

Szepesi Judit

Digitalizálás a Németh László Városi Könyvtárban*

„Óvakodnunk kell tehát az izolálódástól, és ügyelnünk kell arra, hogy sikeresen illesszük mozgalmunkat a létező, komplex folyamatokba. Ha nem tartunk lépést a korral, könnyen a múlt vitrinjében rekedhetünk. Fennmaradásunkhoz azonban múltbéli értékeinket is meg kell őriznünk.”

(Carlo Petrini)

Az MTA SZTAKI és a szegedi iKRON Kft. 2002-ben fejlesztette ki új integrált könyvtári rendszerét, a Huntékát, amelynek egyik része az e-könyvtár – vagy JaDoX – modul. A hódmezővásárhelyi Németh László Városi Könyvtár digitalizálási munkálataihoz megfelelő háttérrel nyújt. A Digitalizáló Műhely létrehozta saját, a weben hozzáférhető elektronikus gyűjteményét a JaDoX szoftver használatával. A cikk a hardver- és szoftverkörnyezet elemeinek sikeres használatáról számol be, hangsúlyozva a megfelelő szabványok alkalmazásának fontosságát.

A könyvek megőrzésére vállalkozó könyvtár mára már mennyiségileg és minőségileg is új szakaszba lépett: nemcsak a nyomtatott információ őrzéséről és megtalálhatóságáról kell gondoskodnia, hanem a legkülönbözőbb médiumokon megjelenő információforrásokat is gyűjtenie, rendszereznie és szolgáltatnia kell. Mindez új feladatokat, egyúttal új lehetőségeket is jelent a könyvtáraknak: megte-remtődtek a technikai feltételei annak, hogy a felhalmozott tudásanyag „kiszabaduljon” az intézmények falai közül, s a számítógépek segítségével a hálózatokon, az interneten keresztül akár a nap 24 órájában, a világ bármely részéről, akár saját otthonunkból is elérhetővé váljon.

A műhely kialakítása

Hardverkörnyezet

Az archivált adatok védelme és a fájlok nagy mérete miatt a könyvtár hálózatán belül létrehoztunk egy gigabites Ethernet-hálózatot, amelyet csak a műhely munkatársai érhetnek el. A műhelyben három munkaállomás, egy fájlserver és szkenne-rek kaptak helyet.

Az archiválandó dokumentumok kiválasztásának szempontjai

A dokumentumokat helytörténész kollégáinkkal választjuk ki, azokat archiváljuk, amelyek iránt

nagy az érdeklődés, de állományvédelmi okokból már nem adhatók kézbe. Minden esetben tájéko-zódunk, hogy a kiválasztott dokumentumról létezik-e digitális másolat. Mindezek mellett nagy hang-súlyt fektetünk a szerzői jogok betartására:

- Az 1999. évi LXXVI. szerzői jogi törvény alapján nem digitalizálunk és szolgáltatunk olyan anya-got, amihez nincs szerzői jogi hozzájárulásunk.
- A Digitalizáló Műhelyben készült képeket elekt-ronikus védjegyükkel látjuk el.
- A Németh László Városi Könyvtár saját szerve-rén helyezzük el és szolgáltatjuk a digitális do-kumentumokat.

A feldolgozó formátumok és szoftverek kiválasztásának szempontjai

Az archiválási munkák megkezdése előtt nagy kérdés volt számunkra, milyen formátumban dol-gozzuk fel a szövegeket úgy, hogy azok az egyes hardver- és szoftverkörnyezetekben egyaránt időtállóak és hordozhatók maradjanak. Ez azért fontos kérdés, mert a hordozhatóság, az adat-kommunikáció kulcsa mindig az alkalmazott fájl-formátumokban keresendő, és sok múlik az adat-formátumokon is.

* A 2006. április 19–21. között Miskolcon megrendezett *Networkshop* konferencián elhangzott előadás szer-kesztett változata.

HTML

Archivált dokumentumainkat először HTML (Hiper-text Markup Language) formátumban dolgoztuk fel, mivel a HTML azt írja le, hogy az adatok hogyan jelenjenek meg a képernyőn és nyomtatásban. A nyelv kifejezetten webes megjelenítésre készült. Munkánk során azonban be kellett látnunk, hogy a szövegek, a metajelek, vezérlőjelek stb. semmit sem mondanak a szöveg tartalmáról, szemantikájáról. Ezért új tárolási formátum után néztünk.

SGML/XML

Az információfeldolgozással kapcsolatban a következő követelményeket fogalmaztuk meg:

- Legyen az információ (a továbbiakban információn szöveges információt értünk) szerkezete tervezhető.
- Legyen olyan elfogadott szabvány, amellyel az ember és a gép számára le lehet írni az információ szerkezetét, meg lehet jelölni a felhasználáshoz a lényeges elemeket.
- Legyen az információ tárolására általánosan elfogadott, dokumentált, gyártófüggetlen szabvány.
- Az információ legyen számítógéppel könnyen feldolgozható.
- A kívánt információt gyorsan, pontosan meg lehessen találni.
- A tárolt információ legyen hordozható.
- A tárolt információ legyen sokoldalúan újra felhasználható.

Ezeknek a kritériumoknak megfelel az SGML (Standard Generalized Markup Language – ISO 8879) nemzetközi szabvány, illetve annak leszármazottja, az XML (eXtensible Markup Language). Az SGML és az XML közös tulajdonsága, hogy elkülönítik a tartalmat, a megjelenítést és az adatok struktúráját, míg a HTML nem.

Szoftverkörnyezet

Könyvtárunk 2003-ban a *TinLib* könyvtári integrált rendszerről áttért egy új fejlesztésű rendszerre, a *HunTékára*, amely az *MTA SZTAKI* és a szegedi székhelyű *iKRON Kft.* közös munkája alapján született meg. Nagy örömünkre a rendszer folyamatosan bővül, egyre több könyvtári munkafolyamatot automatizált, beleértve a digitalizált dokumentumok feldolgozását és webes közzétételét.

Az új modul *JaDoX*, vagy elektronikus könyvtár néven látta meg a napvilágot. Önálló rendszerként is használható, de természetesen integrált a

HunTéka többi moduljával is. (A dokumentumok címléírása történhet a katalógusmodulban, ahol a JaDoX-ban tárolt forrást statikus JaDoX URL-ként adhatjuk meg.)

A JaDoX általános jellemzői

A szoftver részei:

- független, nyílt forráskódú, ingyenes adatbázis-kezelő (PostgreSQL RDBMS);
- platformfüggetlen Java-alkalmazások.

Ezek az alkalmazások a következők: JaDoX-szerver, a JaDoX-szerkesztő, testre szabható webfelület (JSP, TomCat) és egy import alkalmazás (tömeges offline betöltéshez). A szoftver a digitalizálás fázisai közül a következőkben nyújt segítséget: korrektúra, import, feldolgozás, tárolás, közzététel (keresés, böngészés, kapcsolódás egyéb rendszerekhez).

Munka a JaDoX -ban

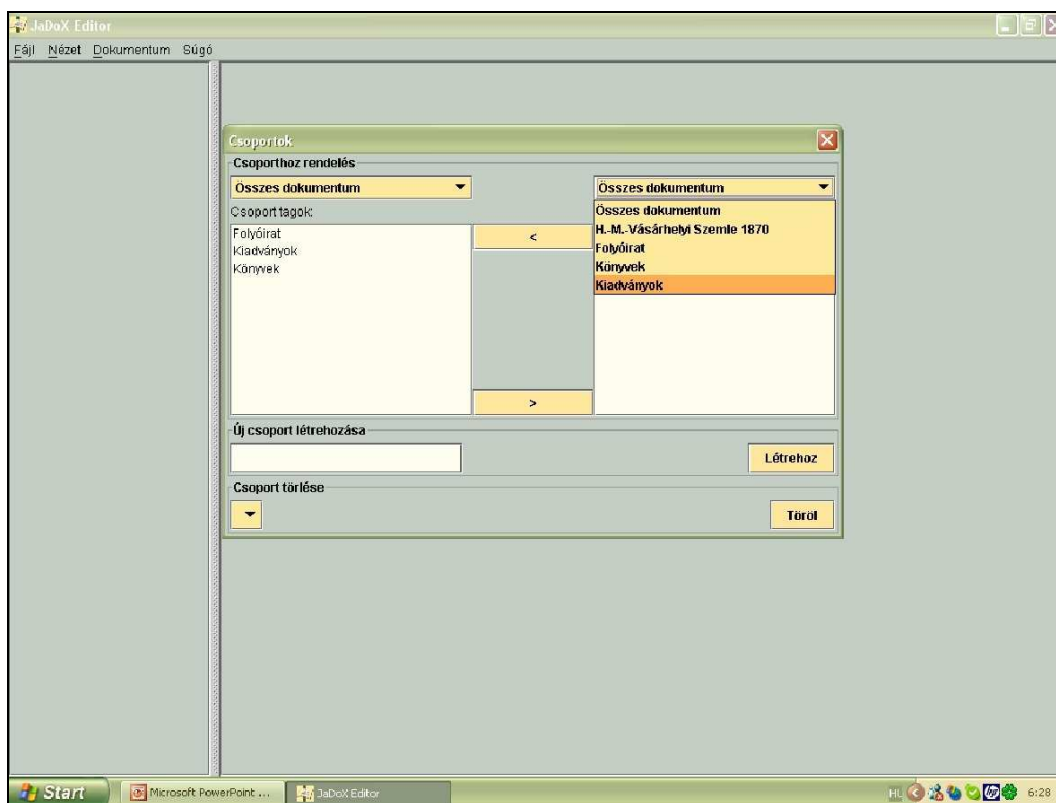
Import

A karakterfelismerő programmal (OCR) felismertett és TXT formátumban elmentett szövegeket importáljuk a JaDoX-ba. A digitalizált dokumentumok importja kétféle módon lehetséges: vagy a szerveroldali alkalmazás segítségével (nevezéktan alapján történő tömeges import), vagy pedig az import menüpont segítségével a szerkesztői felületen oldalanként. A betöltés során a szöveges dokumentumok nyers XML dokumentumként állnak rendelkezésünkre.

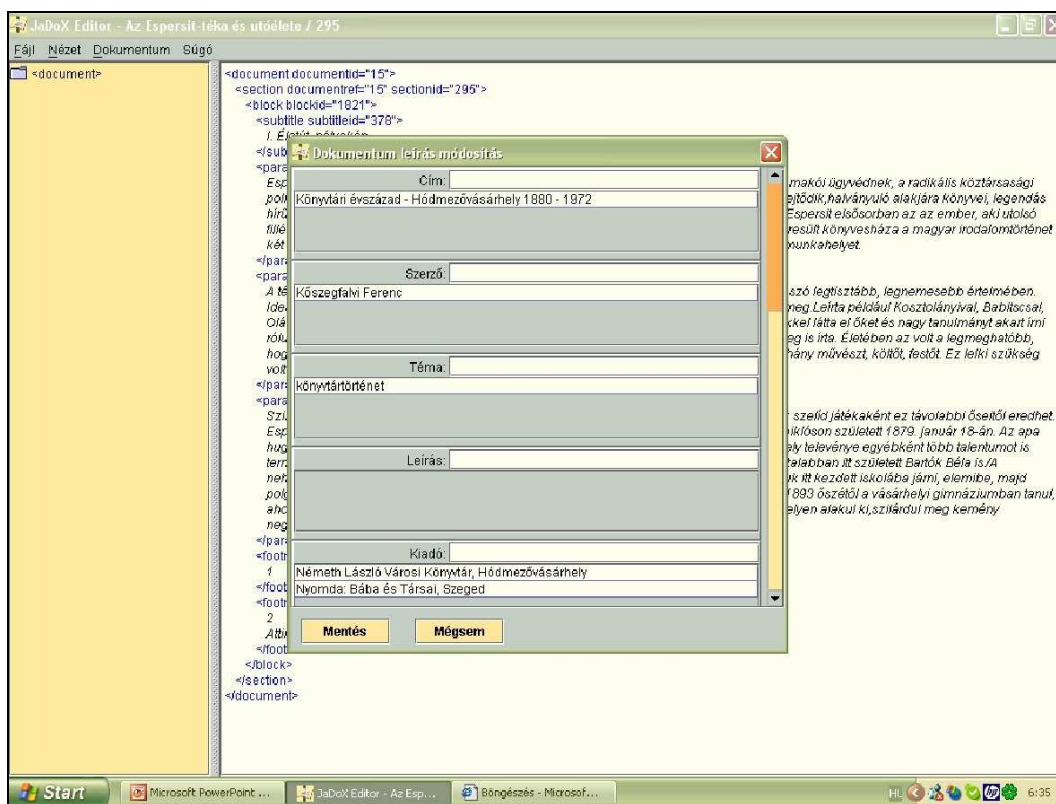
Csoportkialakítás, feldolgozási műveletek, formázási parancsok

Az importálás után csoportokat alakítunk ki, pl. könyvek, folyóiratok, kiadványok (1. ábra). A csoportokat egy ún. *Összes dokumentum* nevű csoport fogja össze. A csoporton belül dokumentumokat hozunk létre (pl. *Vásárhelyi szemle*), ezen belül pedig szekciók (oldalak) jönnek létre. Mindig szekciókkal dolgozunk. Minden dokumentumot feldolgozunk a Dublin Core tizenöt alapelemének segítségével (2. ábra).

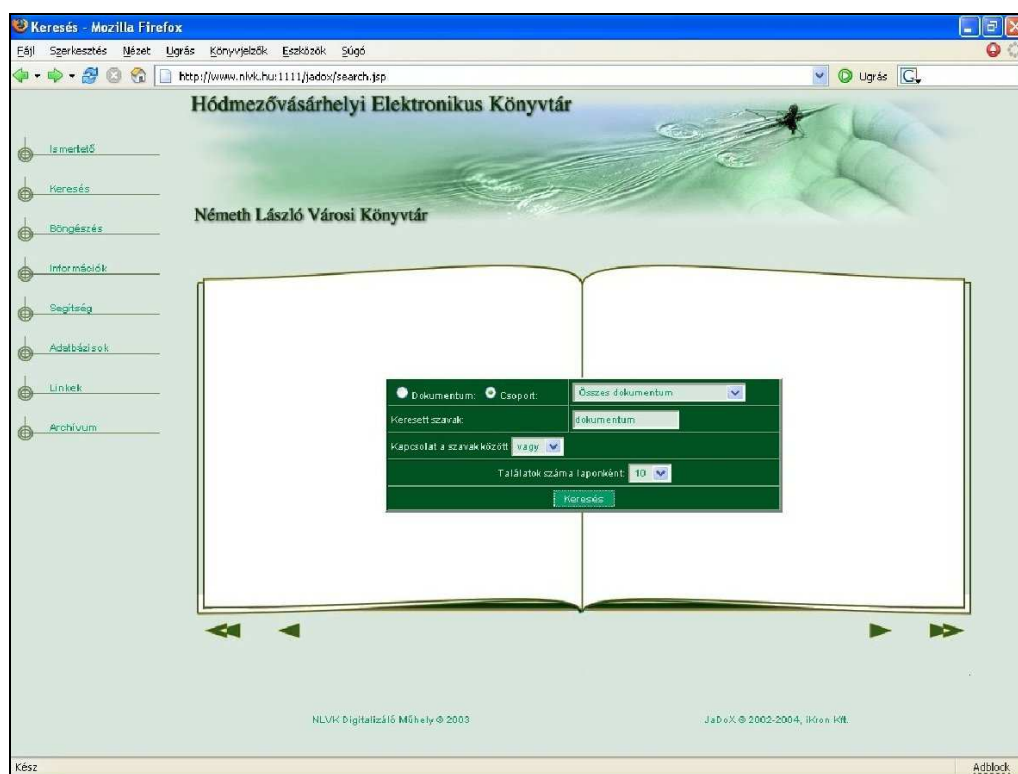
Ezután megnyitjuk a dokumentumhoz tartozó szekciókat. Ekkor a rendszerbe beépített formázási parancsok segítségével formázzuk, tageljük a szöveget, tetszőleges mélységben, akár többször is. A rendszerbe épített DTD szerint strukturálhatjuk a szöveges állományt, megfelelő definíció ese-



1. ábra Csoportkialakítás a JaDoX-ban



2. ábra Feldolgozás a Dublin Core segítségével



3. ábra Keresési lehetőség a webes felületen

tén a tartalomjegyzék automatikusan generálódik a tagek alapján. Az editorban korrektúrázási lehetőség is van.

Tárolás és szolgáltatás

A tárolás a JaDoX relációs adatbázis-kezelőben történik. Emellett – az alkalmazásfüggetlen újrahasznosíthatóság és a nagyobb sebesség miatt – a teljes XML dokumentumot is tároljuk. A weben való közzététel automatikus. A JaDoX webfelülete JSP oldalakból áll, a dokumentumok a következő transzformáció után jelennek meg a felületen:

RDBMS -> XML -> (XSL) -> HTML

A webfelület segítségével kereshetünk az egyes csoportokban vagy az összes dokumentumban csonkolással, vagy nélküle, böngészhetünk a metaadatok és a tartalomjegyzék megjelenítése mellett (3. ábra).

Kapcsolat más keresőkkel

Adatszolgáltatóként metaadatainkat az OAI-PMH protokoll segítségével rendszeresen átadjuk olyan központi adatbázisoknak, mint az NDA (<http://kereso.nda.hu/>) és a SZTAKI (<http://nda.sztaki.hu/kereso/>) keresője.

Tesztelés

A *Publika Magyar Könyvtári Kör* a *Békéscsabai Megyei Könyvtárral* és a *Németh László Városi Könyvtárral* együttműködésben az elektronikus könyvtári szolgáltatások használhatósági vizsgálatának megismertetésére és elterjesztésére dolgozott ki programot. Céljaul tűztük ki, hogy:

- a könyvtári szakemberek minél szélesebb körben ismerkedjenek meg az infokommunikációs rendszerek használatával;
- meg tudjuk határozni azokat a fő eszközöket, módszereket és tényezőket, amelyek alapvető feltételei a használhatóságnak;
- legyen ez a tesztelési fórum visszajelzés a weboldalak készítő szakemberek számára, hol és mit érdemes javítani a rendszeren a rugalmas, sokoldalú felhasználás érdekében.

A program 2006. május 17-én indult. A tizenhét, különböző korú (16 évestől a nyugdíjasig), és különböző felhasználói ismeretekkel rendelkező tesztelő két webes keresőt tesztelt:

- a békéscsabai Körös-Maros kulturális örökség oldalát (<http://koros-maros.bmk.hu/>),
- a hódmezővásárhelyi elektronikus könyvtár oldalát (<http://www.nlvk.hu/jadox/guide.jsp>).

A tesztelésre szánt idő egy-egy óra volt. Két tesztelési célt fogalmaztunk meg:

- mennyire vonzó a két tesztelt honlap, milyen világosan jelzik szerkezetüket, céljaikat,
- mennyire használható a webhely a bennük való információkeresésre.

Összességében elmondhatjuk, hogy elektronikus könyvtárunk a felhasználóktól pozitív visszajelzéseket kapott. A kért információkeresést el tudták végezni, az oldalt kellemes színhatásúnak, felhasználóbarátnak ítélték. Negatív visszajelzés volt, hogy a betűk túl kicsik, ezért a szöveg összeolvad. Ezen a későbbiekben javítanunk kell.

Összegzés

A digitalizálás módszertanának hatalmas szakirodalmi van. Ennek egyik oka, hogy még mindig a kísérletezés és útkeresés korszakában élünk, ahol most kristályosodnak ki a legjobb megoldások; a másik ok, hogy a technikai lehetőségek naponta bővülnek, ami egészen új perspektívákat jelent. Azt is látni kell azonban, hogy szellemi termékeink digitális jövőbe való átmentése pénzügyi kérdés is. A pénzügyi befektetések jelentős kockázattal járnak, ha a digitalizálás nem a megfelelő technikai

eszközök és szabványok alkalmazásával valósul meg.

Irodalom

BÍRÓ Szabolcs: A szövegfeldolgozás modern eszközei – az SGML és XML nyelvek. = Tudományos és Műszaki Tájékoztatás, 51. köt. 10. sz. 2004. p. 453–459.

A HunTéka elektronikus könyvtár modulja. <http://www.hunteka.hu/uploads/490/333/networkshop-2006-jadox-eloadas-hu.htm>

Publika Magyar Könyvtári Kör. <http://www.ki.oszk.hu/publika/>

SGML (Standardized General Markup Language). <http://www.sil.org/sgml/sgml.html>

XML (Extensible Markup Language). <http://www.w3.org/TR/2004/REC-xml-20040204/>

Beérkezett: 2006. VI. 15-én.



Szepesi Judit

a Németh László Városi Könyvtár
Digitalizáló Műhelyének vezetője.
E-mail: szjutka@nlvk.hu

Tegye egyszerűbbé munkáját az EBSCO-val!



Az elektronikus folyóiratok beszerzése összetett feladat, melyben szívesen támogatjuk Önt. A kiadóknál megérdeklődjük az aktuális ajánlatokat, tisztázzuk a folyóirat beszerzésének és megjelenésének módját, valamint segítjük Önt a licencszerződések kidolgozásában.

Engedje, hogy mi végezzük el ezt a feladatot Ön helyett!

Szolgáltatás. Megbízhatóság. Tájékoztatás. Megoldás.
Csak az EBSCO Information Services-től.



www.ebsco.com

CUSTOMERFOCUSEDCONTENTDRIVEN