

Drótos László

eleMEK

Metaadat-kezelő rendszer digitális gyűjteményekhez

Összeáll-e valaha is „világkönyvtárrá”, „tudásbázissá” az az óriási dokumentum- és információhalmaz, amely a weben elérhető? Vagy már örökre lekéstünk róla? Mindenesetre a szemantikus web kitalálói még reménykednek benne, és a hazai Nemzeti Digitális Adattár létrejötte is egy fontos lépés ebbe az irányba. Ezek a rendszerek a digitális objektumokat kísérő metaadatokon alapulnak, csakhogy ilyenek nálunk alig-alig vannak, különösen nem szabványos adatsere-formátumban. Olyan szoftverekre van tehát szükség, amelyekkel ezek a metaadatok könnyen előállíthatók, konvertálhatók, továbbadhatók más rendszereknek. Lehetőleg ingyenes programokra, hogy minél többen használják őket.

Amikor a nemrég lovaggá ütött *Tim Berners-Lee* a kilencvenes évek elején kitalálta a webet, maga sem sejtette, hogy milyen szellemet enged ki a palackból. Az eredetileg elosztott dokumentumkezelő rendszerből az írásos kommunikáció legforradalmibb eszköze lett a nyomtatás megjelenése óta. És a legnagyobb információs káosz is a civilizáció történetében. Az immár 8 milliárd weblapon kereső Google 40 ezer találattal önti nyakon a használóját a „Neumann János” kérdésre, és nem lehet megmondani neki, hogy csak azt a néhány digitalizált könyvet és cikket kérjük, amelyeknek a híres matematikus az írója. Nem véletlen hát, hogy egy évtizeddel a korszakalkotó találmánya után *Tim Berners-Lee* egy továbbfejlesztett, „jelentéssel bíró” világháló kidolgozásán és elterjesztésén fáradozik. Az ígéretek szerint a W3 Konzorcium által összefoglalóan „szemantikus web”-nek nevezett új technológiák jelentik majd azt az Ariadné-fonalat, amely kivezet minket az információs labirintusból – lehetővé teszi az értelmes keresést az interneten, összekapcsolhatóvá a legkülönbözőbb online alkalmazásokat, és eddig megvalósíthatatlanak tűnő „intelligens” szolgáltatásoknak teremti meg az alapjait. Az új technológiák azonban nagyrészt csak a nevükben újak; a dokumentalisztika, a könyvtártudomány és az általános információtudomány már a hagyományos dokumentumok esetében szembesült a hatékony információkeresés problémájával, és többek közt a dokumentumtipológia, az osztályozás, a bibliográfiai leírás, az analitikus feltárás és a teauruszok fegyvereit állította csatasorba. Ezek térnek most vissza „metaadat-sémák”, „webontológia” és hasonló nevek alatt, ezekkel próbálja a W3C – az előbbi hasonlatnál maradván – felvenni a tíz éve elvesztett fonalat.

Ha a web fejlesztői nem is voltak elég előrelátóak, és nem építették be már a kezdet kezdetén kötelező jelleggel a hatékony visszakereséshez szükséges elemeket a rendszerbe, legalább a könyvtáros szakemberek élen járhattak volna az opcionálisan azért már elég régóta rendelkezésre álló, a megtalálhatóságot segítő megoldások alkalmazásában. Mégis, ha körülnézünk a magyar interneten, alig találunk olyan könyvtári oldalt, ahol például a leg-egyszerűbb Dublin Core metaadatok benne lennének a dokumentumok fejlécében, vagy volna valamilyen stabil hivatkozhatóságot, megtalálhatóságot biztosító azonosítjuk (URN, URI vagy PURL). A legtöbb közgyűjtemény megelégszik azzal, hogy a nyilvánosságnak szánt digitalizált dokumentumait kiteszi a honlapjára, a többit meg rábizzá a Google-ra. Jobbik esetben készít róluk egy-egy leírást a saját könyvtári rendszerében, amelyek így ugyan elvben a webOPAC-on át az interneten is visszakereshetővé válnak, de a gyakorlatban az átlagfelhasználók számára elvesznek a nyomtatott könyvek tömegében, és egy adatbázisba zárva a keresőgépek számára is elérhetetlen mélységekbe kerülnek. Az elmúlt években örvendetesen megsaporodó hazai digitalizálási pályázatok és projektek is csak a digitális másolatok előállítását mint végcél szorgalmazzák, és nem terjednek ki ezek szolgáltatásának, visszakereshetőségének módjára. A 2003-ban indult *Nemzeti Digitális Adattár* (www.nda.hu) az első fontos lépés az intézményi honlapon való elektronikus publikálástól a profeszszionális digitális archívumok és az együttesen kereshető elektronikus könyvtárak felé vezető úton. Az NDA felvállalt feladatai közé tartozik többek között szabványos névterek, vagyis besorolási adatok (pl. testületi nevek, földrajzi nevek) létreho-

zása és az Open Archives Initiative (OAI) ajánlásai alapján a magyar interneten levő digitális dokumentumok metaadatainak begyűjtése és kereshetővé tétele. Már ha lennének ilyen begyűjthető metaadatok...

A Magyar Elektronikus Könyvtár 1994 óta élen jár hazánkban a digitális dokumentumok kezelésében és szolgáltatásában. 1995 óta a Nemzeti Információs Infrastruktúra Program, 1999-től pedig az Országos Széchényi Könyvtár keretében folyik a fejlesztése, és ezeknek a támogatásoknak köszönhetően 2001 és 2003 között elkészült egy minden tekintetben korszerűnek mondható, a nemzetközi könyvtári és internetes szabványokhoz és szokásokhoz igazodó, több technikai újdonságot elsőként honosító új keretrendszer: a MEK 2-es verziója, amely a *mek.oszk.hu* címen működik, és az egyik leglátogatottabb hazai tartalomszolgáltatás az interneten. A MEK gyűjteménye az elsők között vált elérhetővé az NDA közös keresőjében, de a különböző adatcsere-formátumokban letölthető metaadatoknak köszönhetően dokumentumaink visszakereshetők a MOKKA-ban, a Kistéka-ban és néhány más könyves adatbázisban is. És természetesen a Google is indexeli a gyűjtemény metaadatait a dokumentumok teljes szövege mellett – újabban már 2-3 napos gyakorisággal.

A több mint kétéves munka során összegyűlt rengeteg tapasztalatot szeretnénk volna közkinccsé tenni, így merült fel egy – a MEK „maximalista” igényeihez képest – egyszerűbb, rugalmasan konfigurálható, platformfüggetlen, moduláris keretrendszer kifejlesztésének ötlete. Az *eleMEK* projekt 2003 végén indult, és megvalósítását az IHM, az NIIF és a MEK Egyesület támogatja. Az 1.0-s verzió 2004 első felében készült el, jelenleg már ennek továbbfejlesztése folyik. Az *eleMEK* GNU GPL licenc alatt terjesztett szabad szoftver, amelyet nonprofit célokra bármely hazai és határon túli magyar intézmény, illetve magánszemély ingyenesen használhat (például az Erdélyi Magyar Műszaki Tudományos Társaságnál már üzemszerűen működik). Mivel a Java forráskód is hozzáférhető, ezért akár a saját igényeinek megfelelően tovább is fejlesztheti, azzal a feltétellel, hogy a bővítéseket elérhetővé teszi a többiek számára is a projekt honlapján.

Az *eleMEK* szolgáltatásai a következők:

1. A digitális dokumentumok metaadatainak rögzítése (1. ábra)

A rendszer alkalmas a legkülönbözőbb típusú (szöveges, képi, hang stb.) digitális gyűjtemények nyilvántartására; összesen 23 adatcsoportban mintegy

The screenshot shows a web browser window with the URL *mek-2-001644* and the page title *Tárgyszó*. The main content area is a metadata entry form with the following fields and buttons:

- Cím:** Tárgyszó / kulcszó: nyomdászat története [INDEX]
- Szerző:** [INDEX]
- Testületi szerző:** [INDEX]
- Közreműködő:** [INDEX]
- Testületi közreműködő:** [INDEX]
- Kiadó:** Sárospatak [INDEX]
- Dátum:** [INDEX]
- Név:** [INDEX]
- Sorozat:** [INDEX]
- Eredeti kiadvány:** [INDEX]
- Dokumentumtípus:** 17-20. sz. [INDEX]
- Származási hely:** [INDEX]
- Jogkezelés:** [INDEX]
- Téma:** [INDEX]
- Tárgyszó:** [INDEX]
- Leírás:** [INDEX]
- Kapcsolatok:** [INDEX]
- Dokumentumazonosító:** [INDEX]
- Formátum:** [INDEX]
- Verziószám / kiadásjelzés:** [INDEX]
- Megjegyzés:** [INDEX]
- Státusz:** [INDEX]
- Feldolgozó:** [INDEX]

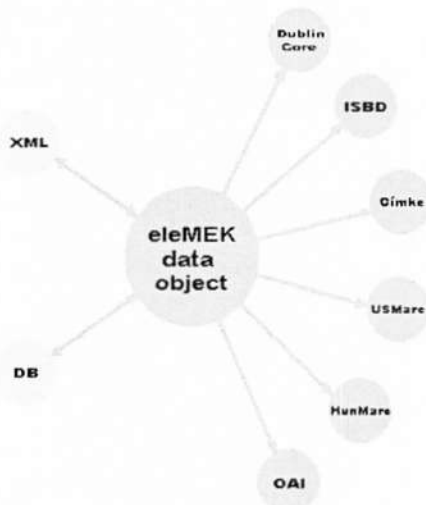
At the bottom of the form are buttons for **MEHET** and **TÖRLÉS**. On the right side, there is a **Törlés** button for each field and a **Tézauszus linkek** section with links to **OSZK**, **Geotaurusz**, **Matriksz**, **MOKKA**, and **OSZK**.

1. ábra

száféle adat írható le egy-egy dokumentumról. Természetesen ezeket nem kell mind kitölteni, mindössze csak 4 kötelező adatmező van. Egyes mezőkhöz kötött listák definiálhatók az egységesítés érdekében (pl. témakörök, nyelvek, dokumentumtípusok); telepítéskor ezek a MEK-ben használt adatokkal vannak feltöltve, de igény szerint bővíthetők, átírhatók. A tárgyszavazáshoz, a szabványos személy- és intézménynevekhez külső teauruszok és névterek linkelhetők. A rögzített metaadatok XML állományokban, illetve opcionálisan egy SQL adatbázisban is tárolhatók. Utóbbi nagyobb, több ezer tételes archívumokhoz ajánlott a gyorsabb visszakeresés érdekében.

2. Keresés a metaadatok között

Az eleMEK tartalmaz egy gyorskereső modult, amellyel a legfontosabb adatok (szerző, cím, téma) alapján egyszerű keresések végezhetők. Biztosít továbbá egy minden metaadatra kiterjedő, összetett kérdésekre is használható könyvtárosi katalógust, valamint egy rugalmasan konfigurálható felhasználói OPAC-ot. Lehetőség van különböző eleMEK adatbázisokban való együttes keresésre is. A találati listákban megjelenő dokumentumokhoz való hozzáférés IP-cím vagy digitális aláírás alapján az egyes dokumentumok szintjén egyedileg korlátozható, így nem nyilvános tételeket tartalmazó archívumokhoz is lehet használni.



2. ábra

3. Metaadatok importja és exportja (2. ábra)

Az eleMEK képes XML formátumban importálni metaadatokat, vagyis átvehetők az esetleg korábban már más rendszerben készített dokumentumleírások. Képes továbbá exportálni a bevitt adatokat különböző szabadon definiálható kimeneti (cédula-, illetve címke-) formátumokba, valamint szabványos adatsere-állományokba (XML, Dublin Core, USMARC és HUNMARC). Az NDA által használt OAI protokollhoz az illesztés folyamatban van (3. ábra).



3. ábra

4. Állománykezelő és karbantartó funkciók

Az előzőekben ismertetett főbb modulok mellett az eleMEK több kiegészítő elemet, segédprogramot is tartalmaz. A rendszer működését, a kimeneti formátumok és a kötött listák konfigurálását biztosító modulon kívül vannak benne hibakeresésre, linkellenőrzésre, biztonsági mentésre, illetve globális adatmódosításra használható segédprogramok, valamint egy, a gyűjtemény összetételét mutató statisztikai modul. Tovább illeszthetők hozzá különböző ingyenes, Java alapú keresők is a dokumentumok szövegében való kereséshez. Az egyes modulokhoz rendszerdokumentáció tartozik, az adatbeviteli modulhoz pedig egy használati utasítás is, amely a mezők helyes kitöltését magyarázza el.

Az eleMEK nem tartalmaz felhasználó felületet az archívum szolgáltatásához. Ezt minden alkalmazó stílusállományok segítségével maga alakíthatja ki a saját igényeinek megfelelően, vagy beépítheti az eleMEK szolgáltatásait a már meglévő kezelőfelületébe.

Mivel a háttérben futó programok Java nyelven íródtak, így az eleMEK egyaránt telepíthető Linux és Windows alapú szolgáltató gépekre. A telepítésben és a rendszer használatában természetesen segítséget nyújtunk, ha valaki ezt igényli. Min-

den fontos információ, a teljes dokumentáció, valamint a letölthető programkód megtalálható a projekt elemek.oszk.hu honlapján. Egy szabadon használható teszt- és oktatási változat elérhető az elemek.niif.hu oldalon.

Irodalom

DRÓTOS László–PERLAKI Attila: Moduláris könyvtári rendszer elektronikus dokumentum-gyűjtemények kezeléséhez: az eleMEK projekt. = Networkshop 2004 konferencia, Győr, 2004 április 5–7. <http://elemek.oszk.hu/ismertetok/netwshp2004.htm>

FÜLÖP Csaba–KOVÁCS László–MICSIK András: A metaadatsémák és a szemantikus web: egységesítés és specializáció a metaadatok világában. = Tudományos és Műszaki Tájékoztatás, 51. köt. 7. sz. 2004. p. 276–284.

Beérkezett: 2004. X. 11-én.



Drótos László

az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár osztályán főkönyvtáros. A Magyar Elektronikus Könyvtárért Egyesület elnökségi tagja.

E-mail: mekdl@iif.hu

Országos Doktori Adatbázis

Az Oktatási Minisztérium a doktori képzésben részt vevő egyetemekkel és doktori iskolákkal együttműködve, a doktori képzésről és a doktori fokozatszerzésről szóló 51/2001. (IV.3.) korm. rendelet (Doktori kormányrendelet) 40. § és 41. § előírásai alapján közkinccsé teszi a magyarországi egyetemeken a rendelet hatályba lépése (2001. április 11.) után doktoráltak dolgozatainak bibliográfiai adatbázisát és a megvédett értekezések téziseit. A rendszer folyamatos feltöltését az egyetemek/doktori iskolák és az Oktatási Minisztérium felelősei összehangoltan végzik. A doktori értekezéseket is tartalmazó teljes adatbázis 2003. január 1-jétől működik.

Elérhető az alábbi címeken, útvonalon:

<http://phd.om.hu> <<http://phd.om.hu/>>

<http://www.om.hu/phd>

www.om.hu <<http://www.om.hu/>> Főoldal / Felsőoktatás / Felsőoktatás és tudomány / Doktori képzés / Országos Doktori Adatbázis

Az adatbázis publikált része mindenki számára hozzáférhető egy ingyenes regisztráció keretében. (A regisztráció csak statisztikai adatgyűjtési célokat szolgál.)

/Katalist, 2004. december 14./

(Téglási Ágnes)