

Cikkarchívum fejlesztése

Burmeister Erzsébet
Miskolci Egyetem, Könyvtár, Levéltár, Múzeum, erzsi@uni-miskolc.hu

Cséfai János
Monguz Kft., jcsefai@monguz.hu

Előzmények

A Miskolci Egyetem Könyvtára a TIOP 1.2.3 Tudásdepó Express pályázat többek között keretében szervereket és szoftvereket vásárolt. A Monguz Kft. Nyerte meg a közbeszerzési pályázatot és a könyvtár megvette a HunTéka integrált rendszert, a JaDoX digitális dokumentumokat kezelő, a Portál24 közgyűjteményi portálok kialakítására szolgáló szoftvert és a Monguz „információ bróker”, amelyet távoli információforrások közös felületen való lekérdezésére fejlesztettek ki.

A könyvtár számára a JaDoX szoftver alapján repozitóriumot dolgoz ki a Monguz fejlesztő gárdája a jelenlegi MIDRA (Miskolci Egyetemi Digitális Raktár és Adattár) felváltására. Ennek az előadásnak azonban nem az egyetemi digitális dokumentumok tárolására és szolgáltatására való repozitórium a témája, hanem a szintén JaDoX alapokon nyugvó cikkarchívum fejlesztése. Ehhez a támogatást a TAMOP 3.2.4 Tudásdepó Express pályázat adja.

A cikkarchívum célja

Egész általánosan fogalmazva: digitális, főleg szövegeket tartalmazó dokumentumok (cikkek, részdokumentumok) tára, melyet magyar könyvtárak konzorciuma tartja majd fenn, építi, használja és szolgáltatja. A cikkarchívum létrehozása része egy átfogó, országos méretű könyvtári fejlesztési stratégiának, és mint ilyen, a kapcsolódó fejlesztési tervekkel szoros együttműködésben valósul meg.

A kész cikkarchívumot fogja használni az Országos Széchényi Könyvtár Tájékoztatási és Olvasószolgálati Igazgatóság Reprográfiai Szolgálat az elkészült cikkmásolat rendelések tárolására. Jelenleg is jelentős mennyiségben zajlik a részdokumentumok digitalizálása megrendelések alapján, és a tevékenység során összegyűjtött digitális tartalmak betöltendők a közös archívumba. Várhatóan a jövőben is a nemzeti könyvtár lesz a központi másolatszolgáltató hely, ezért a Cikkarchívumot elérhetővé kell tenni az OSZK-ból, betöltésre, használatra és letöltésre egyaránt.

A Debreceni Egyetem Egyetemi és Nemzeti Könyvtára működteti az ODR (Országos Dokumentumellátási Rendszer) szolgáltatást és szintén a TAMOP 3.2.4 Tudásdepó Express pályázat teszi lehetővé az ODR megújítását. A Cikkarchívum úgy vesz majd részt az ODR szolgáltatásban, hogy tartalmával hozzájárul a másolatkérések gyors és hatékony teljesítéséhez. A Cikkarchívum tartalma bővül is majd az ODR könyvtárközi kölcsönzési tevékenységének másolatkérési funkciója által, hiszen az ott elkészült, az archívumban még nem található részdokumentumokról készült másolat is gazdagítja a cikkarchívum tartalmát. Ezért gondoskodni kell a Cikkarchívum és az ODR szolgáltatás együttműködéséről, a gördülékeny adatcseréről, az ODR felületről a Cikkarchívum eléréséről.

A cikkarchívum szoros kapcsolatot tart majd fenn nagy magyar folyóiratok, periodika cikk adatbázisokkal: MATARKA, EPA, HUMANUS. Ezek leartott metaadatai a cikkarchívumon keresztül is kereshetők lesznek.

Digitális objektumok kezelésére vonatkozó követelmények

Általános követelmények

- export, import, de saját beviteli felület is
- adatbeviteli felületen helyzetfüggő súgó
- csoportos rekordátvétel
- többszintű keresés, böngészés
- hozzáférési szintek szerzői jogok betartása alapján
- normál felhasználó védett tartalmat nem tölthet le
- könyvtárak könyvtárközi kölcsönzésre letölthetik

Digitális objektumok kezelése

- *szöveges és képes dokumentumok fogadása* (lehetőleg archív formátumban, nagy felbontásban)
- *szövegek, OCR-ezett szövegek kezelése* DOC vagy TXT formában, vagy PDF-ben (egy vagy kétrétegű).
- *képek tárolása* TIFF formátumban, szolgáltatásuk TIFF vagy JPG formátumban
- *bélyegképek, szolgáltatási verziók létrehozása* automatikus konverzióval a mester formátumból
- szükség esetén tetszőleges *vízjel-kép beállíthatósága*, a vízjelezést a rendszer automatikusan végzi

Adatkezelői jogosultsági rendszer

- *Adminisztrátor*: adatbázis karbantartása, mentése, visszatöltése, mezőnevek módosítása, a többnyelvű felület kialakítása és módosítása, felhasználói jogosultságok beállítása
- *Vezető könyvtáros*: adatok egységesítése, törlés, a rekordok jóváhagyása, a feldolgozás elvégzése, az export és import funkciók működtetése
- *Katalogizáló könyvtáros*: adatbevitel, adatmódosítás - rekordok törlése és jóváhagyása nem megengedett

Felhasználói jogosultsági rendszer

- *Felhasználók*: alapvetően keresés, feltöltés, letöltés használhatják az adatbázist, adatok módosítása nem megengedett
- *Partner könyvtárak*: digitális dokumentumok és leíró adatok, fel- illetve letölthetése, többszintű keresés az adatbázisban
- *Látogatók/olvasók*: a keresési és böngészési funkciók használata, csak a szabadon szolgáltatható tartalmak megjelenítése

Metaadatok és digitális objektumok fogadása

- *Metaadatok*: MARC, XML, RIS
- *Digitális objektumok*:
 - ⇒ szöveg formátumban: SGML, XML
 - ⇒ kép formátumban: PDF és egyéb képformátumok

Gyűjtemények és adatfeldolgozás

- több gyűjtemény kezelése
- új gyűjtemény konfigurálhatósága adatfelvivő által
- egyes gyűjteményeknek külön keresőfelület, saját URL-el, illeszkedve a honlap designjához
- külön (telepítendő kliens nélküli) webes beviteli felület az egyes gyűjteményekhez
- gyűjteményenként eltérő adatszerkezet és beviteli felület lehetősége; kötelező és opcionális mezők használata
- több szintű beviteli űrlapok
- külső adatszolgáltató számára egyszerűsített űrlap; a felvett adat még nem publikus
- automatikus értesítés a felvitelről
- az egyszerűsített felvitel továbbszerkeszthetősége
- a felvitel munkafázisának követése (külső adatszolgáltató csak addig módosíthat adatot, amíg a könyvtár nem kezdi meg a feldolgozást)
- kötegetelt adatmódosítási lehetőségek (pl. URL változtatás, státuszváltoztatás)

Lekérdező felület

Három fajta: egymező, egyszerű és haladó

- *Egymező keresés* (Pl. google): a beírt szavakat és kifejezéseket a szoftver a szerzők nevét, címeit, tárgyszavakat együttesen tartalmazó kombinált indexben keresi le.
- *Egyszerű keresés*: felkínálja fel a szerzők, címek és tárgyszavak betűrendes, kulcsszavas indexét és ezek tetszőleges kombinációját, lehet idő és nyelv szerint szűkíteni.
- *Haladó keresés*: választható mezőkből engedni akár több szerző és cím kulcsszó Boole algebra szabályai szerinti kombinálását. Lehet idő és nyelv szerint szűkíteni.

Megjelenítés

- A találati halmaz tovább kereshető.
- A találati lista kiválasztott szempontok szerint (általában idő, szerzői betűrend, cím betűrend) sorrendbe rendezhető.
- A találati lista többféle formában megjeleníthető:
 - ⇒ rövidített (szerző, cím, tartalmazó dokumentum legfontosabb adatai, oldalszám)
 - ⇒ bővített (rövidített + tartalmazó dokumentum közreadója, ISBN vagy ISSN szám, esetleges tárgyszavak, terjedelmi adatok, stb.)
 - ⇒ MARC formátumban
 - ⇒ XML formátumban
- Találatok csoportos kijelölése, exportálása (plain text, MARC, OAI, XML, Dublin Core)

Statisztikák

Generálható, részletes statisztikák napi, havi és éves szintű kimutatásokkal az archívum gyarapodására és látogatottságára, valamint a szolgáltatás honlapjának látogatottságára vonatkozóan az alábbi részletezettséggel:

- *gyarapodás*: az archívumba bekerülő cikkek száma napi, havi és éves szinten,
- *gyarapodás*: az archívumba bekerülő cikkek száma feltöltő könyvtárak szerinti bontásban napi, havi és éves szinten,
- *gyarapodás*: a többi folyóirat adatbázisból (EPA, MATARKA, HUMANUS) automatikus adatcserével bekerülő cikkek száma napi, havi és éves szinten,

- *látogatottság*: a Cikkarchívum honlapjának és egyes menüpontjainak látogatottsága napi, havi és éves szinten,
- *látogatottság*: a lekérdezések figyelemmel kísérése a beírt szavak megjelenítésével együtt napi, havi és éves szinten

A fejlesztői munka

A fejlesztői munka meghatározó része a Monguz Kft. JaDoX és Monguz szoftvereinek továbbfejlesztése. A Cikkarchívum is ezen két program alapjaira épül.

- Az adatok sémájának elkészítése a 3 adatforrás alapján: Egy dokumentum hierarchia rendszer kidolgozása, amely alkalmas arra, hogy a 3 forrás adatait letárolja és kereshetően visszaadja.
- MATARKA OAI-DC elkészítése a Monguz szoftver segítségével.
- A MATARKA, EPA és HUMANUS adatainak (XML vagy bináris MARC) letöltése OAI-n keresztül egy közös tároló helyre.
- A bináris MARC-ot maga a JaDoX program dolgozza át XML formátumra. Az XML-eket XSL segítségével konvertáljuk át olyan formátumra amely a JaDoX számára már megfelelő, feldolgozható.
- A feldolgozott adatokban szerepelnek bizonyos azonosítók, URL-ek, amelyek felhasználásával tovább tudjuk bővíteni a meglévő információ mennyiséget. Például a periodikák leírásában szerepelnek az OSZK-s AMICUS-os azonosítók, amely alapján egy általunk megírt Z-kliens bizonyos időközönként letölti és beimportálja a JaDoX rendszerbe a megfelelő rekordokat. Egy másik, szerver alkalmazás segítségével, pedig a képeket és pdf-eket töltjük le, de csak olyan helyekről, ahonnan jogunk van. A PDF-ek külön feldolgozásra kerülnek, amely során indexeljük tartalmukat, kereshetőség céljából.
- A projekt legnehezebb része a duplum rekordok kiszűrése. A konvertált adatok egy duplum ellenőrző folyamaton mennek át, ha már a megfelelő adatok szerepelnek az adatbázisban, vagy esetleg változás történt a forrásban, akkor egy meghatározott hierarchia sorrendet követve módosul az adatbázis, ha nem akkor új rekordként bekerülnek a közös rendszerbe.

Web

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.

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.

Ennek a továbbfejlesztett verziója látja el a cikkarchívum feladatát.

A régi funkciók (keresés, böngészés, űrlapok szerkesztése.) kibővítésre kerültek, úgy hogy alkalmasak lettek a sémák kezelésére, valamint a jogosultsági szintek mélyebb definiálhatósága valósult meg.