

Burmeister Erzsébet – Kiss Andrea

Miskolci Egyetem, Könyvtár, Levéltár, Múzeum

Tartalomjegyzék-szolgáltatás nyolc könyvtár együttműködésével*

Nyolc magyar könyvtár munkatársai a Miskolci Egyetem, Könyvtár, Levéltár, Múzeum vezetésével tartalomjegyzék-adatbázist építenek három népszerű szabad felhasználású szoftver (Apache webservert, MySQL relációs adatbázis-kezelő és PHP HTML-be ágyazható parancsértelmező nyelv) segítségével, amelyben több mint 100 magyar kiadású folyóirat tartalomjegyzékét dolgozzák fel. Magyarországon ez az első olyan tartalomjegyzék-szolgáltatás, amely nemcsak a folyóiratszámok tartalomjegyzékének böngészését, hanem a szerző és/vagy a címben levő szavak keresését is lehetővé teszi. Elérhetősége: <http://marki.lib.uni-miskolc.hu/kezdo.php>

Bevezetés

A tudományok legfrissebb eredményeiről, a műszaki fejlődés újdonságairól köztudottan leggyorsabban a szakfolyóiratokból és ma már az internetről szerez tudomást a kutató, a könyvtáros és az érdeklődő ember. Az interneten a megbízható (lektorált, ellenőrzött) információk többnyire előfizetéses adatbázisokon keresztül érhetők el. Közöttük fontos szerepet játszanak a folyóiratok cikkei és egyéb kiadványokat (monográfiák, konferenciák, kutatási jelentések stb.) bibliográfiai módszerekkel feldolgozó adatbázisok, amelyek sok esetben vagy maguk is tartalmazzák a dokumentumok teljes szövegét, vagy különböző interfészekon keresztül biztosítják ezekhez a hozzáférést.

Magyarországon is sok, szakfolyóiratok cikkei feldolgozó adatbázis létezik (a Fővárosi Szabó Ervin Könyvtár bibliográfiai adatbázisai, az Országgyűlési Könyvtár Pressdoc és Hundok sajtófigyelő adatbázisa, a BME OMIKK műszaki cikkeket feldolgozó adatbázisai stb.). Ezek az adatbázisok mind nagyon jól használhatók a tájékoztatásban, de mindegyik csak egy-egy részterületet dolgoz fel.

Az alaposabb feldolgozású bibliográfiai adatbázisok általában szelektáltan dolgozzák fel a folyóiratokat, de feltétlenül csak a cikkekre koncentrálnak. Ezzel szemben az egyszerűbb leírású tartalomjegyzék-adatbázisokba egy folyóirat teljes tartalma bekerül, olyan részek is, amelyek teljes egészében kimaradnak az előzőekből.

A mi célunk egy olyan adatbázis létrehozása volt, amely a gyors tájékozódásra és tájékoztatásra jó, és felöleli a magyar kiadású folyóiratok egyre nagyobb hányadát. A bibliográfiai adatbázisok építése sok munkát és komoly könyvtárosi szaktudást igényel, emiatt többnyire csak szűk szakterületre korlátozottan valósítható meg. Ezért úgy döntöttünk, hogy a folyóiratok cikkei nem bibliográfiai, hanem tartalomjegyzék szinten dolgozzuk fel. Így egységnyi idő alatt sokkal nagyobb mennyiségű anyag dolgozható fel. Mivel a cikkek címének többsége utal a témájára, a gyors tájékozódáshoz gyakran elegendő a címben való keresés. A szerző(k), a cím és a folyóirat ismeretében a tájékozott könyvtáros már el tudja dönteni, hogy ajánlhatók-e a szerző és/vagy cím kulcsszavas kereséssel megtalált cikkek az adott témában. Szemünk előtt a SwetsBlackwell közismert SwetsnetNavigator tartalomjegyzék-szolgáltatása lebegett, amely most az Oktatási Minisztérium EISZ (Elektronikus Információ Szolgáltatás) programja keretében a felsőoktatási és tudományos kutatóintézetekben elérhető (<http://www.om.hu/eisz/>). A SwetsnetNavigator több mint 17 000 folyóirat tartalomjegyzékét dolgozza fel. Sajnos magyar kiadású folyóirat nincs közte.

Magyar kiadású folyóiratok tartalomjegyzékerei rengeteg helyen megtalálhatók az interneten,

*A cikk anyaga részben elhangzott előadás formájában az *Informatika a felsőoktatásban 2002* konferencián, Debrecenben.

- a kiadónak (egy példa: a Műszaki és Természettudományi Egyesületek Szövetsége = MTESZ tagegyesületeinél, pl. az Országos Magyar Bányászati és Kohászati Egyesületnél, <http://www.ombkenet.hu/>),
- a könyvtáraknak (két példa: a BKÁE Egyetemi Központi Könyvtár, <http://www.lib.bke.hu/elfir2.html#mgf>, Országos Széchényi Könyvtár, <http://www.oszk.hu/szerv/ki/tartj/tartjegy.htm>) a gondozásában.

A MEK (Magyar Elektronikus Könyvtár) is gyűjti az internetes tartalomjegyzék-szolgáltatásokat (<http://www.mek.iif.hu/porta/virtual/magyar/efolyir/toc/link>). A legtöbb ilyen szolgáltatás azonban csak statikus HTML oldalakat kínál, ahol egyes folyóiratszámok tartalomjegyzéke böngészhető, de szerző és cím szerinti keresés az összes szám tartalomjegyzékében egyszerre nem lehetséges.

A szolgáltatás története

2001-ben az NKÖM (Nemzeti Kulturális Örökség Minisztériuma) Olvasás éve telematikai pályázatán tartalomjegyzék-szolgáltatás létrehozását pályáztuk meg, és nyertünk egymillió Ft támogatást. Célunk akkor az volt, hogy 15 magyar nyelvű és kiadású tudományos (műszaki és természettudományi) folyóirat tartalomjegyzékét dolgozzuk fel havi frissítéssel, és 10 évre visszamenően is saját erőből. Vagyis ingyenesen felhasználható szoftvereket (Apache, MySQL, PHP) alkalmazunk UNIX operációs rendszer felett, és magunk készítjük el az adatbeviteli, illetve lekérdező programokat.

2002. január elején kezdtünk el dolgozni. Az alábbi feladatokat kellett elvégezni:

- a szolgáltatás céljának meghatározása;
- az adatbázis megtervezése, létrehozása;
- a beviteli űrlapok és programok elkészítése;
- a folyóiratok kiválasztása;
- a beviteli segédlet összeállítása;
- a könyvtárosok betanítása, a beviteli munka megszervezése, lebonyolítása.

A Miskolci Egyetem, Könyvtár, Levéltár, Múzeum 12 könyvtárosa február végén már elkezdte az adatbevítelt. A pályázat teljesítésének határideje szorított, igyekezni kellett. Közben folyamatosan finomítottuk a szoftverünket, elkezdtük a lekérdező végző programok megírását. Szolgáltatásunk májusban már működött. Azóta a szolgáltatás továbbfejlesztése folyik:

- A keresési lehetőségek finomodtak: idő (év) szerinti szűkítés, szakterület szerinti szűkítés, a kiválasztott folyóiratra való szűkítés.
- Ha a cikk teljes szöveggel megtalálható az interneten, akkor ugrópont vezet a teljes szöveghez (Közgazdasági Szemle, Műszerügyi és Méréstechnikai Közlemények).
- Konvertálóprogram készült a már HTML formátumban meglévő tartalomjegyzékek bevitelére: így kerül be folyamatosan az adatbázisba a BKÁE Egyetemi Központi Könyvtár által feldolgozott 22 gazdasági folyóirat tartalomjegyzéke (a cikk írásának időpontjában 4 folyóirat bevitele már megtörtént).

2002-ben újból pályáztunk könyvtárunk az NKÖM-nél, de most már nyolc együttműködő könyvtár nevében, és 2,5 millió forint támogatást kaptunk. A nyolc könyvtár vállalja 2003. március 31-ig 75 további folyóirat tartalomjegyzékének 10 évre visszamenő bevitelét. Így a folyóiratok száma $15+22+75=112$ -re emelkedik. A közös adatbevítelt az alábbi könyvtárak végzik:

- *Berzsenyi Dániel Főiskola Központi Könyvtára* (Szombathely)
- *Budapest Főváros Levéltára* (Budapest)
- *Márton Áron Szakkollégium* (Budapest)
- *Miskolci Egyetem, Könyvtár, Levéltár, Múzeum* (Miskolc)
- *MTA Nyelvtudományi Intézet* (Budapest)
- *Nyíregyházi Főiskola* (Budapest)
- *SZTE Szegedi Élelmiszeripari Főiskolai Kar* (Szeged)
- *Váti Magyar Regionális Fejlesztési és Urbanisztikai Közhasznú Társaság könyvtára* (Budapest)

A távolságok ellenére a közös munka e-mail és telefon segítségével könnyen megvalósítható. A feldolgozó könyvtárosoknak két részletes segédlet készült:

- Feldolgozási szabályok.
- Útmutató a tartalomjegyzék-szolgáltató program adatbeviteli űrlapjainak használatához.

A felhasznált szabad szoftverek

Apache. A weboldalakat szolgáltató webszerver. Honlapjuk: <http://www.apache.org>.

MySQL. A világ legnépszerűbb nyílt forráskódú relációs adatbázis-kezelője, több mint kétfélmillió helyen telepítették. A MySQL AB svéd cég készíti, és ingyenesen hozzáférhetővé teszi. Honlapjuk: <http://www.mysql.com>.

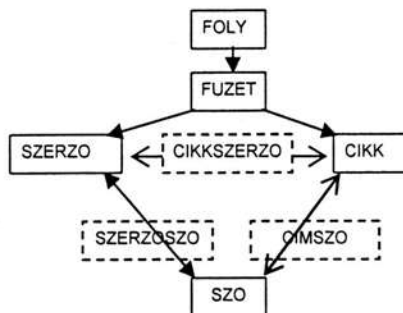
PHP. Eredeti neve: Personal Home Page Tools, mai neve: Hypertext Preprocessor. Olyan önállóan használható programozási nyelv, amely képes nagyméretű webes adatbázis-alkalmazások működtetésére is. Tulajdonképpen kiszolgáló oldali programozási nyelv, amit jellemzően HTML oldalakon használnak. A hagyományos HTML oldalakkal szemben azonban a kiszolgáló a PHP parancsokat nem küldi el az ügyfélnek, azokat a kiszolgáló oldalán a PHP értelmező dolgozza fel. A programokban lévő HTML elemek érintetlenül maradnak, de a PHP kódok lefutnak. A kódok végezhetnek adatbázis-kezelést (feltöltés, lekérdezés), dinamikusan létrehozhatnak képeket, fájlokat olvashatnak és írhatnak. Honlapjuk: <http://www.php.net>.

Az Apache + PHP + MySQL szoftverek együttese közkedvelt megoldásnak számít, igen sok honlap készül ezzel a technikával. Az Apache és a PHP biztosítja a stabil HTTP szerveret és a programozási nyelvet, mellyel dinamikus weboldalakat készíthetünk, míg a MySQL adatbázis-kezelő rendszer az adatok tárolását teszi lehetővé.

Az adatbázis létrehozása és a PHP programok elkészítése

A legfontosabb feladat bármilyen alkalmazás létrehozásánál az alapos, átgondolt tervezés. A cél ismeretében legelőször az adatbázist kell megtervezni. Az 1. ábrán láthatók az adatbázis táblái és a

köztük levő kapcsolatok. A táblák felvétele nem tart sokáig, de a táblák szerkezetének, egymással való kapcsolatának kialakítása komoly megfontolást kíván. Ezzel sok későbbi bosszúságot és felesleges munkát takaríthatunk meg.



1. ábra. Az adatbázis tábláinak kapcsolata

A szaggatott vonallal rajzolt táblák kapcsolótáblák, azonosító számokat rendelnek egymáshoz, például a cikkszerzo tábla minden sora egy cikk számához hozzárendeli a szerzőinek számát. Egy cikkhez maximum 5 szerzőt engedünk meg, vagyis egy cikkszámhoz maximum 5 sor tartozhat a táblában.

Az 1. táblázat mutatja a relációs adatbázis tábláit. A vastag betűs mezőnevek egyedi azonosítók, ún. kulcsok, amelyekre a tábla indexelve van, vagyis az adatok gyorsan kereshetők.

1. táblázat

A Tartalom adatbázis táblái

A tábla neve	mező	mező	mező	mező	mező	mező	mező	mező	mező	mező	mező
Foly	fsz	fcim	ISSN	sorozat	kozre- ado	bib- leiras	konyv- tar	szak	fcim- besor		
Fuzet	fuzetsz	fsz	vol	ev	szam						
Cikk	cikksz	fuzetsz	cikkcim	cimfoly	elso_ oldal	sorszam	ev	url	file- meret	file- format	tipus
Cimszo	cikksz	szosz									
Szerzo	szerzosz	szerzo- nev	n_ szerzo								
Szerzoszo	szerzosz	szosz									
Cikkszerzo	cikksz	szerzosz	rend								
Szo	szosz	szo									

Az adatbázis folyamatosan növekedik. A cikk írásának időpontjában (2002. december) *tartalma*: 41 494 cím (ebből 33 988 címhez tartozik szerző is), 17 423 szerző, 43 170 cikkszerző kapcsolat, 63 417 szó, 3653 ugrópont teljes szövegre.

Eddig több mint 50 php, illetve inc kiterjesztésű fájl készült. Van még ezenkívül 8 jpg fájl az egyedi, csinos megjelenítéshez. A beviteli programok elkészítése sokkal több munkát vett igénybe, mint a lekérdezések programjai. Szükség volt segédprogramokra is a konvertáláshoz, hibakereséshez, indexeléshez stb.

Adatbevétel

Először készültek el az adatbevételhez szükséges HTML oldalak a beágyazott PHP sorokkal, hogy a könyvtárosok elkezdhessék a munkát. A folyóira-

tok, a kötetek, füzetek (vagy számok) és cikkek felvehetőek, módosíthatók, törölhetőek. A 2. ábrán látható a cikkbeviteli űrlap.

Lekérdezés

A bevitt tartalomjegyzékek egyrészt böngészhetőek a folyóirat adott kötetének adott száma szerint, vagy kereshetők a címben levő szavak, illetve a szerzők szerint. A lekérdező oldalon kétféle keresési lehetőségre van mód. Egyszerű kereséssel a szerzők és kulcsszavak böngészhetőek, majd a hozzájuk tartozó cikkek megjeleníthetők, összetett kereséssel a szerző(k) neve és/vagy a címben levő szavak kombinálhatók, és úgy képezhetőek a találati halmazok. Az összetett keresési módnál lehet idő, szakterület és folyóirat szerint szűkíteni. Egyszerű keresésénél csak idő szerint lehet szűkíteni. A 3. ábra egy találati lista felső részét mutatja.

Cikk adatai

Közigazgatási szemle / Magyar Tudományos Akadémia (Budapest). Közigazgatástudományi Bizottság, ISSN 0023-4488. 48. évf. 11. sz. / 2001

Sorszám:

Szerző1: Szerző2:

Szerző3:

Szerző5:

Rovat:

Cimzatok:

Oldalszám:

Teljes szöveges elérés:

URL:

Típus:


Fajlméret:

Fájlformátum:

2. ábra Cikkbeviteli űrlap

**Munkaki Egyetem
Könyvtár, Levéltár, Múzeum**

FOLYÓIRATOK TARTALOMJEGYZÉKE


MINISZTERISÉG

Szerzők: *Kolló János, Nagy Gyula*
Berek a munkanélküliség előtt és után. Teljes szöveg (PDF)

*In: Közigazgatási szemle, ISSN 0023-4346
1995 (42. évf.) 4. sz. 325. old.*

Szerzők: *Kertesi Gábor, Kolló János*
Reálberek és keresési egyenlőségek, 1986-1996. A bérszerkezet alakulása Magyarországon. Létesítési szöveg (PDF)

*In: Közigazgatási szemle, ISSN 0023-4346
1997 (44. évf.) 7-8. sz. 612. old.*

3. ábra Szerzőre keresés találati oldala

A folyóiratok feldolgozása

Elvek

A feldolgozás során ragaszkodtunk a tartalomjegyzék sorrendiségéhez, akkor is, ha az nem követte a cikkek oldalszám szerinti sorrendjét az adott folyóiraton belül. A tartalomjegyzékben szereplő *rovatcímeket* először elhagytuk, de most már lehetséges a rovatcímek felvétele is. Az *oldalszámokat* viszont sosem a tartalomjegyzék szerint vettük fel, hanem a folyóirat átvizsgálása során állapítottuk meg a tényleges oldalszámot. A *szerek* nevét nem egységesítettük.

Tapasztalatok

Pozitívumok

- A felhasznált szoftverek ingyenesen elérhetők. A kereséshez és adatbevitelhez egyaránt elegendő az internetkapcsolat és egy böngészőprogram. Ezáltal az adatbázis bővítésébe könnyen bevonhatók más intézmények, a szolgáltatás akár országos jellegűvé is tehető.
- Az adatok bevitele a felhasznált módszerekkel rendkívül gyors, a bevétel módja egyszerű, könnyen elsajátítható, nagyobb szakképzettséget nem igényel.
- A kulcsszó szerinti keresés biztosítja, hogy egy adott témában kereső felhasználó nagy mennyiségű folyóirat átvizsgálása nélkül is megtalálja a témába vágó cikkek halmazát.

Negatívumok

- A szolgáltatás csupán kiindulópontot nyújt egy adott témában való kereséshez, de a találati halmaz leszűkítése a valóban hasznos információkra további helyben történő kutatást igényel.
- A webes bevétel miatt a vegyjelek hű tükrözése nem megoldott. Az idegen nevekben esetlegesen használt speciális karakterek ugyan unicode-os formában tükrözhetők, viszont az ilyen nevek nem kereshetők.
- A gyors adatbevitel nem jelent feltétlenül pontos és egységes bevittet. Hibaforrások pl.: elírások, helyesírási hibák, adatkihagyások, problémás esetek nem egységes kezelése, alcímek és párhuzamos címek nem szabályszerű felvétele (ha nem képzett könyvtáros a feldolgozó).

Problémák

A feldolgozás során az alapvető gondot az okozza, hogy a tartalomjegyzékek messzemenően nem tükrözik az adott folyóirat tartalmát.

El kell döntenünk, hogy a tartalomjegyzékhez szigorúan ragaszkodva felvállaljuk-e az előforduló (nem kevés!) pontatlanság átvételét, avagy teljes egészében a folyóirat tartalmához igazodva történjen az adatfelvétel. Utóbbi esetben azonban lényegesen több cikk (köztük számos, információforrásként lényegében használhatatlan kisebb hír) lenne feldolgozva, ami nagyságrendekkel meghosszabbítaná a feldolgozás folyamatát, és irreleváns információkkal telítve az adatbázist, megnehezítené a keresést.

A harmadik lehetőség, amihez végül mi folyamodtunk, szintén számos buktatót rejt magában. E módszer lényege, hogy a tartalomjegyzékhez ragaszkodtunk annyiban, hogy csak azokat a cikkeket dolgoztuk fel, amelyeket ott feltüntettek, és abban a sorrendben, ahogy ott szerepeltek. De az ezekben az adatokban szereplő hibákat igyekeztünk kijavítani. Gyakran azonban nehéz eldönteni, hogy mennyiben szükséges tükrözni a tartalomjegyzéket, és mennyiben célszerű az adatok pontosítására időt és energiát fordítani.

A felbukkanó problémák alapvetően három csoportba oszthatók:

1. A tartalomjegyzék kevesebb, mint a tényleges tartalom:
 - pl. kimarad egy szakcikk a tartalomjegyzékből;
 - az adott rovatához tartozó cikkek közül csak egyet tüntet fel vagy nem mindegyiket;
 - egy rovat több helyen is megjelenik a lapon belül, de csak egy oldalszám van hozzá feltüntetve,
 - csak rovatcím van feltüntetve, a hozzá tartozó, releváns információt tartalmazó cikkcímek pedig nem;
 - különszám (akár belekötve is egy rendszeres számba): a különszám címét a tartalomjegyzék csak egyszer tünteti fel, holott valamilyen módon jelezni kellene, hogy a megfelelő cikkek ehhez tartoznak;
 - a szerző neve csak rövidített alakban van feltüntetve;
 - csak a rovatcímhez van rendelve oldalszám, a hozzá tartozó címekhez (noha fel vannak tüntetve) nincs.

2. A tartalomjegyzék több, mint a tényleges tartalom:

- olyan cíkcím is szerepel a tartalomjegyzékben, amely nem található meg a folyóiratban;
- „semmitmondó” rovatcímek (Egyesületi hírek, Szemle);
- feltünteteti az idegen nyelvű tartalomjegyzékeket is;
- egy vagy több idegen nyelven is feltünteteti a cikk címét;
- a szakirodalmi ismertető címe megtévesztő lehet: a cikk csupán az írást ismerteti, a téma, amelyre a cím utal, nincs részletesen tárgyalva.

3. A tartalomjegyzék más, mint a tényleges tartalom:

- a cíkcímek sorrendje más, mint a tényleges;
- a cikk címe teljesen vagy csak néhány szóban eltérő;
- a cikk címe csak magyarul, vagy több nyelven is fel van tüntetve, a cikk pedig idegen nyelvű (ez különösen súlyos gond, mert megtévesztő lehet a felhasználó számára, aki esetleg olyan cikk után kutat a szolgáltatás alapján, amelynek szövegét végül nem is érti meg);
- a tartalomjegyzékben hibásan (elírás vagy helyesírási hiba) van feltüntetve valamelyik adat.

Összefoglaló

A szolgáltatást a közelmúltban megkereszteltük a szolgáltatás további kibővítését tervezzük, hiszen az adatbevitel nagyon gyorsan megy, és a felsorolt negatívumok ellenére a szolgáltatás sok mindenre használható:

- gyors irodalomkutatás adott témában magyar nyelvű cikkekben,

- adott szerző cikkeinek összegyűjtése,
- adott folyóirat számai tartalomjegyzékének böngészése.

A szolgáltatást a közelmúltban megkereszteltük, hogy a felhasználók könnyebben megtalálhassák a kezdőlapot az interneten. A domainnév regisztrálása folyamatban van. Így az elérhetőség hamarosan megváltozik: www.matarka.hu. A MATARKA név feloldása: MAgyar TARtalomjegyzékek KEReshető Adatbázisa. Eredetileg a TARKA nevet szerettük volna, de ezt már mások más célra lefoglalták.

Mivel adatrögzítésre nem mindig lehet pályázni, ezért egyre nagyobb hangsúlyt szeretnénk fektetni az automatizálásra, vagyis már digitális formában meglévő tartalomjegyzékek konvertálására, illetve fel szeretnénk venni minél több kiadóval a kapcsolatot, hogy ők – már gyakran a megjelenés előtt – küldjék el nekünk megfelelő formátumban a tartalomjegyzékeket, melyeket azonnal be tudnánk vinni az adatbázisba, és azok így már a megjelenés előtt elérhetővé válnának.

Irodalom

LÁSZLÓ József: Dinamikus weboldalak, CGI programozás Windows és Linux rendszereken. 2002.

ZANDSTRA, Matt: Tanuljuk meg a PHP4 használatát 24 óra alatt. 2001.

MOULDING, Peter: PHP haladóknak. Fekete könyv. 2002.

TÓTH Gábor: Ingyen hozzáférhető tartalomjegyzék-szolgáltatások a Weben. <http://www.mek.iif.hu/porta/szint/tarsad/konyvtar/forras/tocs/> = Tudományos és Műszaki Tájékoztatás, 46. köt. 3. sz. 1999. p. 91–100.

Beérkezett: 2002. XI. 4-én.

Kazinczy Online 2002 díj

Második alkalommal rendezték meg a legjobb magyar online és offline multimédiás megjelenések – tehát CD-ROM-ok, DVD-k, weboldalak – versenyt. A novemberben tartott eFestivalon négy kategóriában hirdettek győztest. A nyelvhelyességi verseny győztese a www.konyvkereso.hu című honlap. A legjobb weboldalakat benevezték a Cannes-ban minden évben megrendezendő nemzetközi versenyre is.

/Népszabadság, 2002. november 27. p. 26./