

Tapolcai Ágnes

Országos Széchényi Könyvtár MEK Osztály

Nyílt könyvtári gyűjtemények az interneten. Szabványos metaadatok: átjárhatóság*

Az internet világában a digitális formában (is) létező alkotások gyűjteményei egyre nagyobb kulturális értéket képviselnek. A világháló növekedésével egyenes arányban azonban egyre kevésbé triviális feladat egy mű, vagy akár egy gyűjtemény megtalálása. Ehhez kínál új, nemzetközi szabványokon alapuló megoldásokat az információtechnológia.

Bevezetés

A digitális könyvtárak, dokumentumarchívumok nagy emberi munkabefektetéssel, a világ különböző pontjain jönnek létre, és a hálózat azon adottsága folytán, hogy mindenhol egyaránt elérhető, természetesen merül fel az igény, hogy egyben lássuk a tartalmilag összetartozókat. Tapasztalhattuk, hogy ami nem érhető el valamilyen nagyobb rendszer keretében, ma már elvesz az internet óceánjában. Az elérhetőség továbbá azt is magával hozza, hogy nagyobb szabású feldolgozó munkát eloszthatunk számos intézmény és közönség között. Az egységes végeredmény eléréséhez azonban bizonyos közös szabványok betartása szükséges.

Néhány évvel ezelőtt a könyvtárosok között vita folyt arról, hogy legyen-e egy közös integrált könyvtári rendszer a magyar könyvtárakban vagy sem. A közös rendszer megkönnyítette volna az együttműködést, de a kivitelezés megvalósíthatatlannak bizonyult. Még az azonos rendszert használó könyvtárak körén belül sem tudtunk igazán egységes feldolgozást kialakítani. Az internet elterjedésével a kérdés úgy merül fel, hogy térben távoli, esetleg más-más nyelven feldolgozott gyűjtemények együttes kereshetőségét kellene megvalósítani. Arról persze szó sem lehet már, hogy közös rendszert találjunk, vagy megbeszéljük, hogyan is végezzük a feldolgozást, és a gyűjtemények is nagyon különbözőek lettek: a szöveges dokumentumok mellett képek, hangok, filmek, virtuális kiállítások, műtárgyak fényképei, térképek, szoftverek lehetnek az elemei.

Ebben a helyzetben a következő eszközök lehetnek segítségünkre:

- Nyílt rendszerek használata.
- XML leíró nyelv használata adatcseréhez.
- Metaadatok szabványosítása.
- Összetett rendszerek: archiválás, közös szolgáltatás.

Ezek az alábbiakban részletezett technikai megoldások összességükben olyan lehetőséget jelentenek, amely az interneten az elektronikus formában meglévő katalógusokat, teljes dokumentumokat és egyéb alkotásokat témák szerint, jól használható módon megjelenítheti. Az egyes helyi rendszerek nagy szabadsággal rendelkeznek, hogy saját igényeik szerint alakítsák belső szerkezetüket, miközben ha az átjárókat, az adatcserét biztosító közös elemeket is beleépítik, integrálódnak az országos, vagy még nagyobb szolgáltatásokba.

A nyílt rendszerek

A nyílt rendszer olyan részekből álló szoftvert jelent, amelyben a részek egységes, gyártófüggetlen, nemzetközileg elfogadott szabványokon alapuló felületet nyújtanak. Ennek következtében rugalmasan összekapcsolhatók, kombinálhatók. A különböző gyártóktól származó komponensek egymással kompatibilisek, a felhasználó minden összetevőt ott szerezhet be, ahol a legjobb terméket találja saját céljaira. Ami talán még fontosabb, az új, fejlettebb programokra is áttérhet rendszerösszeomlás vagy költséges adatkonverzió nélkül. Időtálló megoldásoknak ígérkeznek tehát a nyílt rendszerek, és természetesen képesek egymással adatokat cserélni, kommunikálni.

* Az OSZK-ban a *Digitális könyvtári kezdeményezések* c. szakmai napon elhangzott előadás alapján.

Az XML leíró nyelv

Az XML [1] a nyílt rendszerek kedvelt adatsereformátuma. Úgynevezett platformfüggetlen formátum, vagyis minden számítógép, bármilyen hardvert és szoftvert használ, tudja értelmezni. Egyszerű karakteres szöveg, teljesen egyértelmű addig, amíg csak angol karakterek vannak benne. A szövegek ebben az ábrázolási formában könnyen továbbíthatók a hálózaton, és nem csak homogén szövegfolyamként: a HTML nyelvhez hasonló, magában a szövegben elhelyezett, szintén karakteres elemek strukturálják. A formai megkötések nyílt szabványban az interneten olvashatók, azokat bármely programba beleírhatják, így módon egyetlen gyártó sem tudja kisajátítani magának, valamint komoly reményt nyújt arra, hogy a jövőbeli programok is el tudják olvasni.

Némi egyszerűsítéssel például ilyen lehet egy könyv leírása:

```
<konyv>
<szerzo>
<vezeteknev>Neil </vezeteknev>
<keresztnev>Bradley</keresztnev>
</szerzo>
<cim>Az XML kézikönyv</cim>
<kiado>Szak Kiadó</kiado>
<hely>Budapest</hely>
<ev>2000</ev>
</konyv>
```

Ideális eszköz tehát a tartalom alkalmazások közötti átvitelére. Az így elraktározott adatokat át lehet alakítani, ha más alkalmazás kéri, néha többször is. Például kiszűrhetjük a fölösleges adatalemeket, megváltoztathatjuk az elemek neveit, megjelenítési formátumot alakíthatunk ki. Ez utóbbira persze nincs szükség akkor, ha adatbázis kéri az adatot. Valójában eléggé elválík egymástól a logikai rendszer és a külső megjelenítés, ami többféle is lehet ugyanazon adathalmazról.

Az angol karakterektől eltérő betűk kódolására különböző szabványok vannak, ez nem érinti az XML szerkezetet, mindössze a fejlécben szereplő bejegyzés árulja el, hogy melyiket használtuk.

Összefoglalva, az XML formátum fő előnyei alkalmazások közti adatátvitelnél:

- platformfüggetlen, minden rendszer tudja alkalmazni, időtálló;
- nyílt szabvány;
- szöveges adat, könnyű a hálózati továbbítás;
- egyszeri adatfelvitel, sokféle formába átalakítható.

A metaadatok

A metaadatok jelentése: strukturált adatok az adatokról. Ilyenek a könyvtári katalógusok, tárgyszóindexek, újabb formájukban a weben található dokumentumok leíró adatai. A könyvtári világban bibliográfiai adatok néven ismerjük, gépi szabványuk a MARC. A web gyors növekedése sürögösen szükségessé tette, hogy nemzetközi metaadatszabványokat dolgozzanak ki minden területre, ez a folyamat 1995 körül indult el.

Az egyes szabványok között eleinte zavar támadt, mivel párhuzamosan is folytak a munkák, és akik valamilyen szabványhoz akartak igazodni, például a nyílt rendszerek gyártói, nem tudták, melyik lesz a végső. Manapság már tisztább a kép: a hasonló projektek elkezdtek együtt dolgozni, és az egyes területeken kiemelkedtek a szabványként elfogadott formátumok. A SCHEMAS projekt [2] éppen ennek a folyamatnak a rendszerezett internetes publikálását végzi. A szabványok fejlődésének folyamatát követi, mégpedig a programfejlesztők szemszögéből, akik rendszereik átjárhatóságát ezek alapján biztosítják. Az interneten folyamatosan publikálják a legújabb verziókat.

Megszülettek a domináns formátumok közti konverziók, illetve keresztábrák is (például a Dublin Core és a MARC között) a domének közti átjárhatóság biztosítására. Az átjárhatóság különböző szinteken valósulhat meg: a közel rokon gyűjtemények között részletes, mély megfeleltetés végezhető el, ugyanakkor nagyobb körben is integrálhatók a szerverek, amikor is a megfeleltetés (vagyis a keresettség) kevésbé részletes.

Az elektronikus dokumentumok leírására az igen egyszerű, mindössze 15 elemből álló Dublin Core bizonyult a legelterjedtebb szabványnak.

A Dublin Core-adatelemek [3]:

1. Title (cím)
2. Creator (alkotó)
3. Subject and Keywords (téma és kulcsszavak)
4. Description (leírás)
5. Publisher (kiadó)
6. Contributor (közreműködő)
7. Date (dátum)
8. Type (dokumentumtípus)
9. Format (fájlformátum)
10. Identifier (azonosító a hálózaton)
11. Source (származási hely)
12. Language (nyelv)
13. Relation (kapcsolatok)

14. Coverage (hatókör)

15. Rights (jogok)

Az adatelemeket XML formátumba konvertálják adatcsere céljára. Sematikusan ábrázolva ilyenforma egy lista:

```
<record-list>
<dc-record>
<title>cím</title>
<creator>első szerző</creator>
<creator>második szerző</creator>
...
</dc-record>

<dc-record>
<title>másik cím</title>
<creator>másik szerző</creator>
</dc-record>
...
</record-list>
```

Archiválás és közös lekérdezés

A fenti technológiai háttérén érdekes megoldások jöttek létre az elektronikus dokumentumok kezelésére. A probléma két fő köre: különböző szerverekről származó gyűjtemények közös lekérdezése és az archiválás/rendszerezés. A nyílt rendszerek és a szabványos metaadatok felhasználásával a közös lekérdezés jól megoldható.

Az egyik ilyen kezdeményezés, az Open Archive Initiatives (OAI) társadalmi kérdésként kezeli az akadémiai szférában létrejövő dokumentumok közkincként való kezelését, vagyis a nyilvános hozzáférést. Keretrendszert ajánl a csatlakozó intézményeknek gyűjteményeik közös kereshetőségére. Kétféle résztvevőt különböztet meg:

- data provider, adatszolgáltató szerver és
- service provider, a nyilvános szolgáltatást nyújtó szerver.

Középpontjában az OAI Metadata Harvesting Protocol áll. A dokumentum az adatszolgáltató szerverén marad, sőt itt vannak a metaadatok is, akár adatbázisban, akár más módon szervezve. A központi szerver átadja a kérést, és az adatszolgáltató szerver válaszként megadja a metaadatok listáját. A metaadatok között többnyire szerepel egy ugrópont (link) a teljes dokumentumra is. A Dublin Core elemei között az Identifier (azonosító) szolgál erre a célra. Az azonosító adatelembe optimális esetben az URL-nél állandóbbat szántak, például URN-t, DOI-t, elektronikus dokumentumra

kiadott ISBN-t vagy ISSN-t. Amíg azonban nem rendelkezik ilyennel egy dokumentum, az elérési utat jelentő URL kerül ide.

Mindazonáltal nem kötelező az elérést, vagyis az ugrópontot beletenni, az adatszolgáltató dönthet úgy, hogy csak a metaadatokat adja át, ha például a tartalom nyilvános közlésére nem jogosult. Az elv hasonló a Z39.50-hez, de annál kevésbé átfogó, ezért egyszerűbb, könnyebben telepíthető. A közös készlet a Dublin Core 15 alapeleme, XML formában. Saját adatelemeiket ehhez rendelik hozzá az adatszolgáltató szerverek, de párhuzamos megjelenítési formátumokat is készíthetnek, ha azok jobban illeszkednek a gyűjtemény természetéhez. Az alternatív formátumok adatelemeit is XML formában kell megadni.

A felépítés azt is lehetővé teszi, hogy egy kisebb, kevés technikai lehetőséggel ellátott könyvtár például ne építsen ki saját szolgáltató felületet, hanem a dokumentumok gyűjtésére koncentráljon, azt rendezett módon elhelyezze szerverén, és csak a szolgáltatást nyújtó felületen jelenjen meg. A szoftver készen van, és ingyenesen letölthető a webről. A gondosan tesztelt 2.0-s verziót 2002 nyarán hozták nyilvánosságra.

Az elektronikus dokumentumok archiválása és szolgáltatása

Az elektronikus dokumentumok archívumának kezelésére több ingyenes szoftver is elérhető. Ezek egyike az eprints.org szoftver (<http://www.eprints.org>) [4]. A szoftver két egyetemen már kísérleti jelleggel működik. A közös lekérdezést az Open Archives protokollal végzik.

Egyetemi, kutatóintézeti e-archívum létrehozása különösen hasznos, mert a legfrissebb eredmények így sokkal jobban hasznosulnak, valamint archiválásuk is megoldódik. Ahhoz azonban, hogy jól használható és ismert legyen egy archívum, kell egy bizonyos mennyiségű tartalom. Ezután már önmagától gyarapszik, mert a közismert helyekre szívesen felteszik az emberek alkotásait.

Összefoglalva: a digitális könyvtárak lehetőségei jól kibővültek a technika oldaláról, és Magyarországon is értékes anyagok gyűlnek a hálózaton. A közelmúltban a Digitális Könyvtári kezdeményezések c. szakmai napon például a már ismert Neumann-házon és Magyar Elektronikus Könyvtáron

kívül megismerhettük a Jász-Nagykun-Szolnok Megyei Verseghegy Ferenc Könyvtár legújabb eredményeit, a Debreceni Egyetemi Könyvtár ebben az évben nyilvánossá tett részeit, és az egri Bródy Sándor Megyei Könyvtár weboldalain közzétett gazdag, helyi képeslapgyűjteményt.

Az új technológiák talaján most már létrehozhatnánk egyetemi publikációs archívumokat és alternatív elektronikus folyóiratokat is, metaadatszinten összekapcsolhatnánk a meglevő hazai digitális könyvtárak gyűjteményeit, közös lekérdező felületet vagy közös elosztott adatbázist lehetne csinálni a magyar cikkbibliográfia és a TOC adatbázisokhoz; és még ki tudja hány speciális feladatra alkalmasak a nyílt rendszerek. Első lépésként az lenne a legfontosabb, hogy minden rendszer a nyílt megoldásokat használja, vagy legalább adatcsereszinten képes legyen ezek kezelésére, mert az együttműködő rendszerek jelentik a közeljövő fejlődési útját.

A szövegben található *webhelyek* internetes címei:
Open Archives Initiatives:
<http://www.openarchives.org/>
Elektronikus dokumentumok archívumának kezelésére szolgáló ingyenes szoftver:
<http://www.eprints.org>

Az eprints.org szoftver két kísérleti archívuma:
Nottinghami egyetem:

<http://www-db.library.nottingham.ac.uk/eprints/>

Glasgow-i egyetem: <http://eprints.lib.gla.ac.uk/>

Neumann-ház: <http://www.neumann-haz.hu>

Magyar Elektronikus Könyvtár:

<http://www.mek.iif.hu>

Jász-Nagykun-Szolnok Megyei Verseghegy Ferenc Könyvtár, Elektronikus Könyvtár:

<http://www.vfmk.hu/vfek/>

Debreceni Egyetem Egyetemi és Nemzeti Könyvtár: <http://fulltext.lib.unideb.hu/>

Az egri Bródy Sándor Megyei Könyvtár képeslapgyűjteménye:

<http://www.brody.iif.hu/WEB/Web.htm>

Irodalom

- [1] BRADLEY, Neil: Az XML kézikönyv. Bicske, Szak Kiadó, 2000. 564 p. ISBN 963 9131 21 0
- [2] Forum for Metadata Schema Implementers. <http://www.schemas-forum.org>
- [3] Dublin Core Metadata Initiatives. <http://dublincore.org/>
- [4] PINFIELD, Stephen—GARDNER, Mike—MacCOLL, John: Setting up an institutional e-print archive. = Ariadne, 2002. 31. sz. április 11. <http://www.ariadne.ac.uk/issue31/eprint-archives/>

Beérkezett: 2002. VI. 24-én.

Könyvjegyzék Központ

2000 júliusában David Wilson Taylor, egy nyugdíjas könyvtáros, elkezdte építeni a *Booklist Center* nevű webhelyet, amelyet a „legjobb és díjat nyert könyvek” gyűjteményének szánt. Jelenleg ez a katalógus több mint 250 jegyzéket tartalmaz, amelyet 62 kategóriába szerveztek. A könyvek jegyzéke kézikönyvek, elektronikus könyvek, tudományos és műszaki művek stb. listáiból áll. A díjnyertes művek között Pulitzer-díjas, Nobel-díjas és más díjakat elnyert könyvek találhatók.

A Booklist Center <http://home.attbi.com/~dwtaylor/index.htm> alatt található.

/IATUL-hírlevél, 2002. augusztus, 1 p./

Virtuális égbolt

A *Virtual Sky Viewer* nevű portál bepillantást enged az éjszakai égbolt csillagaira, változásaira. Több mint népszerű-tudományos album, mivel a teljes északi félteke égboltját különböző részleteiben és változó nagyításokkal lehet szemlélni, zoomolás igénybevételével. A különféle jeleneteket az égbolton kattintással lehet előhívni, zoomolással nagyítani vagy kicsinyíteni, és így nagyobb területeket befogni. A csillagászati koordinátát is jelölik a képeken.

További információ: <http://virtualsky.org>

/Information World Review, 182. sz. 2002. július/augusztus, p. 17./

(R. P.)