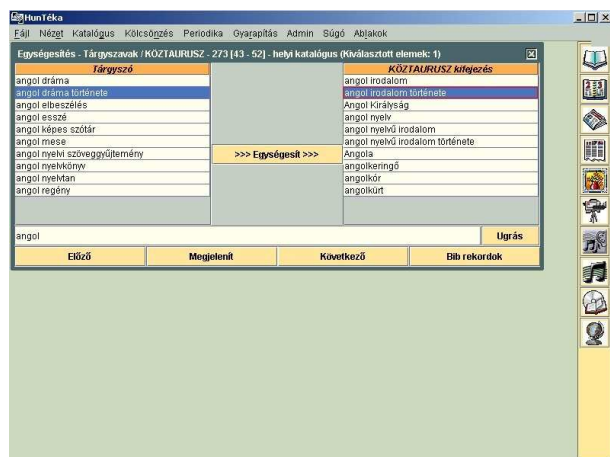
A két tárgyszórendszer (a saját és a KÖZTAURUSZ) az összehasonlíthatóság érdekében nem keveredhetett, vagyis a KÖZTAURUSZ nem kerülhetett a szabvány szerinti helyére – a 150-es adatmező besorolási rekordjai közé. A KÖZTAURUSZ így ideiglenesen a 154-es adatmezőre épülő besorolási állományként került a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

1. Meg kellett vizsgálni, hogy vannak-e olyan tárgyszavak, melyek pontosan megfeleltethetők a KÖZTAURUSZ valamelyik lexikai egységével (deszkriptorával vagy nemdeszkriptorával).
2. Egyezés esetén az eddigi saját tárgyszóra mutató bibliográfiai rekordok tárgyszóhivatkozását át kell irányítani a KÖZTAURUSZ-t tartalmazó besorolási állományban szereplő lexikai egységre.
3. A 650-es adatmezőből 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 (a 150-es, illetve a 650-es adatmezőkbe) kell átmozgatni.
6. A további tárgyszavazásnak a KÖZTAURUSZ-nak immár a 150-es/650-es adatmezőkben szereplő lexikai egységeivel kell történnie.

A két tárgyszórendszer összevetését megnehezítette, hogy az *Országos Széchényi Könyvtár*tól kapott KÖZTAURUSZ korábbi változataiban a deszkriptorok többnyire (de nem mindig) nagybetűs írásmódban szerepeltek, ezért azt többször is újra be kellett tölteni. A kifejezések összehasonlí-

tása először a normalizált értékek alapján történt, „megölve ezzel az ékezeteket”, így érdekes párok, illetve rekordkapcsolatok 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.

Karaktersorozat szerinti összehasonlítással 4340 tárgyszó volt megfeleltethető a KÖZTAURUSZ valamelyik lexikai egységével, ami közel 50 ezer (48 234) bibliográfiai rekordot érintett. 2332 olyan régi tárgyszó volt, amelyet nem lehetett automatikusan hozzárendelni egyetlen lexikai egységhez sem a KÖZTAURUSZ-ban. Ebben benne voltak a cikkekhez kapcsolódó, korábban már említett tárgyszavak is (1391 rendsző, 64 576 érintett cikk-rekord), amelyek elkülönültségét a könyvtár továbbra is fenn kívánta tartani. Az összevetésről tehát el lehet mondani, hogy több mint 90%-os volt, ami meglehetősen hatékony eredmény [3]. A maradék közel 1000 tárgyszó KÖZTAURUSZ-szal történő megfeleltetéséhez olyan funkciót alakítottunk ki, amely 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 a KÖZTAURUSZ egyezőnek ítélt lexikai egységéhez (5. ábra).



5. ábra HUNTEKA Tárgyszó – KÖZTAURUSZ egységesítését szolgáló felülete

Feltételezhetően lesznek olyan saját tárgyszavak, melyeket a KÖZTAURUSZ mellett is szeretnének megőrizni; ezeknek a szabad szavas tárgyszavak (ún. „szabad tárgyszavak”) közé kell kerülniük.

A „köztauruszosítás” problémái

Az eddig ismertetett lépések nem voltak problémamentesek. Még az összehasonlítás során merült fel a rendsző szerint egyezőnek ítélt, ám a saját tárgyszórendszer nem megfeleltethető tagjára mutató kapcsolattal rendelkező rekordok problémája. Más szóval olyan rekordok tárgyszavairól volt szó, mely tárgyszavak nem szerepeltek a KÖZTAURUSZ-ban lexikai egységként. 338 ilyen régi tárgyszó volt. Mivel ezek a tárgyszavak a dokumentumrekordokban szerepeltek, a dokumentum tartalmára vonatkozóan továbbra is értékes információt képviseltek, ezért ezeket nem töröltük, hanem rejtetté tettük. Ennek eredményeként az ilyen tárgyszavak böngészéskor nem jelennek meg a mutatóban, de a megjelenített dokumentumrekordban ott láthatók.

A Huntéka OPAC felülete több szempont szerinti keresést, illetve böngészést tesz lehetővé. A KÖZTAURUSZ közel 40 ezer tezauszszócikke közül pillanatnyilag nagyjából 33 ezerhez nem kapcsolódik bibliográfiai rekord. Ez a kereséskor nem okoz gondot, mivel a különféle tárgyszavak közös mutató alapján kereshetők, és az olvasó nem szembesül a teljes tárgyszókészlettel, vagyis az összes lehetséges tárgyszó listájával. A böngészésben azonban meglehetősen zavaró a 30 ezer „üres” kifejezés, nem beszélve a hozzájuk kapcsolódó hasonló mennyiségű „lásd” utalásról, amelyek alapján ugyancsak böngészni lehet. A használt tárgyszavak kiszűrése és böngészésben való megjelenítése tehát mindenképpen problémát jelent. Nem tekinthető kielégítőnek az a megoldás, miszerint a böngésző mutatóban minden kifejezés mellett szerepeltetjük az adott tárgyszó előfordulásait, vagyis a hivatkozott (talált) bibliográfiai tételek számát, mert a jelenlegi kihasználtságot figyelembe véve oldalakat kellene böngésznie az olvasónak, míg találattal rendelkező kifejezésekre lel [4]. Ha a KÖZTAURUSZ adott lexikai egységéhez csak a hozzá tartozó struktúrát (más szóval a tezauszszócikét) emelnénk be a böngészésbe, még mindig szép számmal lennének olyan tárgyszavaink, amelyek automatikusan átemelődtek, de bibliográfiai rekord nem tartozik hozzájuk. Ha azonban meg tudjuk oldani, hogy a tárgyszómutatóban csak azok a tárgyszavak jelenjenek meg, melyeknek van bibliográfiai rekordkapcsolatuk, vagyis „élő” tárgyszavak, akkor már csak az „élő” tárgyszó tezauszszócikkében szereplő kifejezések között lesznek „üres” (találatok nélküli) tárgyszavak (azaz a KÖZTAURUSZ egy élő lexikai egységére kattintva láthatók lesznek mind a hozzá kapcsolódó to-

vábbi, találattal rendelkező, mind pedig a találat nélküli kifejezések). Ezeknek a találatok nélküli kifejezéseknek a megléte viszont szükséges, különben kiszakítjuk a tárgyszavakat az eredeti, a KÖZTAURUSZ-ban megadott hierarchiájukból. Az „élő” tárgyszavak közötti böngészés hátterében az összes besorolási rekordot tartalmazó fiktív táblára tett üres böngészőmutató áll, amelyet egy SQL-script kétóránként feltölt a hivatkozott besorolási rekordokkal.

Végül, de nem utolsósorban, gondoskodnunk kellett arról, hogy az export/import során a tárgyszavak a megfelelő helyre kerüljenek. Egyfelől tehát meg kellett oldani, hogy a KÖZTAURUSZ ideiglenesen szabálytalan, 654-es HUNMARC adatmezőben helyet foglaló lexikai egységei a HUNMARC/USMARC export során (pl. MOKKA-feltöltés) a helyükre, a 650-es adatmezőbe kerüljenek. Másfelől viszont felül kellett vizsgálni az import (HUNMARC, Z39.50) folyamatát. A KÖZTAURUSZ integritásának megőrzése érdekében az importtal bejuto tárgyszavak nem kerülnek be a KÖZTAURUSZ struktúrájába, hanem egyelőre továbbra is a saját tárgyszórendszer számára használt 150-es adatmezőbe (besorolási állomány) kerülnek. Az egységesítés végeztével, a saját tárgyszavak teljes kikoptatását követően, ezzel kapcsolatosan további döntésre lesz szükség (esetleg sor kerül szabadon használható tárgyszavak alkalmazására).

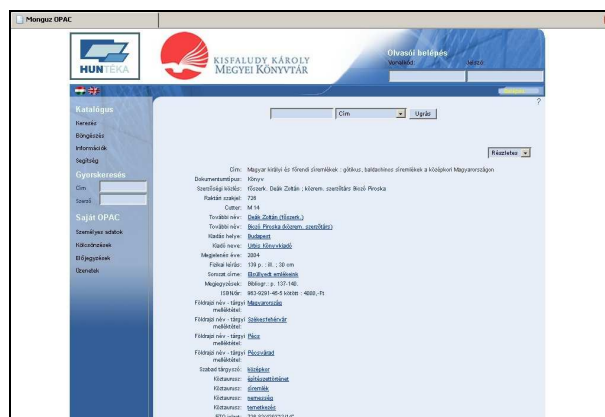
Az olvasó szemszögéből

A könyvtár OPAC-felületén [5] két példával illusztráljuk a KÖZTAURUSZ Huntékába integrálásának eredményességét.

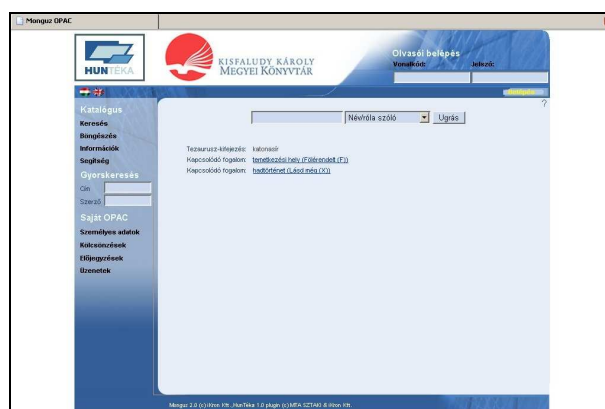


6. ábra Egy KÖZTAURUSZ-deszkriptor
kapcsolatrendszere az OPAC-felületen

Első példánkban olyan lexikai egység teljes, a RELEX által kezelt tezauruszcikkét láthatjuk, amelyet sikerült megfeleltetni létező kifejezéssel az adatbázisban (6. ábra). Az olvasó számára is jól látható a tárgyszó hierarchiában elfoglalt helye, amelyen belül a csatolók (linkek) segítségével bármelyik kapcsolódó kifejezés irányában továbbléphet. A 7. ábrán látható rekordot kiválasztva pedig a KÖZTAURUSZ lexikai egységeinek és régi tárgyszavaink „együttélésére” láthatunk egy példát.

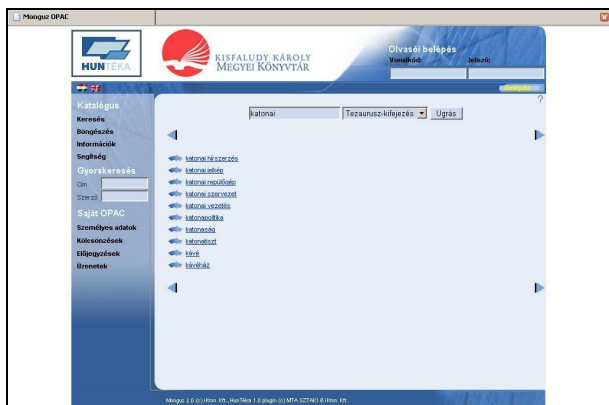


**7. ábra Maradvány tárgyszavak
és KÖZTAURUSZ-kifejezések együttese
egy bibliográfiai rekordban**



8. ábra **Bibliográfiai kapcsolattal nem rendelkező utaló az OPAC-ban**

Az „élő” tárgyszavak közötti böngészés szemléltetéséhez az előző példánál maradva a „temetkezés” deszkriptortól a „temető”, „temetkezési hely” kapcsolódó fogalmakon keresztül el lehet ugyan jutni a „katonasír” kifejezésig, de ahhoz nem tartozik bibliográfiai rekord (8. ábra). Ugyanez a deszkriptor a böngészésből közvetlenül nem található meg (9. ábra).



9. ábra „Élőrekord-böngészés” az OPAC-ban
– csak a bibliográfiai a kapcsolattal rendelkező
kifejezések böngészhetők

Együttműködés a könyvtárosok és a szoftverfejlesztők között

Az együttműködés nem volt zökkenőmentes. A Huntéka 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 beszélnének. Nekünk, könyvtárosoknak is meg kell tanulnunk az informatikusok fejével gondolkodni, igényeinket egyértelműen megfogalmazni.

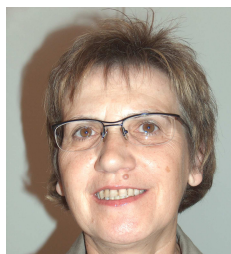
Cikkünkben azt kívántuk bemutatni, hogy létezik Magyarországon könyvtári rendszer, a Huntéka, amely a könyvtárosok és a szoftverfejlesztők szoros együttműködésének köszönhetően a szoftverrel dolgozó könyvtárosok és a használók számára is lehetővé teszi a gyors és megbízható, ugyanakkor korszerű, tezuauruszt alkalmazó információkeresést. Igennel válaszolhatunk a címben feltett kérdésre: lehet következetes tartalmi feltárás szoftvercsere mellett is, de ehhez olyan együttműködés kell, amely a Huntéka fejlesztői és a Kisfaludy Károly Megyei Könyvtár munkatársai között kialakult [6].

Gáncsné Nagy E.–Lengyel M.: Következetes tartalmi...

Irodalom és jegyzetek

- [1] PRÓKAI Margit: Fogalmi választás és nyelvi vonzás a tárgyszavas információkereső nyelvi modellekben. = Tudományos és Műszaki Tájékoztatás, 52. köt. 5. sz. 2005. p. 218–234.
- [2] www.mvkkvar.hu/koztaurusz/kozt/index.phtml
- [3] A százalékos érték az eredeti és a jelenlegi adatbázisban szereplő tárgyszó / KÖZTAURUSZ által hivatkozott rekordokból következtethető vissza.
- [4] GRÄFF Zoltán–TÉGLÁS György–THÉK György: A magyarországi könyvtárügyi fejlesztések és a Textlib. = Tudományos és Műszaki Tájékoztatás, 52. köt. 9. sz. 2005. p. 392–398.
- [5] <http://infoserver.kkmm.hu/katalogus/index.jsp>
- [6] Ezúton fejezzük ki külön köszönetünket az Ikron Kft. munkatársainak, Szabó Mihálynak és Sóti Róbertnek, akik adatbázisunk „köztauruszosításában” közreműködtek.
- [7] UNGVÁRY Rudolf: Tezaurusz a felhasználói felületen: az optimális tezauruszmegjelenítés problémái. = Tudományos és Műszaki Tájékoztatás, 48. köt. 3. sz. 2001. p. 99–108.
- [8] UNGVÁRY Rudolf: Tezauruszkezelő programok és a RELEX. = Tudományos és Műszaki Tájékoztatás, 48. köt. 1. sz. 2001. p. 3–15.

Beérkezett: 2006. VI. 7-én.



Gáncsné Nagy Erzsébet

a győri Kisfaludy Károly
Megyei Könyvtár
igazgatóhelyettese.
E-mail: e.gancs@mail.kkmm.hu



Lengyel Monika

az MTA SZTAKI Könyvtár-
informatikai csoportjának
tudományos munkatársa,
könyvtári informatikai
szaktanácsadó,
konverziós szakértő.
E-mail: Imoni@sztaki.hu