

Adatbázis transzformációalapú kereshetőségjavítása: ETO-kereshetőség a BME OMIKK Aleph rendszerében

Az informatika fejlődése a dokumentumrekordokat, illetve a szöveges állományokat tartalmazó adatbázisok egyre sokoldalúbb kereshetőségét eredményezi. A természetes nyelvű kifejezéseken alapuló információkereső nyelvek jól használható megoldást nyújtanak. Kérdés azonban, hogy teljes értékű alternatívát nyújtanak-e, vagy csupán kiegészítik a mesterséges nyelven alapuló információkereső nyelvek használatát az osztályozásban és a keresésben? A komplex kapcsolatok (pl. az ETO-ban a viszonyítás) keresőtámogatása pezsgő kutatási terület éppúgy, mint az ontológiaalapú kereséstámogatás vagy a természetes nyelvű gépi fordítók. Jelenleg hibrid rendszerekben kell gondolkodnunk! Az egyik legfontosabb kérdés az, hogyan fér meg egy keresőrendszerben a mesterséges és a természetes nyelveken alapuló osztályozás.

Bevezetés

A dokumentumrekordokat, illetve a szöveges állományokat tároló adatbázisok számítógépes kezelésének alapvető feladata, hogy megoldja az online katalógusban tárolt információk pontos és gyors keresését. A gyorsaság egyik feltétele, hogy a keresést, és vele az információkereső nyelv használatát könnyen, rövid idő alatt lehessen megtanulni, hogy a használó egyszerűen fogalmazhassa meg a keresőkérdést (állíthassa össze az ún. keresőképet vagy -profil). A gyorsaság másik aspektusa többnyire kimondatlan elvárás: a találati halmazt gyorsan állítsa elő a számítógép. Az adatbáziskezelő-rendszer a keresőképet hasonlítja össze az ún. dokumentumképpel vagy -profilal, amely az egy dokumentumrekordhoz kapcsolt ismérvek összessége. Ezek az ismérvek származhatnak valamely információkereső nyelv szavaiból, de lehetnek a dokumentum vagy a dokumentumrekord szövegének szavai is (az ún. szövegszavak). A keresés hatékonyságának értékelését fejezi ki a relevancia, mely a találati halmaz minőségi jellemzője: mekkora a keresőkérdés alapján a halmazba tartozó tételek aránya a találati halmaz méretéhez viszonyítva. A találati halmaz teljessége pedig azt fejezi ki, hogy mekkora a megtalált, releváns dokumentumok aránya az adatbázisban található, a keresőkifejezésre illeszkedő dokumentumképekhez viszonyítva (ideális esetben 100%).

A szabadon választott szavakon alapuló szabadszavas keresés térhódítása a tanulási folyamat rövidítésének irányába mutat, hiszen bárki saját szavaival végezheti a keresést. Nem biztos azonban, hogy ez a megközelítés önmagában kellően releváns és teljes találati halmazt eredményez (nem biztos ugyanis, hogy a választott keresőszó szövegszóként, és a használó által szándékolt értelemben szerepel). Az automatikus szemantikus és szerepviszony-fókuszált¹ támogatás, amely elengedhetetlen, hogy ez a megközelítés jól működjön, még fejlődőben van. A különböző nyelvek és nyelvváltozatok közötti jó minőségű gépi fordítás, amely szintén ennek a megközelítésnek a hatékonyságához járulhat hozzá, ugyancsak gyerekcipőben jár.

A keresés minősége ma úgy javítható, hogy természetes vagy mesterséges nyelven alapuló információkereső nyelvet (az utóbbit nevezik osztályozási rendszereknek is) alkalmazunk, és ezek szavai (mesterséges nyelven alapuló osztályozási rendszer esetében a jelzetei) a rekordok ismérvei. Az előbbiekhöz tartoznak a tárgyszavas rendszerek és a teauruszokban strukturált deskriptoros nyelvek. Az utóbbiak közül a legismertebb, Közép-Európában is alkalmazott információkereső nyelv vagy osztályozási rendszer az egyetemes tizedes osztályozás (ETO).

Cikkünkben az információkeresés hagyományosan két változata, a betűrendes mutatót használó ún. *böngészés* és a keresőképből közvetlenül találati halmazt előállító *keresés* közül az utóbbival foglalkozunk. Az ETO felépítésének rövid bemutatása után kitérünk a (kizárólag) természetes nyelvű szavakkal történő keresésre épülő rendszerekben tapasztalható problémákra. Ezután bemutatjuk az Aleph integrált könyvtári rendszer dokumentumrekordok kereshetőségét előkészítő és a keresőkérdeket feldolgozó eljárását: azokat a kereteket (vagy korlátokat), amelyek között keresőrendszerünket meg kell valósítani, majd elemezzük a kereshetőségi követelményeket az Aleph lehetőségeinek tükrében, és ismertetünk egy megoldási lehetőséget. Ezt követően példákon keresztül elemezzük a megoldást, és rámutatunk annak ismert hiányosságaira. A cikk befejező részében összefoglaljuk az eredményeket.

Az ETO-jelzetrendszer bemutatása

Az ETO-jelzeteket két nagy csoportba sorolhatjuk. Megkülönböztethetünk ún. *egyszerű jelzeteket*, amelyek a főtáblázatokban szerepelnek (nevezik ezeket főtáblázati számoknak is), illetve *összetett jelzeteket* (nevezik ezeket összekapcsolt jelzeteknek is), amelyeket több egyszerű jelzet összekapcsolásával, illetve a főtáblázatban szereplő jelzetek és a közös alosztások összekapcsolásával hozhatunk létre. A *közös alosztások* három típusa különböztethető meg: az önálló általános alosztások, a nem önálló általános alosztások és a nem önálló speciális alosztások. Az *általános alosztások* bármely főtáblázati számhoz hozzáfűzhetők. A *nem önálló speciális alosztások* csak abban a táblázatban használhatók, amelyekben felsorolták őket, illetve jelzik felhasználhatóságukat. Attól függően, hogy melyik főtáblázatban használják, az egyes speciális alosztások mást-mást jelenthetnek.

Vegyük most sorra, hogy az ETO jelzetszerkesztési szabályai alapján hogyan képezzük az ETO-jelzeteket, mit jelentenek az egyes jelzetek, és milyen keresési igények merülnek fel, ha e jelzetek, illetve az összetett jelzeteket alkotó jelzetek szerint kell keresni az adatbázisban.

Egyszerű ETO-számok

Egyszerű fogalmaknak felelnek meg és a főtáblázatban található. A számrendszer tíz számjegyből épülnek fel (innen az osztályozási rendszer nevében a „tizedes” jelző), és a tizedes törtek elve

szerint tagoltak (azaz az egyszerű jelzeteket három számjegyenként pontokkal tagolják). Az így tagolt elemek egyes részei szerint nincs értelme a keresésnek.

Az összetett ETO-számok képzése

Az összetett ETO-számokat alapvetően három módon képezhetjük: főtáblázati számok összekapcsolásával, általánosan közös alosztások és speciális alosztások alkalmazásával. A következő szakaszokban példákkal illusztráljuk a lehetőségeket.

Főtáblázati számok összekapcsolása

- Két, tartalmilag egymástól független fogalom vagylagos egymás mellé rendelése, összekötése a + (plusz) jellel.
 - 669.35+669.715** Rézötvözetek és alumíniumötvözetek.
 - (436+439)** Ausztria és Magyarország: egynemű önállóan közös alosztások kapcsolata egymás között.
 - Online katalógusban a + jellel összekapcsolt jelzeteket elég külön-külön megadni a dokumentumképben, mert a **vagy** logikai művelettel a keresőképben összekapcsolhatók.
- A fogalmak egymást követő sorozatának egyik fogalomtól a másikig való mellérendelő összefoglalása a / (...-tól...-ig) jellel (ennek a neve kiterjesztés).
 - 531/534** Mechanika, beleértve a szilárd testek, folyadékok, légnemű testek mechanikáját és a mechanikai rezgések, hangtan témakörét is. Ily módon az **53**-as Fizika szakcsoporton belül egy közbenső fölérendelt fogalmat alkottunk az alosztályok sorában.
 - A keresés során a / jelet tartalmazó jelzeteket egységes egészként kell kezelni.
- Két fogalom közötti kapcsolat jelölése (mellérendelő viszonyítás) a : (kettőspont) segítségével.
 - 548.0** Általános kristálytan, illetve **548.0:53** Kristályfizika
 - (436:439)** Összehasonlítás vagy kapcsolat Ausztria és Magyarország között: egynemű önállóan közös alosztások kapcsolata.
 - Online katalógusban a : jellel összekapcsolt jelzeteket elég külön-külön megadni a dokumentumképben, mert az **és** logikai művelettel a keresőképben összekapcsolhatók.

Főtáblázati számok összekapcsolása önálló és nem önálló általánosan közös alosztásokkal

1. Nyelvi alosztások: az írásmű nyelvének jelölése az = (egyenlőség) jel és az azt követő számjegy segítségével.
 - a. =30 Német nyelvű
2. Formai alosztások: az írásmű formájának jelölése a (0) (zárójel nulla) jel segítségével.
 - a. (038) Szótár
 - b. 55(038) Geológiai szótár
3. Földrajzi alosztások: egy fogalom földrajzi (térbeli) vonatkozásának jelölése a (1/9) segítségével.
 - a. (439) Magyarország
 - b. 55(439) Magyarország geológiája
4. Etnikai (népi) alosztások: népek, népfajták jelölése a (=) (zárójel egyenlő) jel segítségével.
 - a. (=30) Németek
 - b. 392.5(=30) Házassági szokások a németeknél
 - c. 392.5(439=30) Házassági szokások a magyarországi németeknél
5. Idő szerinti alosztások: a fogalom időbeli vonatkozásának jelölése az "... " (idézőjel) segítségével.
 - a. "196" Az 1960-as évek
 - b. 53"196" Fizika az 1960-as években
6. Szempont szerinti nem önálló alosztások²: annak a szempontnak jelölése, amelyből valamely fogalmat szemlélünk a .00 (pont nulla nulla) jel segítségével.
 - a. .001.573 Matematikai modell
 - b. 612.8.001.573 Az idegrendszer matematikai modellje
7. Általános ismérvek szerinti nem önálló alosztások. Valamely fogalommal kapcsolatos egyéb sajátosság, például az anyagok vagy a személyek jelölése a (-0) (kötőjel nulla) jel segítségével.
 - a. -033.6 Kerámia
 - 903.23-033.6 Kerámia edények a régészetben
 - b. -053.6 Fiatal személy
 - 316.62-053.6 Fiatalok társadalmi viselkedése

Főtáblázati számok összekapcsolása nem önálló speciális alosztásokkal

1. Kötőjeles, speciális alosztások (-1/-9)
 - a. 62-2 Gépek álló és mozgó részei
 - b. 82-2 Színdarab
2. Pont nullás speciális alosztások (.01/.09)
 - a. 53.08 Mérőműszerek. Mérési eljárások. Elméletük

- b. 82.08 Irodalmi műfajok technikája. Retorika. Stilisztika
3. Aposztrófós speciális alosztások (*)
 - a. 669.35 Rézötvezetek
 - b. 669.35*24 Réz-nikkel ötvezetek
4. Számvégződéses speciális alosztások³ (...1/...9)
 - a. 539.12 Elemi részecske
 - b. 539.12...14 Spin és momentum
 - c. 539.121.4 Elemi részecskék spinje és momentuma

Ugyanazok a speciális alosztások az egyes főtáblázati számokhoz kapcsolva mást-mást jelentenek. Emiatt a keresés során irreleváns találatok kerülhetnek a találati halmazba. Ez elkerülhetetlen, hiszen a keresőkérdés pontos értelmezése sem adható meg általában, csak konkrét főtáblázati számokhoz kapcsolva.

Külső forrásból (nem ETO-ból) származó alosztások: egyes személyek vagy fogalmak jelölése betűvel (A/Z) vagy (*) jellel és sorszámmal. Rendezése betűrendben vagy a sorszámok (nem ETO-számok) szerint történik.

1. Nevek:
 - 1(091)Arisztotelész Arisztotelészről szóló filozófiai mű
2. Betűk, betűcsoportok:
 - 061.1(100)ENSZ Egyesült Nemzetek Szervezete
3. Számok, sorszámok, rendszámok:
 - 546.42.027*90 a Stroncium 90-es tömegszámú izotópjá

Itt jegyezzük meg, hogy az összekötés (+) a logikai vagy kapcsolatnak, a kiterjesztés (/), a mellérendelő viszonyítás (:) és az alosztásokkal való összekapcsolás pedig logikai és kapcsolatnak felelnek meg.

A mesterséges nyelven alapuló jelzetek kereshetősége természetes nyelvre szabott rendszerekben

Könyvtárunkban, a BME OMIKK-ban, a dokumentumok túlnyomó többsége ETO alapján osztályozott. Ezért nemcsak az adatbázisban található egyszerű és összetett jelzeteknek, hanem az összetett jelzetek részjelzetei szerinti pontos keresésnek is nagy jelentősége van. Az ETO mesterséges nyelven alapuló jelzetrendszerében történő keresés a természetes nyelvben nem, vagy kisebb súllyal jelentkező problémákat vet fel. Ilyenek a

homonimakezelés vagy az írásjelek kezelésének kérdése.

A természetes nyelven alapuló információkereső nyelveket használó keresőrendszerek egyik közös tulajdonsága, hogy a nyelvben rejlő redundanciát kihasználva *igyekeznek* oly módon átalakítani a keresőkérdéseket, hogy azok információtartalmát ne csökkentsék. Tekintsük például az „adatbáziskezelő” kifejezést, és egy elterjedt átalakítási lépést, az ékezetmentesítést⁴. Az eredmény – adatbáziskezelő – továbbra is hordozza az eredeti kifejezés jelentését. Ezek az átalakítások sajnos mégsem mindig teljesítik az információ megtartásának kritériumát. A *kérés* kifejezés ékezetmentesítésére (*keres*) a keresőkérdést tartalmilag megváltoztatja: megkapjuk a *keres*, azaz az ékezetmentes verzióra adandó találatokat is, és fordítva. A *kérés-keres* homonimáppal illusztrált hiba hatását csökkenti, ha több szóból áll a keresőkérdés, mivel ezek a szavak szövegösszefüggést teremtve mintegy értelmezik a homonim keresőszót.

Akárcsak a természetes beszélt nyelv, az információkereső nyelv is sok homonimát tartalmaz. A mesterséges nyelven alapuló osztályozási rendszerekben elvileg ugyan nincsenek homonimák, mivel nincs két azonos alakú, de eltérő jelentésű egyszerű vagy összetett jelzet. Ha azonban az összetett jelzeteket alkotó általános, illetve közös alosztások szerint kell keresni, ez a probléma már megjelenik, mivel ezeknek a részjelzeteknek a számai homonimák lehetnek.

Az írásjeleket és más metanyelvi jeleket más-más módon kell kezelnünk a természetes és a mesterséges nyelven alapuló információkereső nyelveknél. A földrajzi nevek vagy történelmi személynevek helyes írásmódjában különös jelentősége van a kötőjeleknek vagy más központosításnak. A keresőkérdés megválaszolásakor a szigorú helyesírási szabályoktól azonban el kell tekintenünk annak érdekében, hogy jól használható keresőeszközt ajánlhassunk a felhasználóknak. Amikor a használó *Szent-Györgyi Albert* Nobel-díjas biokémikus neve alapján keres, a találatokat éppúgy meg kell adja a keresőrendszer, ha a hibás *Szentgyörgyi* vagy *Szent Györgyi* alakok valamelyikét használja.

Államalapító királyunk nevének szabványos megjelenítési formája⁵ *István (Magyarország: király), I., Szent*. A használó találatként elvárja a *Magyarország, a király, vagy az I István* keresőkifejezésekre éppúgy, mint az *I István Magyarország: király* formára az első királyunkról szóló dokumentumrekor-

dot. A kettőspont, a zárójelek, vagy a sorszámot jelző pont jelenléte vagy hiánya nem befolyásolhatja ezt a keresést.

Az ETO összetett jelzetein belül az alosztások részjelzeteit meghatározott (szóhatároló) jelek (kerek zárójelek, idézőjelek, pont nulla nulla stb.) értelmezik. Ezek a mesterséges nyelven belül afféle meta- vagy metanyelvi jelek (l. az ETO-jelzetrendszer bemutatása c. szakaszt). A részjelzetek környezetfüggetlen⁶ (azaz a metajeleket figyelmen kívül hagyó) elemzése félreértésekhez vezet, és ilyen felhasználásuk a kereséskor is félreértésekhez vezet, irreleváns találati halmazokat eredményez. Példaként tekintsük a 30 számot. Zárójel-egyenlő környezetben *népi alosztást* jelöl: (=30) – Németek. *Főtáblázati számként* a 30 jelentése „A társadalomtudományok elmélete, módszertana és módszerei”. A keresőkérdésben vagy a dokumentumképben szereplő összetett jelzeteknél a részjelzetek szóhatároló vagy kontextuális metajeleinek elhanyagolásakor az összes olyan dokumentumot megkapjuk találatként, amelyek az előbbi két kategória *valamelyikébe* beleesnek.

A példák illusztrálják, mennyire más megközelítést igényelnek a természetes nyelven és a mesterséges nyelven alapuló információkereső nyelvek az írásjelek kezelésében. Míg bizonyos jelek elhanyagolása az egyik esetben minőségjavulást (l. *Szent-Györgyi* és *Szentgyörgyi* azonos kezelése), addig a másik esetben komoly romlást eredményez (l. (=30) és 30 közötti különbségek).

Persze a két megközelítés elméletben remekül megfér egymás mellett: különböző feldolgozási eljárással a két eset kezelhető. Vizsgálataink mégis azt mutatják, hogy az adatbáziskezelő (katalogus)rendszerek többsége, köztük az Aleph sincs felkészítve arra, hogy az ETO-jelzetek részjelzetei szerint is keresni lehessen. Miből adódik a különbség elmélet és gyakorlat között? Egyes rendszerek nem képesek a két eset kellő mértékű szétválasztására, mások legalábbis nincsenek kellő gondossággal beállítva. Azokban a rendszerekben, ahol a két eset kezelése nem választható szét a szükséges mértékben, a keresési lehetőségek beállításaakor a gyorsabb tanulási görbével jellemezhető, természetes nyelven alapuló információkereső nyelvek javára kötnek komoly kompromisszumot a beállításokat végző szakemberek. A következő részben az Aleph példáján fejtjük ki részletesebben a szétválasztás hiányát, majd az azt követő szakaszban ismertetünk egy, az Alephre szabott megoldást.

Keresési alapelvek az Aleph rendszerben

A keresés, pontosabban a keresőkép és a dokumentumkép összehasonlításának (rendszerszintű) támogatására általánosan használt módszer az ún. indexek építése. Ez nem jelent mást, mint a dokumentumrekordok, illetve a szöveges állományok olyan átalakítását (transzformációját), hogy az indexek segítségével a különböző szempontok szerint végzett kereséseket gyorsan kiszolgálhassuk. Eredményül pedig releváns találatokat kell, hogy kapjon a használó.

Ebben a szakaszban az Aleph keresőrendszer elemeinek tárgyalását nem korlátozzuk az ETO-jelzetekkel végzett keresésre. Az általánosság oka az, hogy egyetlen keresőrendszerben kell együtt élnie az ETO-jelzetek és más szempontok szerint végzett kereséseknek. Rendszerben kell tehát gondolkoznunk ahhoz, hogy az ETO-jelzetekkel történő keresés megoldása ne befolyásolja számottevően a többi keresési lehetőséget.

Különböző indexek más-más adatkör alapján

Az indexek elkészítése a felhasználás célja és a forrásadatok egyedi azonosító jellege szerint más és más módon történhet. Alapvetően két módszert különböztetünk meg az Aleph 500 integrált könyvtári rendszerben.

Az első módszer elsődleges és másodlagos azonosítók, egységesített besorolási adatok egészére teszi lehetővé a keresést. Technikailag további két változatot különböztetünk meg. A művek elsődleges azonosítói közül azok egyedi azonosításra alkalmas, valamint az egyedi azonosításra nem alkalmas összetett kifejezések kezelése. Az egyedi azonosításra alkalmas elsődleges azonosítókra példa a monografikus művek ISBN kódja, folyóiratok ISSN száma, vagy kiadói jelzetek, mint például a jegyzetazonosító. Ezek részben hierarchikus felépítésűek⁷ is lehetnek, de csak az azonosító egészére kiadott pontos keresésnek⁸ van igazi jelentősége. Az effajta jelzetek kezelése éppen ezért egyszerű.

A második lehetőség a szóalapú indexelés. Ez egyrészt jelenti az adott szöveg minden egyes, közzel elválasztott (vagy egyéb kritériumok szerint önállóként kezelendő) szavai szerinti keresést, illetve a szabványosított (kötött szótárba foglalt tárgyszavas, deskriptoros) információkereső nyelv összetett szavainál az egyes összetevő szavak szerinti keresést. A dokumentumrekord termé-

zetes nyelvű, szöveges adatelemei, mint például a cím vagy a szerzőségi közlés is kezelhető ilyen módon, lehetővé téve a keresést kevés információ birtokában: a cím jellegzetes kifejezése, vagy a szerző néveleme szerint (extrém esetben ui. az adott névelemről – például nem európai neveknél – azt sem tudja a kereső, hogy vezetéknevről, keresztnévről stb. van-e szó). A szóalapú keresethez szükséges biztosítás az olyan információkereső-nyelvi szavak esetén nagy jelentőségű, ahol az alkotóelemek önállóan is hordoznak jelentést. Ez a helyzet például az ETO jelzeteinél.

A továbbiakban a szóalapú indexekkel kiszolgált keresőkérdésekről lesz szó, hiszen a vázolt lehetőségek közül cikkünk tárgya, az ETO részjelzetei szerinti kereshetőség oda illik a leginkább: ahogy elvárható, hogy a leíró adatok (pl. a dokumentum szövege, az annotáció, a cím, a megjegyzés) szemantikailag önálló (azaz önálló jelentést hordozó) elemei/szavai alapján keresni lehessen, úgy elvárható, hogy ugyanaz lehetséges legyen a természetes vagy a mesterséges nyelven alapuló másodlagos adatok (a tárgyszavak, deskriptorok, ETO-jelzetek) esetében is.

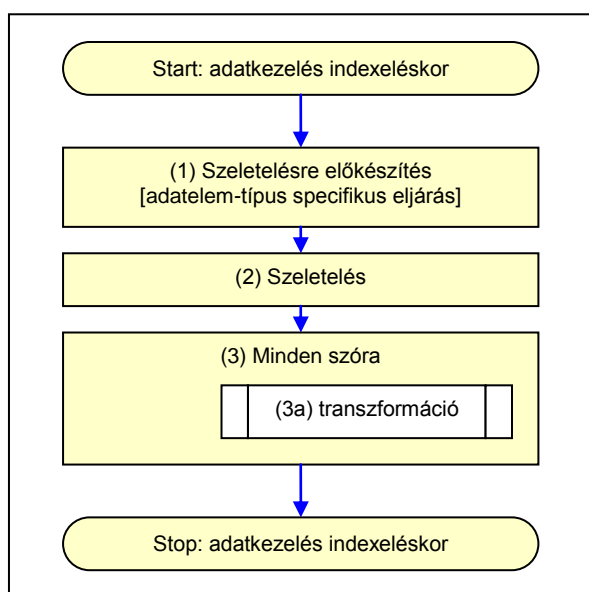
Szóalapú indexek építése

A szóalapú indexek definiálásának első lépéseként meghatározzuk, hogy a dokumentumrekord mely összetevőiből (mely ismérvekből) készüljenek az egyes indexek. Ez leginkább az alkalmazott katalogizálási szabványtól és konvencióktól függ. A HUNMARC adatcsere-formátumot tekintve cikkünk tárgya, az ETO-jelzet, a 080 mező \$a almezőjében szerepel, így ebből építjük az ETO szóalapú keresőindexünket. Az itt ismertetett folyamat természetesen más adatelemekre is érvényes.

Az 1. ábra mutatja az ismérvek (adatok) útját a dokumentumrekordtól a szóindex-bejegyzésig. A feldolgozási folyamat központi művelete, amikor az összetett jelzeten vagy kifejezésen/szövegen belül annak önálló jelentést hordozó részeit minősítik (1) és elkülönítik (2). Ezt a műveletet a továbbiakban szelektálásnak nevezzük.

A szóindex-bejegyzések elkészítésekor az indexelendő adatelem előfeldolgozáson megy keresztül, amelynek célja kettős. Az adatelem szavakra bontása mellett igény szerint az ékezetmentes vagy kis- és nagybetűtől független stb. keresést lehetővé tevő transzformáción is át kell esnie a szavaknak. A folyamat első lépésében (1) a rendszer tisztítja az indexelendő adatelemet. Az önálló je-

lentést hordozó elemeket elválasztó szóhatároló, vagy kontextuális metanyelvi jeleket egységesen szóközzé⁹ alakítja. A feldolgozásnak ebben a fázisában adatelem-típusonként (pl. ETO-jelzet, egységesített szerzői név stb.) más és más eljárás alkalmazható, ami kedvez a különféle esetek: a természetes vagy a különféle mesterséges nyelven alapuló ismérvek kezelésének. Az elválasztó jelek „felismerésekor” jelentkezik a rendszer egyik korláta: néhány specializált eljárástól (amelyek köre nem bővíthető) eltekintve karakterek feltétel nélküli törlése vagy szóközzé alakítása valósítható meg.



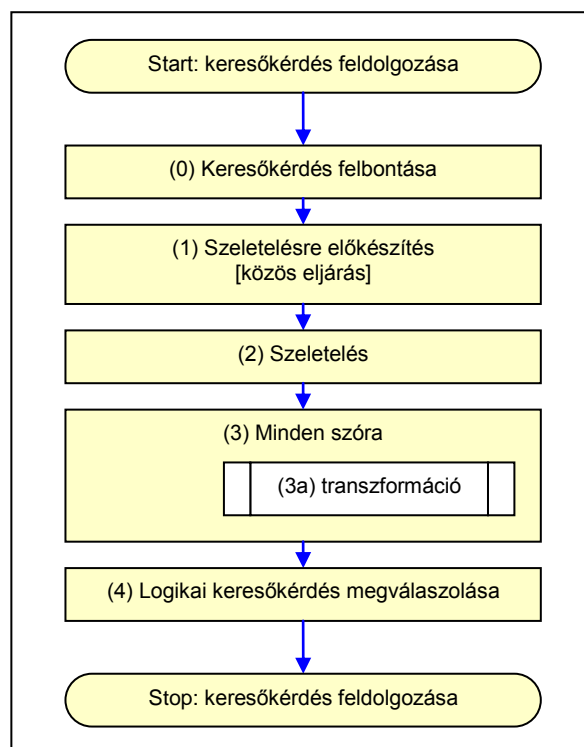
1. ábra Adatelemek feldolgozása szóalapú indexeléshez

A feldolgozási folyamat folytatásaként szeleteli (szavakra bontja) a rendszer a bejövő adatelemet (2. lépés). A szeletelő eljárás végtelenül egyszerűen működik: a szóközöket tekinti elválasztó jeleknek, és úgy készíti el a szóhalmazt. Ez a szóhalmaz lesz a következő lépés bemenete. Ez az eljárás sem változtatható meg a rendszerben. Minden keletkező szó karakteralapú transzformáción megy keresztül (3a lépés): a szavak minden karakterét, azok környezetétől függetlenül, egy legfeljebb öt karakter hosszú sorozatra cserélhetjük. Az így átalakított szóhalmaz kerül a szóalapú indexálmányokba.

Keresőkérdések szóalapú indexekre

A dokumentumok ismérvei mellett a keresőkérdést is transzformálni kell ahhoz, hogy az indexek támogatásával megtörténhessen a kereső- és a

dokumentumképek összevetése. Az általunk vizsgált és használt Aleph 500 rendszerben ez a folyamat a 2. ábrán látható módon történik a szóalapú indexeken végzett keresések esetén.



2. ábra Szóalapú keresőkérdések értelmezési folyamata

Az 1, 2, 3, 3a lépések az 1. ábrán bemutatott folyamat megfelelő lépéséhez hasonlóan működnek.

A szóalapú keresőkérdések feldolgozási folyamatának bemenete a keresendő kifejezés, és az adatkör, ahol a keresendő kifejezést értelmezi a használó. A keresendő kifejezés állhat keresőszavak egy vagy több elemű, szóközzel elválasztott sorozatából (a továbbiakban elemi keresőkérdés vagy elemi keresőkifejezés), vagy azok logikai műveletekkel összekapcsolt, opcionálisan zárójellezett sorozatából. A folyamat előkészítő, 0. lépésében a keresőkérdés szintaktikai elemzése történik meg. A logikai műveletekkel összekapcsolt, illetve zárójellezéssel csoportosított kifejezést a rendszer felbontja elemi keresőkifejezésekre, amelyeket külön kezel egészen a feldolgozás 4. lépéséig. A szintaktikai elemzés sajátossága, hogy az egyenlőségjelet (=), idézőjelet ("), valamint a kerek zárójeleket a keresést vezérlő karaktereknek tekintik, és csak jól meghatározott környezetben fordulhatnak elő. Máskülönben szintaktikai hibához vezetnek, megakadályozva ezzel a keresés végrehajtását.

A 0. lépésben előállt elemi keresőkifejezésben szereplő keresőszavak bizonyos esetekben (tipikusan a mesterséges nyelven alapuló információkereső nyelvek kifejezései esetén) tovább bonthatók a kontextuális metajelek mentén önálló jelentést hordozó egységekre¹⁰. A keresőkérdés feldolgozása során az önállóan is jelentést hordozó alkotóelemeket elválasztó különböző metanyelvi jeleket egységesen szóközzé alakítja a rendszer (1. lépés). Az Aleph 500 esetében ez rendszer szinten közös eljárás – beállításai függetlenek attól, hogy milyen adatelem-csoportra (tehát indexállományra) vonatkozó keresőkérdést dolgoz fel.

A szeletelés (2. lépés) az elemi keresőkifejezés transzformált (ti. szóközökkel tagolt) alakjának szavakra bontása. Hasonlóan a szóindexbejegyzések elkészítéséhez, a szeletelő eljárás itt is a szóközöket tekinti elválasztó karaktereknek. Ennek a lépésnek az eredménye keresőszavak halmaza, amely a következőkben az indexbe történő behasonlítás alapja lesz.

A szavakra bontott elemi keresőkifejezés minden szavát egy karakteralapú, környezetfüggetlen transzformációnak veti alá a rendszer (3., 3a lépések). Ez szintén rendszer szinten közös eljárás.

A feldolgozás 4., záró lépésében a rendszer az egyes elemi keresőkifejezések szavakra bontott alakját keresi az indexekben – ezzel közvetett módon a dokumentumképpel végez összehasonlítást. Az elemi keresőkifejezésekre kapott halmazokat a kereséskor megadott logikai műveletek (és, vagy műveletek) szerint összeveti a kereső, és kialakul a kimenet: a találati halmaz, amely visszajut a használóhoz.

Az ETO-kereshetőség kihívása a rendszer kényszereinek tükrében

Az előzőekben vázoltuk a természetes és a mesterséges nyelveken alapuló információkereső nyelvekkel végzett keresések alapjait, és bemutatottuk, hogyan történik a dokumentumrekordok előfeldolgozása, majd a tényleges keresés folyamán a keresőkérdések feldolgozása az Aleph rendszerben. Láttuk, hogy az egyes lépések bizonyos keretek között (rendszer szinten vagy adatelem-specifikusan, nem bővíthető építőelem-készletből összeállítva stb.) konfigurálhatóak, míg mások nem (pl. a *szeletelés*, amely csak a szóközök men-

tén tud dolgozni). Azonban egyik lépés sem szabható a feladathoz tetszőleges programrészeket beillesztésével: az algoritmusok adottak, amelyeket különböző trükkökkel (amelyek pusztán technikai fogásnak tűnhetnek) kell a feladathoz igazítani – ez adja az ETO részjelzetei szerinti kereshetőség biztosításának igazi kihívását.

Az adatelemek előfeldolgozása (1. ábra), valamint a keresőkérdés feldolgozási folyamata (2. ábra) jól párhuzamba állítható egymással. Az egyező számosságú lépések célja közös, de legalábbis hasonló. A markáns különbség a *szeletelésre előkészítés* (1. lépés) fázisban található. Míg az előfeldolgozáskor adatelem-specifikus eljárást használhatunk, addig a keresőkérdések kezelésekor egyetlen, rendszer szinten közös eljárás van a kezünkben. A fő probléma tehát az, hogyan hangoljuk ezt a rendszer szinten közös eljárást úgy, hogy mindegyik keresési lehetőség (ETO, szerzői nevek elemei stb.) számára optimálisan teljesítsen.

A bemutatott környezetben az jelenti a kihívást, hogy az ETO jelzetei, illetve részjelzetei szerinti kereshetőséget minél magasabb szinten biztosítsuk a többi keresés minőségének érintése nélkül. További, az előzőekben nem kellően hangsúlyozott kényszer az, hogy a szeletelésre előkészítő 1. lépés többnyire csak karakteralapú, állapotmentes¹¹ transzformációra képes: speciális esetektől¹² eltekintve csak bizonyos karakterek környezetfüggetlen törlése vagy szóközzé alakítása oldható meg a rendelkezésre bocsátott, és nem bővíthető építőelem-készlettel.

Válasz a kihívásra

A rendszer sajátosságaiból adódó feladatok megoldásait vesszük itt sorra. A csoportosítás szempontja az előzőekben bemutatott feldolgozási folyamatok egyes lépései.

A szeletelésre előkészítés (1. lépés) fázisban az ETO-ban használt azon jeleket kell vizsgálni, amelyek az önálló jelentést hordozó részeket képviselő szavak szerinti bontást segítik, vagy az elhagyásuk más módon járul hozzá a keresési minőség javításához. Az itt kezelt jeleket két csoportra bonthatjuk:

1. amelyek eltűnése nem befolyásolja, vagy kifejezetten növeli a találati halmaz relevanciáját
 - a. a + jel (jelentése a fogalmak halmazelméleti uniója, VAGY kapcsolat) bizonyult ilyennek,

2. párban álló, zárójeltípusú jelek, amelyek nyitó és záró karaktere állapotmentesen megkülönböztethető (ti. azok különböznek). Ekkor a nyitó karakter mentén szeletelhetünk, míg a záró karaktert meghagyva megmarad a részjelzet információtartalma: például a (439) – Magyarország részjelzetből 439 lesz, ami továbbra is hordozza azt az információt, hogy földrajzi alosztásról van szó.

- a. (...) (kerek zárójel – a különféle alosztások jele). A nyitó karakter mentén szeletelhetünk. A kereséskor helyette a kapcsos zárójel használandó, l. a következő részben.
- b. bár az "..." (idézőjel – időbeli közös alosztás, a fogalom időbeli vonatkozásának jelölése) is zárójeltípusú, a nyitó és záró karakter csak a környezetét is figyelembe véve különböztethető meg. Tekintve, hogy a szeletelésre előkészítés lépésben a karakterek környezetét nem tudjuk figyelembe venni (l. korábban), ezért a "16" – 17. század részjelzetből az idézőjel elhagyásával 16 lenne, ami nem különböztethető meg a 16 – Logika¹³, logisztika, ismeretelmélet fő táblázati számától. Ezért a "..." jelet másik lépésben kell kezelünk.

A keresőkérdések feldolgozási lépésében (0. lépés) felvetett problémákra (ti. a kerek zárójelek, az egyenlőségjel és az idézőjelek speciális, keresésszintaktikai jelentése) a jelzetek kezelése során adható válasz. Két csoportba sorolhatók a karakterek:

1. A szóindexek építésének (3a) lépésében lecserélt karakterek: az egyenlőség (=) és a dupla idézőjeleket (") rendre az osztás (÷) és a keresztszorítás (×) jelekre cseréljük. A felhasználónak kell gondoskodnia a csere karakter hasz-

nálatáról kereséskor (l. 1. táblázat), hiszen az egyenlőségjel (=) használata szintaktikai hibát eredményezne, míg az idézőjel (") speciális jelentésű a kereső számára.

2. A keresőkérdések feldolgozásának 3a részében visszacserélt karakterek. Az ETO esetében csupán a kapcsos zárójelet¹⁴ ({}), amely a kerek zárójel helyettesítésével az alosztások alapján történő keresést teszi lehetővé. A keresőkérdés megfogalmazásakor az alosztásokat jelző kerek zárójel kapcsos zárójellel történő helyettesítéséről a felhasználónak kell gondoskodni.

A keresést végrehajtó felhasználó szempontjából nincs különbség a két csoport között¹⁵: az 1. táblázat 2. oszlopában megadott karaktereket kell használni a kereséskor.

Keresési példák, hiányosságok

A keresési módszer alkalmazásával új, az adatbázisban eddig rejtetten meglévő információkhoz juthatunk el. Erre szeretnénk néhány példát mutatni három példagyűjteményben. A keresést az Aleph integrált könyvtári rendszerrel végezzük a BME OMIKK adatbázisában¹⁶, és megadjuk a releváns találatok becsült arányát¹⁷. Ahol másként nem jelezzük, ott 100% a becsült relevancia.

Tematikus példák: Magyarország

Ebben a gyűjteményben Magyarország osztályozására szolgáló földrajzi alosztásból kiindulva térképezzük fel az adatbázisunkat.

Magyarország (439) jelzetét keresőszóként {439} formában a 3. ábra szerint írhatjuk a keresőbe, melyre 8228 találatot ad a kereső (4. ábra).

1. táblázat

Az ETO keresésekor alkalmazandó csere karakterek, zárójelben a Unicode kódjaik hexadecimális írásmóddal

Eredeti karakter az ETO-jelzetben	Csere karakter kereséskor	Elérése magyar billentyűzeten	Megjegyzés
= (0x003D)	÷ (0x00F7)	AltGr + Ő	
" (0x0022)	× (0x00D7)	AltGr + Ú	
) (0x0029)	} (0x007D)	AltGr + N	Visszacserélt karakter.
((0x0028)	{ (0x007B)	AltGr + B	Rendszerszinten mellőzhető transzformáció, és kereséskor szóközzel helyettesíthető a szeletelés miatt (l. a 14. lábjegyzetet).

Keresés

Írja be a kereső szót/szavakat {439}

Mező a kereséshez ETO-jelzet

Szavak egymás mellett? Nem Igen

3. ábra Magyarország keresőkérdése a BME OMIKK webes katalógusában

Keresőkérdés:
ETO-jelzet új= 439)
Rendezve: Év, Szerző

Eredmény: 8228 rekord (max 1000 rekord jeleníthető meg és rendezhető)

Rekordok: 1 - 20 / 8228

#	Szerző	Cím	Év
1	Cséffán József	A közalkalmazottak jogállásáról szóló törvény és magyarázata	2008
2	Gáthner Petra	Réber László, 1920-2001	2008
3	Gerlei Tamás	Az Ikarus évszázados története	2008
4	Gyarmati Gabriella	Munkácsy képek Kecskeméten : kiállítás Munkácsy Mihály műveiből v. Pákh Imre amerikai	2008
5	Maróty Katalin	Hild József építészete : tanulmányok az építés halálának 140. évfordulóján rendezet	2008
6	Mendöl Zsuzsanna	Forbát Alfréd, 1897-1972	2008
7	Michelberger Pál	Közlekedés a XXI. században jövőről a jelenben	2008
8	Molnár Gyula	Az én vizuális nyelvem	2008

4. ábra Magyarország – találati halmaz (részlet)

1 / 1 rekord

Raktári jelzet	909.199
	B 32023
Könyvtár info	Központi könyvtár
Országkód	hu
Nyelvkód	hun
Cím	Kelet-Magyarország regionális vízellátása
ETO	828.1(439-11) 711.2(439-11) 828.1(439-11)

5. ábra Kelet-Magyarország – találati halmaz

Kelet-Magyarország osztályozására a (439-11) jelzet szolgál, amelyet a {439-11} formában kereshetünk. Eredményként egyetlen találat adódik (5. ábra).

A (436:439) jelzettel azokat a dokumentumokat osztályozzuk, amelyek Ausztria és Magyarország kapcsolatáról szólnak. A keresőszó {436:439}, az adatbázisunkban nyolc dokumentumot találunk (6. ábra).

1 / 8 rekord

Raktári jelzet	B 31783
Könyvtár info	Központi könyvtár
Könyvtár info	Vasútépítés
Országkód	HU
Nyelvkód	hun
Cím	125 éves a Bécs-győri vasút
Akírói szói	Győr-Sopron-Ebenfurti Vasút
Tárgyszó	Vasúti közlekedés
ETO	856.2(436:439)(091)

6. ábra Ausztria és Magyarország kapcsolata – egy rekord a találati halmazból

Az Osztrák-Magyar Monarchia az előbbi speciális esete, leírására a (436/439) jelzetet használjuk. A keresőszó {436/439}, melyre 13 találatot kapunk (7. ábra).

9 / 13 rekord

Raktári jelzet	296.399
Könyvtár info	Központi könyvtár
Országkód	hu
Nyelvkód	hun
Szerző	Csonkaréti Károly
Cím	Császári és királyi hadihajók
Tárgyszó	Hadihajó Osztrák-Magyar Monarchia időszaka
ETO	Első világháború 623.82(436/439) 355.49(4)*1914/1918" 359(436/439)(091)

7. ábra Dokumentumrekord az Osztrák-Magyar Monarchia találati halmazból

A magyarországi németek jelzete (439=30), a megfelelő keresőszó {439=30}, amely négyszer fordul elő az adatbázisban (8. ábra). Osztályozói hibából eredően a jelzet (439:=30) változatban is előfordul. Ezt a {439:=30} keresőkérdésre adott két találat mutatja (9. ábra).

1 / 4 rekord

Raktári jelzet	367.568/20
Könyvtár info	Központi könyvtár
Országkód	hu
Nyelvkód	gerhun
Cím	Beiträge zur Volkskunde der Ungarndeutschen
Sorozat	A magyarországi nemzetiségek néprajza
Tárgyszó	Néprajz Magyarország Németek
ETO	39(439=30)(082)

8. ábra Magyarországi németek – a találati halmaz egy eleme

1 / 2 rekord

Raktári jelzet	421.799
Könyvtár info	Központi könyvtár
Országkód	hu
Nyelvkód	hun
Cím	A magyarországi németek kitelepítése és az 1941. évi népszámlálás
Tárgyszó	Nemzetiségi kérdés Demográfia Statisztika Magyarország Németek Kitelepítés
ETO	325.254.1(439=30) 312.954(439)*1941"

9. ábra Magyarországi németek – osztályozói hibából helytelen jelzettel is szerepel az adatbázisban

Magyarországi események történeti tárgyalása osztályozási jelzete a (439)(091). A megfelelő keresőkérdés {439}{091}, amely 1173 dokumentum-rekordot eredményez (10. ábra). Ennek a találati halmaznak a relevanciája nehezen becsülhető, de magasabb, mint 88%.

Keresőkérdés:
ETO-jelzet uj= 439} 091};
Rendezve: Év, Szerző

Eredmény: 1173 rekord (max 1000 rekord jeleníthető meg és rendezhető)
Rekordok: 1 - 20 / 1173

Ugrás szövegre Ugrás #

#	Szerző	Cím	Év
1	<input type="checkbox"/> Gerlei Tamás	Az Ikarus évszázados története	2008
2	<input type="checkbox"/> Szentgyörgyi Zsuzsa	Mémők, tudós, iskolateremtő : Michelberger Pál és kora	2008
3	<input type="checkbox"/> Alföldi László	Bányászati karsztvízszint-süllyesztés a Dunántúli-középhegységben : rekvium a Dunántú	2007
4	<input type="checkbox"/> Babiczky László	Szabadság tér 17. : a Magyar Televízió tündöklése és...	2007
5	<input type="checkbox"/> Batta István	A magyar fizikai szaknyelv történetéből : A fizikai tudományok hazai irodalmának történet	2007
6	<input type="checkbox"/> Bertók László	Víznyomok Magyarországon	2007
7	<input type="checkbox"/> Fülöp András	Szarvasbikák, gépmadarak	2007
8	<input type="checkbox"/> Hámosi Ferenc T.	Magyarország mozdonyai	2007

10. ábra **Magyarországi események történeti tárgyalása – találati halmaz (részlete)**

A fizika története Magyarországon – ez az **53(439)(091)** jelzet, a keresőszó pedig **53{439}{091}**. Összesen 16 találatot kapunk. A fizika, vagy valamely szakterületének magyarországi történetét az **53?(439)(091)** jelzetcsoport jelenti, a kereső azonban jelenleg nem képes a túl sok illeszkedő fogalom (a fizika túl sok szakterülete) miatt a kérdés kiszolgáltatására (11. ábra).

Az 1956-os magyarországi események jelzete **(439)"1956"**, amelyet a **{439}×1956×** keresőkérdelemmel érhetünk el. A kereső 34 találatot ad (12. ábra). Az idő szerinti alosztások kereshetősége miatt az adatbázisunk kronológiaként is szolgál.

Keresőkérdés:
ETO-jelzet uj= 53 439} 091};
Rendezve: Év, Szerző

Eredmény: 16 rekord (max 1000 rekord jeleníthető meg és rendezhető)
Rekordok: 1 - 16 / 16

Ugrás szövegre Ugrás #

#	Szerző	Cím	Év
1	<input type="checkbox"/> Batta István	A magyar fizikai szaknyelv történetéből : A fizikai tudományok hazai irodalmának történet	2007
2	<input type="checkbox"/> Simonyi Károly tudományos emlékülés 2006 Budapest	Simonyi Károly tudományos emlékülés : születésének 90-ik, halálának 5-ik évfordulójára	2007
3	<input type="checkbox"/> Simonyi Károly	A magyarországi fizika kultúrtörténete, XIX. század : (vázlat)	2002
4	<input type="checkbox"/> Horváth Zsolt	Fizika	2001
5	<input type="checkbox"/> Simonyi Károly	A magyarországi fizika kultúrtörténete, XIX. század : (vázlat)	2001
6	<input type="checkbox"/> Gazda István	A magyarországi fizika klasszikus századai, 1590-1890 : [Segédkönyv a felsőfokú oktatási i	2000
7	<input type="checkbox"/> Gazda István	A magyarországi fizika klasszikus századai, 1590-1890	2000
8	<input type="checkbox"/> Batta István	A magyar fizikai szaknyelv fejlődése : A fizikai tudományok hazai 1999 irodalmának története	

11. ábra **A fizika története Magyarországon – találati halmaz (részlet)**

Keresőkérdés:
ETO-jelzet uj= 439}×1956×;
Rendezve: Év, Szerző

Eredmény: 34 rekord (max 1000 rekord jeleníthető meg és rendezhető)
Rekordok: 1 - 20 / 34

Ugrás szövegre Ugrás #

#	Szerző	Cím	Év
1	<input type="checkbox"/> Láncai András	Korrajz, 2007 : a XX. Század Intézet évkönyve	2007
2	<input type="checkbox"/> András Sándor	1956 : egyetemisták a forradalomban	2006
3	<input type="checkbox"/> Békés Csaba	A forradalom és a magyar kérdés az ENSZ-ben, 1956-1963 : tanulmányok, dokumentumok és kronol	2006
4	<input type="checkbox"/> Bögre Zsuzsanna	Asszonyorsok : ötvenhatos élettörténetek elemzése	2006
5	<input type="checkbox"/> Bratkowski, Andrzej	Bezbronne braterstwo : Kraków - Budapest 1956	2006
6	<input type="checkbox"/> Congdon, Lee	1956: the Hungarian revolution and war for independence	2006
7	<input type="checkbox"/> Csurka Gergely	1956 : ahol mi győztünk : [1956. december 6. Magyarország - Szovjetunió 4-0] : [a "melbourne-	2006

12. ábra **Az 1956-os magyarországi események – találati halmaz (részlet)**

A Magyarországgal kapcsolatos keresési példákat a 13. ábrán foglaltuk össze.

Korábbi keresések eredményei

Keresőkérdés	Rekordok száma
ETO-jelzet uj= 439}	8228
ETO-jelzet uj= 439-11}	1
ETO-jelzet uj= 436 439}	8
ETO-jelzet uj= 436/439}	13
ETO-jelzet uj= 439+30}	4
ETO-jelzet uj= 439 +30}	2
ETO-jelzet uj= 439} 091}	1173
ETO-jelzet uj= 53 439} 091}	16
ETO-jelzet uj= 439}×1956×	34

13. ábra **Az első példaszor (Magyarország) kereséseinek összefoglalása**

Tematikus példák: német nyelv

Második példagyűjteményünkben a nyelvi alosztások közül a **=30** német nyelv jelzetét vizsgáljuk.

Az állományunkban a **=30** jelzettel a német nyelvvel foglalkozó dokumentumokat (szótárak, nyelvtankönyvek stb.) osztályozunk. A megfelelő keresőkérdés a **?=30**, amely 536 találatot eredményez.

A német nyelvű általános értelmező szótárakat a **801.316.4=30** jelzetre keresve találjuk meg. A keresőkérdés **801.316.4+30**, a találati halmaz mindösszesen 6 tételből áll.

A német nyelvű általános vagy valamely szakterület értelmező szótára a **801.316.4?=30** jelzetcso-

porttal írhatjuk le. A keresőkérdés **801.316.4?÷30**, amely 93 találatot eredményez.

A német nyelvvel kapcsolatos keresési példákat a **14. ábrán** foglaltuk össze.

Korábbi keresések eredményei	
Keresőkérdés	Rekordok száma
ETO-jelzet uj= ?÷30	536
ETO-jelzet uj= 801.316.4÷30	6
ETO-jelzet uj= 801.316.4?÷30	93

14. ábra A második példasor (német nyelv) kereséseinek összefoglalása

Tematikus példák: németek

A következő példákban az etnikai (népi) alosztásokra áttérve a **(=30)** németek jelzetéből épített komplex kereséseket mutatunk be.

Az **(=30)** – németek részjelzetre a **{÷30}** keresőkérdéssel kereshetünk: 9 dokumentumot találunk.

Az erdélyi szászok jelzete a **(439.21=30)**, a keresőkérdés **{439.21÷30}**, összesen 4 találatot kapunk.

A németek kultúrtörténelmét a **930.85(=30)** jelzet osztályozza, a keresőkérdés **930.85{÷30}**, amire 3 találatot ad a keresőrendszer.

Németek dalai: **784(=30)**, a keresőkérdés **784{÷30}**. Az adatbázisban egyetlen találatot kapunk.

A németek témaköréhez kapcsolódó keresések eredményeit a **15. ábrán** foglaltuk össze.

Korábbi keresések eredményei	
Keresőkérdés	Rekordok száma
ETO-jelzet uj= ÷30}	9
ETO-jelzet uj= 439.21÷30}	4
ETO-jelzet uj= 930.85 ÷30}	3
ETO-jelzet uj= 784 ÷30}	1

15. ábra A harmadik példasor (németek) kereséseinek összefoglalása

A kötőjeles speciális alosztás okozta hiba

A **72 – Építészet** jelzetre végzett keresés találati halmazában előfordulnak a **-72** nem önálló, kötőjeles alosztást tartalmazó rekordok is (**16. ábra**).

A **72** keresőkérdésre kapott találati halmaz 1693 eleméből 1554-et találtunk relevánsnak, ami 92%-os relevanciát jelent. A **72 NEM ?-72** keresőkérdéssel (**17. ábra**) a találatok száma 1560-ra csökken, megtartva az 1554 releváns találatot. Ez 99%-os relevancia.

Hasonlóképpen: a **16 – Logika, logisztika, ismeretelmélet** főtblázati számra való keresés olyan találatokat is eredményez, amelyekben a **-16** alosztást tartalmazó jelzet fordul elő. Ennek a 237 elemű találati halmaznak 84 eleme releváns. A mintegy 38%-os, kiugróan alacsony relevancia a **16 NEM ?-16** keresőkérdéssel 86%-ra javítható: a 98 elemű találati halmaz tartalmazza a 84 releváns találatot.

5 / 1693 rekord	
Raktári jelzet	BM 22
Könyvtár info	Nukleáris
Országkód	us
Nyelvkód	eng
Szerző	Tu, Jiyuan
Cím	Computational fluid dynamics
Alcím	practical approach
Tárgyszó	Áramlástan
	Numerikus módszer
	Szimuláció
	Operációkutatás
ETO	532.5:519.63-72
	519.63-72:532.5
	519.876.5:[532.51]

16. ábra Irreleváns találat a 72 – Építészet találati halmazban a -72 speciális alosztás előfordulása miatt

Keresés	
Írja be a kereső szót/szavakat	72 NEM ?-72
Mező a kereséshez	ETO-jelzet
Szavak egymás mellett?	<input checked="" type="radio"/> Nem <input type="radio"/> Igen
<input type="button" value="Mehet"/> <input type="button" value="Töröl"/>	

17. ábra A 72 – Építészet találati halmazból kizárjuk a -72 speciális alosztás előfordulásait

Az ETO-jelzetrendszer bemutatása c. szakaszban a nem önálló alosztások, így a kötőjeles speciális alosztások szerint végzett keresésről megállapítottuk, hogy a keresőkérdés értelmezése sem adható meg általában. Itt azonban egy jól értelmezhető keresőkérdésre adott találati halmazba kerülnek

bele nem releváns találatok. A jelenség oka a kötőjeles kifejezések speciális kezelése az Aleph-ben. A részletes magyarázattól itt eltekintünk, azonban fontosnak tartjuk, hogy a megoldás hiányosságai is dokumentáltak legyenek.

Összefoglalás

Cikkünkben az ETO osztályozási rendszer jelzetei szerint végzett keresés kapcsán bemutattuk annak felépítését, és a szabadszavas keresésekre optimalizált keresőrendszerekben felmerülő ETO-kezelési problémákat.

Általános megoldási elveket, és a BME OMIKK-ban használt Aleph 500 integrált könyvtári rendszer esetén konkrét megoldást adtunk egy jó minőségű ETO-jelzetek szerinti keresés megvalósítására. Ez a módszer lehetővé teszi az összetett jelzetek egyes elemeire, az önálló és a nem önálló általános alosztásokra történő keresést. Segítségével kihasználhatjuk azt az előnyt, amit az online katalógusok nyújtanak a cédulakatalógusokhoz képest: a feldolgozott információk gazdagabban tárolnak fel a kereső előtt. Adatbázisunkban jelenleg az 1991-ig Magyarországon kiadott ETO-táblázatok alapján szerkesztett jelzetek szerepelnek. Ha könyvtárunkban áttérünk az új magyar ETO (UDC Publ. No. P057) kiadás szerinti osztályozásra, a keresési módszer továbbra is alkalmazható lesz.

A megoldás nem tökéletes, és használatához néhány egyszerű szabály ismerete elengedhetetlen. A bemutatott keresési példák azonban illusztrálják, hogy a meglévő rendszerekhez képest jobb minőségű ETO-jelzetek szerinti keresést valósíthatunk meg.

Jegyzetek

- ¹ Viszonyítást, vagy más relációt (pl. róla szól, valamely szempont szerint) kifejező természetes nyelvi fordulatokat is figyelembe vevő kereséstámogatás.
- ² A 2005-ös magyar ETO-kiadásban már nem szerepelnek a Szempont szerinti nem önálló alosztások, de a könyvtári adatbázisokban a régebbi ETO-kiadásokból származó jelzetekben továbbra is előfordulnak.
- ³ A 2005-ös kiadásban már a Számvégződéses speciális alosztások sem fordulnak elő.
- ⁴ Felmerülhet a kérdés, hogy a 21. század informatikai környezetében miért van szükség az ékezetmentesítésre. Korábbi technikai színvonalon (esetleg hibá-

san) elkészített dokumentumképpen például „kalapos ő” szerepelhet, vagy idegen nemzeti környezetben nehézséget jelenthet az „ő” helyes bevitele. Másik átalakítási példa a kis- és a nagybetűk közötti konverzió.

- ⁵ A dokumentumrekordban más forma lehet jelen, hiszen rendszerint már a megjelenítő program is átalakítja a dokumentumrekordot.
- ⁶ Egy elemző vagy transzformációs eljárás környezetfüggetlen, ha az általa éppen feldolgozott adatot (karaktert, vagy karaktercsoportot stb.) a környezetéből kiragadva kezeli. Például az egyszerű idézőjel (") környezetének (ti. a mellette levő karakterek) ismerete nélkül nem lehet eldönteni: nyitó vagy záró idézőjellel van-e szó.
- ⁷ Ilyen az ISBN-10 kódok nyilvántartási csoport-, kérelmező- és kiadványelemei. Például az ISBN-10: 963-05-7944-8, ahol 963 Magyarország, 963-05 az Akadémiai Kiadó, 963-05-7944 pedig a konkrét mű azonosítója, amelyhez egy ún. ellenőrző számjegyet csatolnak (itt: 8). Az ISBN-13 szintén hierarchikus, azonban a 978, 979 prefixek kevésbé szemléletes módon illenek a hierarchiába.
- ⁸ Bár lehetséges az Akadémiai Kiadóban megjelentett műveket a 963-05 ISBN-10 prefix alapján keresni, ritkán járunk el így. Sokkal jobban kézre áll az egységesített kiadói nevekben történő keresés, ami ráadásul áthidalja valamely kiadó több ISBN tartománya alapján történő keresés problematikáját stb.
- ⁹ A szóközzé alakításra az előfeldolgozás (2) lépése miatt van szükség: annak tárgyalásakor visszatérünk erre.
- ¹⁰ Felmerülhet a kérdés: miért bontanánk tovább például az 53(439)(091) elemi keresőkérdést (Fizika története Magyarországon). A válasz a keresés módszerében keresendő. A keresőképet nem közvetlenül a dokumentumképekkel, hanem az azokból épített, a keresést támogató indexekkel hasonlítjuk össze. Az indexekben pedig külön bejegyzésként található az 53 – Fizika, (439) – Magyarország és a (091) – Történeti szempontból részjelzetek.
- ¹¹ Egy eljárás (algoritmus stb.) állapotmentes (stateless), ha az adatelemek (karakterek stb.) feldolgozását az előzőektől függetlenül végzi. Ezzel szemben az állapottal rendelkező (stateful) eljárások a korábban látott adatelemekből (karakterekből stb.) egy kivonatot (az ún. belső állapotukat) készítenek, amit felhasználnak az aktuális bemenet feldolgozásához.
- ¹² Ilyen speciális eset például a pontozott formában írt betűszavak, mint például A.B.C.D. átalakítása az ABCD formára, mielőtt a kifejezésben szereplő pont (.) karaktereket a szeletelést lehetővé tevő szóközre cserélnénk. (Ennek az eljárásnak nincs jelentősége az ETO-jelzetek kezelésekor.)

- ¹³ Logika az 1-es, Filozófia táblázaton belül.
- ¹⁴ A szemfüles olvasó észreveheti, hogy a kerek zárójel nyitóeleme a szeletelésre előkészítés során, tehát az 1. lépésben szóközzel helyettesítődik. Hasonlóképpen a keresőkérdés feldolgozásának 1. lépésében a nyitó kapcsos zárójelet szóközzel kell helyettesítenünk. Ezt a helyettesítést a keresést végző felhasználó is elvégezheti: a keresőkép összeállításakor a megfelelő nyitó kapcsos zárójel helyett szóköz írható. Természetesen ez csak opcionális, a {439} és a {439} keresőkérdés egyaránt a várt eredményt hozza.
- ¹⁵ Ha a felhasználó szempontjából nincs különbség, akkor mégis mi az oka a megkülönböztetésnek? A keresést támogató szóindexben a bejegyzések ennél a módszernél kerek zárójelet tartalmaznak, ami megoldás kidolgozása során jelentett segítséget.
- ¹⁶ Az internetes katalógus és kereső elérhető a <http://aleph.omikk.bme.hu/> címen.
- ¹⁷ A releváns találatokat az Aleph-től független, saját fejlesztésű programmal, valamint kézi módszerrel számoltuk.

Irodalom

Aleph Configuration Guide. Ex Libris Ltd., 2004.

The Aleph 500 User Guide. Ex Libris Ltd., 2004.

B. HAJDU Ágnes – BABICZKY Béla: Bevezetés az információkereső nyelvek elméletébe és gyakorlatába. Budapest, Universitas Kiadó, 1998. ISBN 963-9104-20-5

BACH Iván: Formális nyelvek. Budapest, Typotex, 2005. ISBN 963-9132-92-6

Egyetemes Tizedes Osztályozás. Teljes kiadás. Segéd-táblázatok (FID Publ. No. 390)

KÖRNYEI Márta: Könyvtári osztályozás: Tanárképző Főiskolák: Egységes jegyzet. Budapest, Nemzeti Tankönyvkiadó, 1997. J11-1135

UNGVÁRY Rudolf – VAJDA Erik: Könyvtári információkeresés. Budapest, Typotex, 2002. ISBN 963-9326-29-1

UNGVÁRY Rudolf – VAJDA Erik: 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

ZÖLDI Péter: Az ETO az Országos Széchényi Könyvtár számítógépes rendszerében. = TMT, 39. köt. 3. sz. 1993.

http://www.epa.oszk.hu/00100/00143/00007/zoldi_h.html

www.unicode.org, kódpontok, Basic Latin

<http://www.unicode.org/charts/PDF/U0000.pdf>, [2008. 03. 30.]

Beérkezett: 2009. I. 4-én.



Marton József

mérnök-informatikus,
a BME OMIKK informatikus
munkatársa.

E-mail: jmarton@omikk.bme.hu



Prokné Palik Mária

a BME OMIKK osztályozó
könyvtáros munkatársa.

E-mail: mpalik@omikk.bme.hu

Kitüntetés a Magyar Kultúra Napján

2009. január 22-én, többekkel együtt *Bibliotéka Emlékérem*-díjat vehetett át

Bánhegyi Zsolt,

a *TMT Tanácsadó testületének* tagja, a *Magyar Tudományos Akadémia Könyvtárában* végzett több évtizedes és a magyar könyvtári kultúra fejlődése érdekében kifejtett elméleti és gyakorlati munkájáért.

Szívből gratulálunk!

(A TMT szerkesztősége)