

# A HunTéka elektronikus könyvtár modulja

A HunTéka integrált könyvtári rendszer 2002-ben debütált az MTA SZTAKI és az iKron Kft. több éves közös fejlesztésének eredményeként. Jelenleg több mint 44 intézményben működik, továbbfejlesztése pedig folyamatosan zajlik saját terveink és a HunTéka felhasználók visszajelzései alapján.

2005-ben a HunTéka elektronikus könyvtár modullal bővült, mely nagyban megkönnyíti a digitalizálási munkák elvégzését a feldolgozástól a szabványos tároláson keresztül az Interneten való közzétételig. Ez a modul, a JaDoX néven különálló szoftverként is működő, nyílt szabványokon alapuló, platform-független dokumentumkezelő rendszer. A JaDoX alapvető korrektúrázási, szöveg feldolgozási, tárolási és közzétételi funkcióit a 2005-ös Networkshop-on mutattuk be.

Jelen előadás első felében az elektronikus könyvtár modul rövid bemutatása mellett az időközben megvalósult továbbfejlesztéseket részletezzük, a második felében pedig gyakorlati tapasztalatokról esik szó a hódmezővásárhelyi Németh László Városi Könyvtár Digitalizáló Műhelye vezetőjének tolmácsolásában.

## I. rész. Az elektronikus könyvtár modul bemutatása

### Főbb jellemzők

#### ***Mi is az a JaDoX***

A JaDoX egy önmagában is jól működő, önálló elektronikus könyvtári program, amely egyben a HunTéka elektronikus könyvtár moduljának szerepét is betölti. A JaDoX hatékony és teljes körű segítséget nyújt digitalizált dokumentumok feldolgozásához, tárolásához, megjelenítéséhez és visszakereshetőségéhez. E funkciók segítségével teljes mértékben kiszolgálja az elektronikus könyvtárral szemben támasztott igényeket.

#### ***Komponensek, technológia***

A szoftver egy független, nyílt forráskódú, ingyenes adatbázis kezelőből (PostgreSQL RDBMS) és platform-független JAVA alkalmazásokból áll. Ezek az alkalmazások a JaDoX szerver, a JaDoX szerkesztő, a Testre szabható WEB felület (JSP, TomCat) és egy Import alkalmazás tömeges offline betöltéshez.

Az alábbiakban ✓-val jelöltük, hogy ez a szoftver a digitalizálás fázisai közül melyekben nyújt segítséget:

- Szkennelés
- Felismertetés (Recognita, Fine Reader)
- Korrektúra ✓
- Import ✓
- Feldolgozás ✓
- Tárolás ✓
- Közzététel (keresés, böngészés, kapcsolódás egyéb rendszerekhez) ✓

A korrektúra akár egy tetszőleges szövegszerkesztőben akár a JaDoX editorban elvégezhető. A digitalizált dokumentumok importja kétféle módon lehetséges, vagy a szerver oldali alkalmazás segítségével (nevezéktan alapján történő tömeges import), vagy pedig az import menüpont segítségével a szerkesztői felületen.

A feldolgozás során megadhatjuk a dokumentum metaadatait (Dublin Core), tetszőleges csoportokba szervezhetjük őket (fa struktúra), tetszőleges DTD szerint strukturálhatjuk a szöveges állományt (tag-elés tetszőleges mélységben, akár több körben is). Megfelelő DTD esetén a tartalomjegyzék automatikusan is generálódhat a megfelelő tag-ek alapján.

A tárolás a többi komponens szempontjából a JaDoX szerverben történik, a szerver pedig relációs adatbázis-kezelőben helyezi el a dokumentumokat. Erre a célra DTD-független relációs adatmodellt használunk. Emellett a teljes XML-t is tároljuk (az alkalmazás-független újrahasznosíthatóság és a nagyobb sebesség miatt).

A WEB-en való közzététel automatikus. A JaDoX web felülete JSP oldalakkól áll, a dokumentumok az alábbi transzformáció után jelennek meg a felületen:

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

A WEB felület segítségével kereshetünk (csonkolással, vagy anélkül), böngészhetünk a metaadatok és a tartalomjegyzék megjelenítése mellett, illetve eltérő megjelenítést definiálhatunk a különböző dokumentumtípusok esetén (pl.: folyóiratok cikk-alapú és könyvek oldal-alapú megjelenítése. Szöveges, képi és PDF letölthetőség (bővíthető).

## **JaDoX a HunTékában**

A cíMLEÍRÁS (metaadatok felvétele) a HunTéka katalógus moduljában is történhet. Innen egy kattintással felhívható a kapcsolódó JaDoX dokumentum betöltésre, feldolgozásra, megjelenítésre a HunTéka rendszeren belül. A JaDoXból is átléphetünk a HunTékába a dokumentum cíMLEÍRÁSA céljából.

## **Kapcsolat egyéb rendszerekkel**

A cíMLEÍRÁS történhet a könyvtár saját katalógusában, ahol a JaDoX-ban tárolt forrást statikus JaDoX URL-ként adjuk meg.

Adatszolgáltatóként (Data Provider) metaadatainkat az OAI-PMH protokoll segítségével rendszeresen átadhatjuk olyan központi adatbázisok számára mint pl. az NDA.

A szintén önálló programként is funkcionáló, a HunTéka OPAC moduljának szerepét betöltő Monguzon keresztül elektronikus könyvtárunkat egyéb adattárakhoz is kapcsolhatjuk.

## **Az elmúlt év fontosabb továbbfejlesztései**

Megvalósult a dokumentumok hosszú távú azonosítására szolgáló URN-NBN kód támogatása.

A JaDoX korábbi, a szabványos szövegfeldolgozást célzó alap-funkcionalitását kiterjesztettük egyéb szöveges és bináris dokumentumtípusok kezelésére.

Korábban a JaDoX telepítésekor meg kellett adnunk azt a DTD-t, amelyet elektronikus könyvtárunkban egységesen használni kívántunk. Most már szabadon definiálható DTD-eket alkalmazunk a hierarchikus dokumentumtár minden szintjén (dokumentumok, csoportok, gyűjtemények stb.).

Korábban a digitalizált dokumentumok szöveges verziójának szofisztikált feldolgozására helyeztük a hangsúlyt. Ezt a funkcionalitást megtartva, a felhasználók és az érdeklődők visszajelzései alapján az újabb verzióban a tömeges, tisztán kép-alapú archiválást is megkönnyítettük.

## **II. rész. Tapasztalatok a Németh László Városi Könyvtár Digitalizáló Műhelyében**

### **A Digitalizáló Műhely**

Az elmúlt 15 év folyamán a hódmezővásárhelyi Németh László Városi könyvtár hagyományos könyv kölcsönző funkciója átalakult szolgáltató könyvtári funkcióvá. A megváltozott igényeknek megfelelően szolgáltatásainkat folyamatosan fejlesztjük, átalakítjuk és bővítjük, új munkaterületek alakultak, alakulnak ki.

2004-ben, a könyvtárunkban Digitalizáló Műhelyt alakítottunk ki, ahol városunk, könyvtárunk helytörténeti és értékes régi dokumentumai kerülnek archiválásra és szolgáltatásra a Hódmezővásárhelyi Elektronikus Könyvtár felületén.  
<<http://www.nlvk.hu/jadox>>

### **A Műhely kiépítése**

Az adatok védelme miatt a könyvtár hálózatán belül létrehoztunk egy hálózatot, melyet csak a Műhely munkatársai érhetnek el.

Úgy gondoltuk, hogy a szkennelés következtében nagyméretű fájlok mozgatására lehet szükség az egyes munkaállomások és a szerver között, ezért gigabites ethernet hálózatot építettünk ki rendszergazdánkkal az Ikron Kft-vel.

A műhelyben 3 munkaállomást alakítottunk ki.

### **Munka a műhelyben**

#### ***A dokumentumok kiválasztásának szempontjai***

- A felhasználók körében keresett a dokumentum
- A dokumentum állapota nem kielégítő (állományvédelmi szempontok)
- Az országban még nem digitalizálták

A Hódmezővásárhelyi Elektronikus Könyvtár kialakítása a hazai tapasztalatokra támaszkodva történt.

Mintául a Magyar Elektronikus Könyvtár létrehozása és működtetése szolgált, majd ezek egyeztetésre kerültek az intézmény lehetőségeivel és elképzeléseivel.

## ***A digitalizálás és a szolgáltatás menete***

A műhely első munkája a Vásárhelyi Szemle 1870-es évfolyama volt. Összesen 7 füzete jelent meg.

A munka menete:

### 1. Archiválás

A dokumentumok oldalait nagy felbontással (400 dpi) beszkeneltük a Fine Reader-be, itt az oldalakat jpg-be és txt-es formátumba UTF 8-as kódolással mentettük.

### 2. Korrektúrázás

Következett a korrektúrázás, vagyis az eredeti szöveget összehasonlítottuk a szöveges (txt) dokumentummal.

### 3. Szolgáltatás

A txt-es – bizonyos esetekben - a jpg formátumú dokumentumokat beimportáltuk dokumentumkezelő rendszerünkbe, a JaDox-ba. Az import végén a szöveges dokumentumok nyers xml dokumentumként álltak a rendelkezésünkre.

## ***Munka a JaDoX-ban***

Az importálás után *csoportokat* alakítottunk ki (pl. könyvek, folyóiratok, kiadványok stb.) Ezeket, a csoportokat egy ún. összes dokumentum nevű csoport fogja össze. Ennek már nincs őse, fölérendeltje. A csoporton belül *dokumentumokat* hoztunk létre pl. Vásárhelyi szemle, ezen belül pedig *szekciók (oldalak)* jönnek létre. Mindig szekciókkal dolgozunk.

A könyvek feldolgozása oldal a folyóiratok pedig cikk alapúak.

A feldolgozási munkák megkezdése előtt, érdemes a dokumentumokat Dublin Core adatelemekkel ellátni. A rendszerben a 15 adatelem megtalálható.

Ezután megnyitottuk a dokumentumhoz tartozó szekciókat. Ekkor a rendszerbe beépített formázási parancsok segítségével formázhatjuk, tag-geljük a szöveget.

Itt az editorban a szöveget korrektúrázhatjuk is.

A formázási parancsok megadása után érdemes a munkánkat leellenőrizni az interneten.

## ***Közzététel a weben***

Az importálás után automatikusan láthatjuk a dokumentumokat a weben.

### Keresés :

-- csonkolással, vagy anélkül

### Böngészés :

- Metaadatok, tartalomjegyzék megjelenítése
- Eltérő megjelenítés folyóiratok (cikk-alapú) és könyvek (oldal-alapú) esetén
- Szöveges, képi és PDF letölthetőség

## **Összegzés**

A digitalizálás költséges művelet, komoly szellemi és pénzügyi befektetéseket igényel. Ezek a befektetések jelentős kockázattal járnak, ha nem megfelelő technikai feltételekkel és a szabványok alkalmazása nélkül valósulnak meg. Ezért fontos, hogy olyan szabványos, nyílt rendszerek alkalmazására törekedjünk, amelyek biztosítják a tárolt dokumentumok időtállóságát és az információk cseréjét.