

MATARKA - magyar folyóiratok tartalomjegyzékeinek kereshető adatbázisa

*Burmeister Erzsébet, erzsi@marki.lib.uni-miskolc.hu
Miskolci Egyetem, Könyvtár, Levéltár, Múzeum*

*Kiss Andrea, konpinty@uni-miskolc.hu
Miskolci Egyetem, Könyvtár, Levéltár, Múzeum*

Bevezetés

A könyvtári tájékoztatásban nap mint nap találkozunk azzal a problémával, hogy irodalomkutatást kell végeznünk egy témakörben, vagy egy szerző műveinek bibliográfiáját kell összeállítanunk; az ehhez szükséges információk egy része azonban folyóiratokban van elszórva. Az is megtörténhet, hogy az olvasó egy konkrét cikket keres, de nem tudja pontosan, melyik folyóiratban vagy melyik számban jelent meg.

Ilyenkor gyakran csak több évfolyamnyi folyóirat átlapozásával találjuk meg (vagy még úgysem), amit kerestünk. Ez persze hosszú ideig tart és türelmet, figyelmet igényel. Hogyan segíthetnek ebben a munkában az informatikai eszközök? A folyóiratok és az abban megjelenő cikkek analitikus feldolgozásával több adatbázis is épül, ezek egy része CD-ROM-on, mások az interneten keresztül is elérhetőek. Kereső felületeik biztosítják, hogy gyorsan és könnyen eljussunk a minket érdeklő információkhoz. Ezeknek az adatbázisoknak azonban vannak hátrányaik is. Egyrészt egy-egy témakörre vagy kiadványtípusra specializálódtak, és messzemenően nem fednek le minden területet. Ráadásul több közülük csak válogatást ad a folyóiratok anyagából. Másrészt lassan épülnek ezek a szolgáltatások, következésképp a legfrissebb információk gyakran nem találhatók meg bennük.

Egy folyóiratban való kereséskor kiindulópontunk általában a tartalomjegyzék, hiszen gyakran a cikkek címei is eligazítanak azok tartalmáról. Ésszerűnek tűnik tehát, hogy felhasználjuk ezeket a tartalomjegyzékeket egy olyan, általánosan elérhető, gyorsan elkészíthető és frissíthető szolgáltatáshoz, mely megkímél bennünket a folyóiratokban való fásasztó keresgéléstől, ugyanakkor tájékoztatást nyújt azok teljes tartalmáról. A MATARKA elindításakor egy ilyen szolgáltatás létrehozása lebegett a szemünk előtt.

Előzmények

Az interneten már régóta elérhető jónéhány folyóirat tartalomjegyzéke. Eleinte azonban ezek csak statikus oldalak voltak, és ugyanúgy végig kellett böngészni az információkért azokat, mint a nyomtatott kiadványokat. Csak az utóbbi időben kezdtek megjelenni olyan oldalak, ahol valamilyen keresési lehetőséget is biztosítottak, de természetesen csak az adott folyóiraatra vonatkozóan.

Mi viszont úgy kívántuk létrehozni tartalomjegyzék-szolgáltatásunkat, hogy a böngészés mellett több szakterületen, egyszerre több folyóiratban is gyorsan lehessen keresni.

A szolgáltatást 2002-ben indítottuk be két NKÖM pályázat támogatásával.

Ingyenes szoftverek (Apache webservert - PHP HTML-be ágyazható programozási nyelv - MySQL relációs adatbáziskezelő) felhasználásával hoztuk létre az adatbázist és készítettük el az adatbeviteli valamint a lekérdező felületet.

Első lépcsőben a Miskolci Egyetem Könyvtár, Levéltár, Múzeum munkatársai 15 műszaki és természettudományi folyóiratot dolgoztak fel 10 évre visszamenőleg, majd még 7 könyvtár csatlakozásával létrehoztunk egy konzorciumot, hogy bővíthessük a feldolgozott folyóiratok körét.

Az induló konzorcium tagjai:

- Berzsényi Dániel Főiskola Központi Könyvtára
- Budapest Főváros Levéltára
- Márton Áron Szakkollégium, Budapest
- MTA Nyelvtudományi Intézet Könyvtára
- Nyíregyházi Főiskola Központi Könyvtár
- Szegedi Egyetem Szegedi Élelmiszeripari Főiskolai Kar Könyvtára
- VATI KHT (Magyar Regionális Fejlesztési és Urbanisztikai Közhasznú Társaság) könyvtára

Azóta újabb 4 intézmény:

- Debreceni Egyetem Egyetemi és Nemzeti Könyvtára, Társadalomtudományi Könyvtár
- MÁV Dokumentációs Központ és Könyvtár

- MTA Regionális Kutatások Központja Könyvtára
 - Természetvédelmi Hivatal könyvtára
- csatlakozott a közös munkához, melynek eredményeként jelenleg 142 folyóirat, összesen 101093 címe található meg az adatbázisban.

A MATARKA bemutatása

A MATARKA magyar kiadású (ezért elsősorban magyar nyelvű) folyóiratok tartalomjegyzékeit dolgozza fel.

A gyűjtőkörünk:

- agrár
- általános
- gazdasági
- jogi
- műszaki
- nyelvtudomány
- társadalomtudományok
- természettudományok

Ebben az összeállításban a résztvevő könyvtárak gyűjtőköre erősen tükröződik. Így például az agrár szakterületet leginkább az élelmiszeripari folyóiratok jelentik a Szegedi Tudományegyetem Szegedi Élelmiszeripari Főiskolai Kar Könyvtára révén. Ugyanígy Budapest Főváros Levéltárának könyvtárosa segítségével lefedjük a levéltári szakfolyóiratokat, az MTA Nyelvtudományi Intézetének könyvtára pedig a nyelvtudományi folyóiratokkal járul hozzá az adatbázisához.

Tartózkodunk a szépirodalmi vagy szépirodalommal erősen vegyített folyóiratok felvételétől, mert ezekben a címek nem sokat árulnak el a tartalomról /sajnos ez sok társadalomtudományi folyóiraatra is igaz/ és ezért a tartalomjegyzékek legfeljebb szerző szerint lennének hatékonyan kereshetők. Az orvosi folyóiratokat például azért nem dolgozzuk fel, mert létezik a MOB (Magyar Orvosi Bibliográfia). Kezdetben volt egy önző szempontunk is, a saját egyetemünk karainak szakterületeit próbáltuk lefedni. Ennek ellenére nevünkkel összhangban gyűjtőkörünk meglehetősen tarka.

A folyóiratok többsége 5-10 évre visszamenőleg található meg az adatbázisban, de vannak ennél sokkal hosszabb időszakokra visszanyúló feltárt folyóiratok is.

Adatbevitel

Adatbevitelre két módon van lehetőség:

1. Interneten keresztül, tetszőleges böngészővel, felhasználói névvel és jelszóval közvetlenül az éles adatbázisba viszi be az adatokat a könyvtáros. (Létezik egy teszt adatbázisunk is, amely egyrészt a fejlesztések tesztelésére, másrészt az újonnan csatlakozó tagoknak a beviteli folyamat gyakorlására szolgál. Ez utóbbira azért is szükség van, mert ugyan készült egy útmutató a használathoz és a feldolgozási szabályokat is rögzítettük, de munka közben szoktak felbukkanni az esetleges kérdések.) Az adatbevitellel foglalkozó munkatársaknak tulajdonképpen nem kell mást tenniük, mint az interneten használatosakhoz hasonló űrlapokat kell kitölteniük.



1. ábra: Egy adatbevitelhez használt oldal

2. Megadott formátumban tetszőleges szövegszerkesztővel (jegyzettömb, Word, stb.) felveszi a könyvtáros az adatokat, a fájlt csatolva e-mailben eljuttatja nekünk, és mi egy konvertáló program segítségével bevisszük az adatokat az adatbázisba.

Felmerülhet a kérdés, hogy miért van erre a kétféle feldolgozási módra szükség. A konvertáló program készítésének gondolata azért merült fel, hogy be lehessen vinni az interneten nagyon sok helyen megtalálható tartalomjegyzékeket újbóli begépelés nélkül. A BKÁE Egyetemi Központi Könyvtár hozzájárult ahhoz, hogy az általuk HTML oldalakon szolgáltatott gazdasági folyóiratok tartalomjegyzékét leszedjük, és átalakítás után az adatbázisba konvertáljuk. Az átalakítás eredményeképp keletkezik az alább bemutatott formátum, melynek bekonvertálása egy Mehet gomb megnyomását jelenti - ilyenkor egy PHP program fut le. Több könyvtáros inkább így küldi nekünk a tartalomjegyzékeket, pl. az MTA Regionális Kutatások Központja könyvtárosa /Tér és társadalom, Területi statisztika, Település és térségfejlesztés/ vagy a Természetvédelmi Hivatal könyvtárosa /Aquila, Crisicum, Természetvédelmi Közlemények, Denevérkutatás/. De nem könyvtárosok is hajlandóak erre a munkára, pl. a CEO Magazin,

A műszerügyi és mérés technikai közlemények is így jut el hozzánk, sőt a legutolsó számtól kezdve a Könyv, Könyvtár, Könyvtáros is. Ezt a formátumot szkenneléssel és szövegfelismertetéssel is elő lehet állítani.

A formátum egy aránylag rövid tartalomjegyzék esetén:

Könyv Könyvtár Könyvtáros, 2003, 12, 9
* Könyvtárpolitika
Pegán Anita-Vajda Kornél: A könyvtár - kapu a világra 3-7
Fogyatékkal élő emberek a közkönyvtárakban - követelmények és lehetőségek az információs technológia társadalmában 8-9
Skaliczki Judit: A fogyatékkal élők integrálásáért 10-14
Europa Nostra-díjas a Fővárosi Szabó Ervin Könyvtár 14-15
* Fórum
Van-e a könyvtártudománynak hermeneutikája? 16-23
* Műhelykérdések
Moldován István: MEK fejlesztések 2002-2003-ban 24-29
* História
Pogányné Rózsa Gabriella: Klas August Linderfelt és "eklektikus" katalogizálási szabályzata 30-42
Fogarassy Miklós: Kódképek egy könyvtáros múltjából 42-47
* Perszónália
Bartos Éva: Meghalt Kiss Jenő (1933-2003) 48-49
* Extra Hungariam
Pátrovics Péter: Bécsi képeslapok 50-53
* Könyv
A Pest Megyei Könyvtár 50 éve 54-57
Helyismereti könyvtárosok IX. országos tanácskozása 57-60

A formátum valószínűleg nem nagyon szorul magyarázatra, de van részletes dokumentációnk is a bonyolultabb esetek felvételére.

Első sor: a folyóirat szám adatai

További sorok: szerzők kötőjellel elválasztva: cím, két szóköz és oldalszámok

*-gal kezdődő sorok: rovatok

A 142 folyóiratból jelenleg 112 közvetlen adatbevitellel készül, 30 pedig konvertálással.

Saját magunk kipróbáltuk, hogy mennyi munkával jár egy szám tartalomjegyzékének felvétele és annak hosszától függően 15-20 perc jött ki, átlagban egy cím egy perc, természetesen a szerzőkkel és oldalszámmal együtt. Tovább tart a munka, ha a folyóirat tartalomjegyzéke nem elég pontos, mert ilyenkor magát a folyóiratot is fel kell lapozni, és egybe kell vetni a tartalomjegyzéket a cikkekkel.

A kereső felület

A szolgáltatás 2003 elejétől saját domain névvel rendelkezik, és a kereső felület a www.matarka.hu címen érhető el.

A MATARKA lényegében három módon használható.



2. ábra: Nyitóoldal - Keresés ablak

1. Keresés

Keresni lehet a cíkcímekben szereplő szavakra (cím kulcsszavak), a cíkcím egy részletére és a szerzőnevek között. Ezeket kombinálhatjuk is, mivel 2 keresőmező áll rendelkezésre. Ha szerzőnévre keresünk, 1 keresőmezőbe egy szerzőt írhatunk, cím kulcsszavak közül viszont több is beírható ugyanazon keresőmezőbe, anélkül, hogy bármilyen jelet kellene közéjük tenni. Bővíthetjük a találatok számát csonkolással, amelyre a % jel szolgál. Szűkíthetjük a találatok számát azáltal, hogy megadjuk, melyik a legrégebbi év, amelyre vonatkozzon a keresés, vagy melyik szakterülethez tartozó lapokban történjen a keresés, sőt, csak egy adott folyóiratban is kereshetünk. Keresés módban azonnal a találatokhoz jutunk.



3. ábra: Keresés - Találatok listája

2. Tallózás

A cikkek címeiben szereplő szavak ill. a szerzők neveiből egy index készül. A Tallózásban megadott kereső szó alapján egy listát kapunk azokról a szavakról, melyek szerepelnek az indexekben és a megadott keresőszóval kezdődnek, valamint rögtön láthatjuk is, hogy az adott szóra hány találat van. Itt kiválasztva a nekünk megfelelő szót jutunk el a találatokig, az adott cikkekig.



4. ábra: Tallózás - címindex

3. Böngészés

A folyóiratok címeinek betűrendes vagy szakterületre leszűkített listájából kiindulva böngészhetünk az egyes számok teljes tartalomjegyzékeiben.

A fentiekből következik, hogy a MATARKA egyaránt eredményesen használható akkor is, ha a felhasználó pl. egy konkrét folyóirat legfrissebb számának tartalmáról akar tájékozódni, vagy ha egy adott szerző megjelent cikkeit kívánja összegyűjteni, vagy ha egy témához kapcsolódó cikkeket keres. S mindezt nem pusztán egy folyóiratban, vagy egy szakterülethez tartozó anyagban teheti meg, hanem általános keresést is folytathat. Mindamellet le is szűkítheti keresését idő, szakterület vagy folyóirat alapján. Ugyanakkor egy konkrét cikk pontos adatait is gyorsan megtalálhatja. Sőt, esetenként magát a cikket is, mivel az interneten megtalálható cikkek esetében egy ugrópontot találhatunk a cikk adatai mellett a teljes szöveghez.

Statisztika

Munkánk eredményességének figyelésére első lépésben március végén számlálót építettünk be a főoldalra, amely az oldalra szándékosan ellátogató vagy véletlenül eltévedőket számolja. (A fejlesztők két gépe természetesen ki van zárva a számlálásból, mert az igen torzítaná az adatokat.) De ezzel nem voltunk megelégedve, mert így nem láttuk, hogy mennyit, mit, hogyan keresnek, kihasználják-e a felhasználók a tallózási és böngészési lehetőséget is. Ezért április elejétől egy adatbázis táblában statisztikai célból letároljuk a keresések időpontját, módját és a kereső kifejezéseket. Ez nagyon tanulságos volt, mert kiderült, hogy a használóink a jól ismert kereső robotnál, a Google-nál megszokott módon teszik fel kérdéseiket (és a mi szépen leírt súgónkra úgymond fittyet hánynak). Ez arra ösztökélt bennünket, hogy a kereső felületet és a keresési lehetőségeket átalakítsuk. Vagyis most már lehet:

- Egy cím mezőbe több kulcsszót is beírni (előtte csak egyet lehetett a Google-lal ellentétben, több szó esetén több mezőt kellett kombinálni)
- Lehet címrészletre is keresni (Google-ban az idézőjeles keresés)

Ez a keresések hatékonyságát nagymértékben megnövelte.

A még mindig előforduló - sajnos nagyrészt kivédhetetlen - hibák a felhasználók részéről:

- Kevesen csonkolnak (pedig úgy sokkal több lenne a találat)
- Gyakran túl sok szót (az egész címet) írnak be néhány jellegzetes kulcsszó helyett (csak akkor kapnak találatot, ha teljesen helyesen írták be a címet)
- Vesszővel választják el a szavakat
- Kevesen használják a keresések kombinálását, de ha használják, akkor túl sokat írnak be
- Beírják a dr. szócskát a szerző neve elé
- Nem veszik észre, hogy a "Tallóz" lehetőség egy szavas indexbe visz, és több szót írnak be tallózáshoz

A hibák kiküszöbölésén úgy próbálunk segíteni, hogy a "Nincs találat" oldalon egy tömörített súgót építettünk be.

A számlálónk állása március 26-án kezdve a számolást: 27516 (2003. okt. 15.) Mivel a keresések letárolásának módján időközben többször javítottunk és ezért egységesen nagyon nehéz lenne a kiértékelés, ezért az utolsó 10 nap (okt. 6 -15) adatait összesítettük:

Összes belépés: 2537

Összes keresés: 3005

Keresés (cím): 1405

Keresés (szerző): 587

Keresés (címrészlet): 299

Tallózás: 164

Böngészés: 782

Megjegyzések a puszták számokhoz:

- Sok belépést nem követ keresés - vagyis a tényleges keresők átlagban nemcsak egy, hanem több keresést végeznek.

- Alacsony a tallózások száma, az elnevezés megválasztása lenne helytelen?

A MATARKA folyamatos fejlesztés alatt áll, és ez nem csupán a kereső felületre vonatkozik. Nemrég

készült el az új adatbeviteli rendszer. Szükséges volt az átalakítás, egyrészt mivel a jelszavas belépési rendszer nem működött tökéletesen, másrészt minél több folyóirat került feldolgozásra, annál nehezkesebbé vált a régi űrlapok használata. Ezért került sor a teljes felület átalakítására, amely most már biztonságosabb és könnyebben, kevesebb kattintással, görgetéssel kezelhető. Egyúttal igyekeztünk barátságosabbá tenni a megjelenését is.

Felhívás

A MATARKA, mint ahogy a fentiekből már kiderült, együttműködésre épülve gyarapszik. Nincs mögötte szervezet (a NKÖM kétszer támogatott minket), a munka nincs benne a könyvtárosok munkaköri leírásában, nincs "hivatalos" szerkesztő bizottság. A cikk szerzői ketten fejlesztik, és nagyon sokan építik, olyan könyvtárosok, akik hasznosnak találják és használják is ezt a fajta szolgáltatást. Ezért azzal a felhívással fordulunk a 3K olvasóihoz, ha van egy-két olyan folyóirat a könyvtárunkban, melynek feldolgozását hasznosnak találnák, csatlakozzanak hozzánk, hogy minél többen építsük együtt az adatbázist.

További tervek

Most készül a cikkek MARC formátumban történő kiexportálásához PHP script. Ennek segítségével egy kiválasztott cikk, egy adott folyóiratszám cikkei MARC formátumban kiexportálhatók lesznek. Amennyiben az idén a NKÖM-höz beadott pályázatunk sikeres lesz, elkészítjük a cikkmásolat küldő szolgáltatást is és a folyóiratok számát tovább bővítjük.