

Armarius: egy felhasználóbarát digitális könyvtár

Nagy László Tibor¹, Stan Johann², Csiszár Emese³

Témavezető: dr. Egyed-Zsigmond Előd⁴

Kivonat

Napjainkban egyre nagyobb jelentőséget tulajdonítanak a régi, múzeumi értékű kéziratok, könyvek és levelek megőrzésének. A múzeumok és levéltárak ezeket az iratokat rendszerint elkülönített termekben védik, hiszen a fény, a levegő páratartalma, a por, a kíváncsiskodó kezek mind károsíthatják őket. Ezen értékes dokumentumok megőrzéséhez nyújt megoldást a digitalizálás. Azonban a folyamat során készült elektronikus dokumentumok közzététele sem kis feladat. Ebben próbál segítséget nyújtani az általunk kifejlesztett rendszer, mely egy könnyen kezelhető, intuitív kezelőfelületet biztosít a digitalizált dokumentumok böngészéséhez és leírásához. Jelen dolgozatunkban bemutatjuk a rendszert, kitérve egy konkrét alkalmazási területére: Bolyai Farkas és János által hátrahagyott kéziratok digitalizálására vonatkozó kezdeményezésre.

Marosvásárhely nevezetes könyvtára, a Teleki Téka felbecsülhetetlen értékű könyveknek, írott dokumentumoknak ad otthont. Ezen értékes dokumentumok, kulturális örökségünk részei, ezért fontos hogy minél szélesebb közönségnek legyen lehetősége hozzáférni, megtekinteni őket. A Téka egyik legértékesebb gyűjteménye a két híres matematikus Bolyai Farkas és Bolyai János kéziratai, mely mintegy 20.000 oldalt foglal magába. A majdnem két évszázados kéziratok megőrzése fontos és nehéz feladat, az idő múlásával és a lapok forgatásával egyre romlik az írás minősége. Erre a problémára nyújt megoldást egy élő, digitális könyvtár létrehozása.

Élő könyvtárként lehetővé kell tegye nemcsak a dokumentumok internetes megjelenítését, hanem egy egyszerű, könnyen kezelhető felhasználói felület segítségével biztosítani kell a dokumentumok manuális leírását, keresést, kommentálását, tudományos viták és megbeszélések lefolyását. Ezáltal nemcsak a kutatók munkája válna könnyebbé, hanem a Teleki Téka felbecsülhetetlen értékű gyűjteményének egy része elérhetővé válna a nagyközönség számára és ezáltal nőne a Téka és a két Bolyai munkásságának ismeretsége. Ezt a szerepet hivatott betölteni a jelenlegi prototípusrendszerünk.

¹ Babeş-Bolyai Tudományegyetem, Informatika Kar, informatika szak, 4. év

² Institut National de Sciences Appliquées (INSA), Lyon, Franciaország, informatika szak, 4. év

³ Sapientia Erdélyi Magyar Tudományegyetem, Marosvásárhely, Számítástechnika szak, 4. év

⁴ Egyetemi docens, Institut National de Sciences Appliquées (INSA), Lyon, Franciaország, Informatika kar

Online könyvtárak a világhálón

Világszerte egyre több múzeum dönt a digitalizálás és a digitalizált anyagok közzététele mellett. A világhálón három különböző kategóriát különböztethetünk meg: múzeumok honlapjai, online arhívumok, online digitális könyvtárak. Az első kategóriába tartozó honlapok célja a figyelemfelkeltés és a népszerűsítés, az online arhívumok jelentős mennyiségű digitalizált anyagot bocsátanak a felhasználó rendelkezésére, míg az online digitális könyvtárak különféle webes szolgáltatásokat biztosítanak.

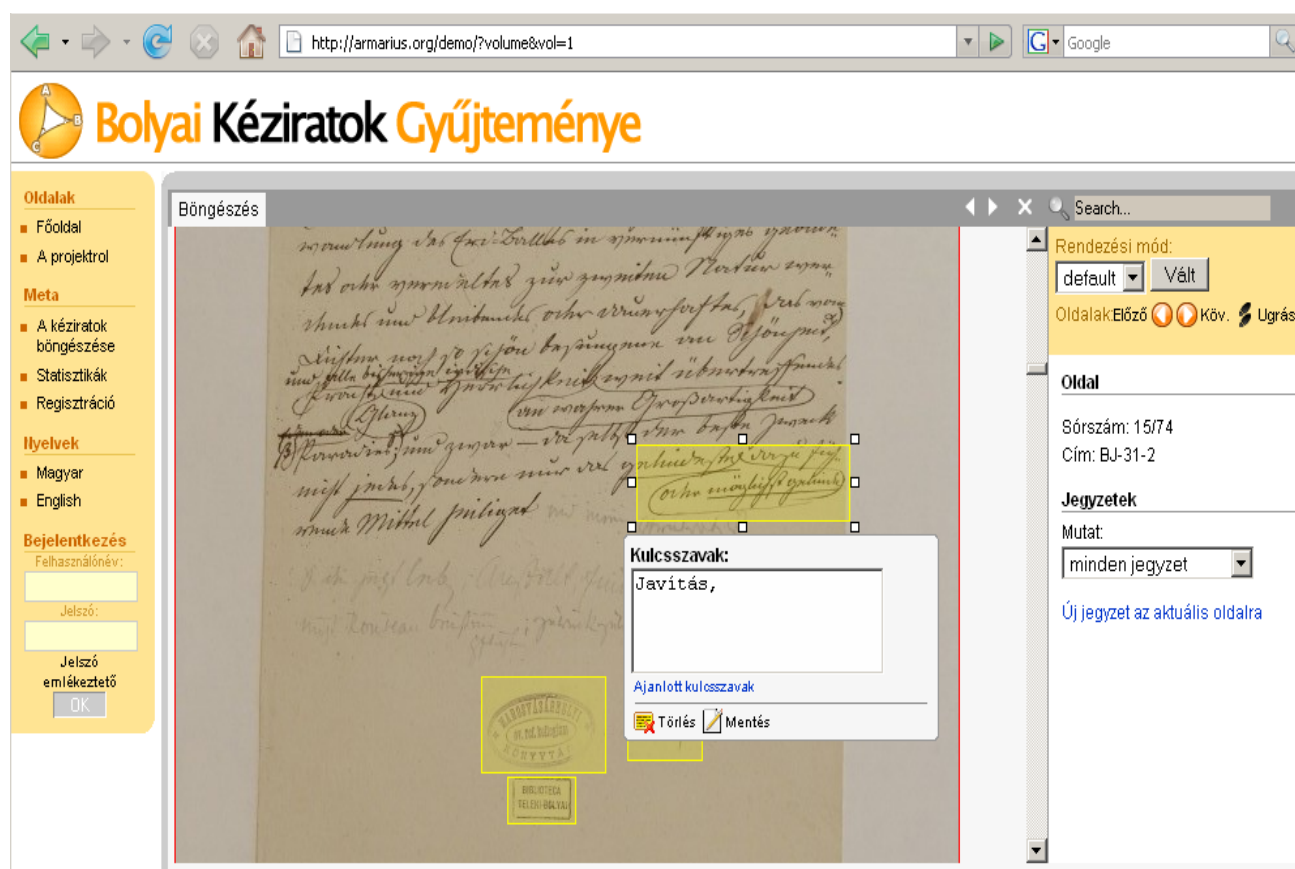
Napjainkban a széleskörű közönség egyre több olyan értékes dokumentumhoz, film- és hanganyaghoz jut hozzá, amiket eddig csak képzett szakemberek tekinthettek meg, ráadásul ezen dokumentumok tanulmányozása a weben keresztül könnyebb és nem helyhez kötött. Ilyen szolgáltatásokat számos múzeum biztosít, néhány közülük: Bibliotheca Corviniana Digitalis (<http://www.corvina.oszk.hu>)- Mátyás király kódexgyűjteménye, William Blake Archive (<http://www.blakearchive.org/blake>) - egy erős keresőfunkcióval rendelkező kéziratosok és rajzok tartalmazó gyűjtemény, a MuseoSuomi projekt – fejlett kulcsszóalapú dokumentum-leírást és böngészést biztosító felület.

Prototípusrendszerünk, Armarius⁵

Prototípusrendszerünk (Ábra 1) teljes egészében open source alapú, Apache webszervert használtunk és MySQL adatbázis-kezelővel dolgoztunk. A képeket egy állományrendszerben tároljuk, a kulcsszavakat pedig adatbázisban.

Rendszerünk egy olyan könnyen átlátható és kezelhető felhasználói felületet biztosít, mellyel a felhasználó böngészheti a kéziratosok, jegyzeteket fűzhet az egyes oldalakhoz vagy annak részeihez, módosíthatja vagy törölheti ezeket a jegyzeteket illetve kulcsszó alapú keresést végezhet a gyűjteményben. Ugyanakkor olyan “élő” szolgáltatásokat is biztosít, mint a kulcsszó-ajánlás gyakoriság és hasonlósági szintek alapján.

⁵ www.armarius.org



Ábra 1

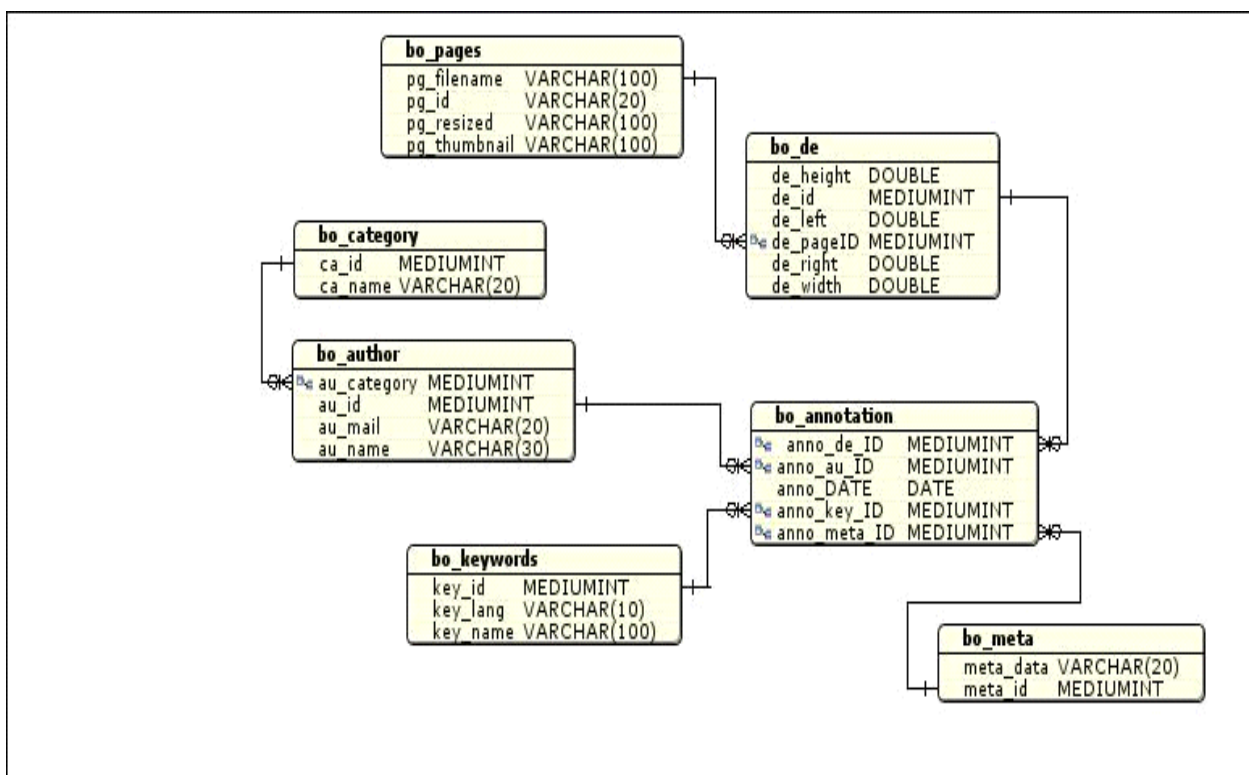
Elméleti megközelítés

Jelenlegi kísérleti rendszerünk céljai között szerepel a digitalizált dokumentumokat megjeleníteni, azok kézi leírását lehetővé tenni és egy gazdag keresőrendszerrel ellátni. A digitalizálás és képfelismerés utáni lépés a publikálási folyamatban a dokumentumokat, a szótárat és a leírásokat tároló adatszerkezet elkészítése. Ez az adatszerkezet figyelembe kell vegye a közzendő adatok sokféleségét: kép, szöveg, ábrák, jelek, numerikus képleírások, vektorok. Lehetővé kell tegye a korabeli kézírás és mai illetve „laikus” magyar nyelvre való fordításának tárolását. Célunk egy olyan web alapú „élő arhívum” kifejlesztése, mely az alapfunkciók mellett, egy interaktív felületet biztosít a felhasználó-felhasználó illetve a felhasználó-rendszer kommunikációra, és mely különféle felhasználó-segítési és -követési módszereket implementál. Mindemellett célszerű ha a rendszer funkciói többnyelvűek.

A digitalizált dokumentumokat egy gazdag keresőrendszerrel szeretnénk ellátni. A digitalizálás és képfelismerés utáni lépés a publikálási folyamatban a dokumentumokat, a szótárat és a

leírásokat tároló adatszerkezet elkészítése. Ez az adatszerkezet figyelembe kell vegye a közlendő adatok sokféleségét: kép, szöveg, ábrák, jelek, numerikus képleírások, vektorok. Lehetővé kell tegye a korabeli kézírás és mai illetve „laikus” magyar nyelvre való fordításának tárolását. Mivel többféle meta-adattal dolgozunk, szétválasztottuk a meta-adattípusokat (pl. cím, szerző, dátum, stb.) a kulcsszavaktól (a konkrét szavak, értékek, mint pl. Bolyai János, 1843, primszám stb.) és külön táblában tároljuk ezeket. Ezáltal a rendszer rugalmasan tudja kezelni a meta-adatokat, bármikor könnyen új típust lehet felvenni, a táblák struktúrájának vagy a már meglévő bejegyzések módosítása nélkül.

A Bolyai kéziratok struktúrája egyáltalán nem homogén. Bolyai nagyszerű matematikus volt, de sajnos nem gondolt arra, hogy munkáját az utókor számára könnyen hozzáférhetővé tegye. Ebből a tényből kiindulva tesszük azt a javaslatot, hogy egy kézirat esetén egy tetszőleges méretű téglalap jelentsen egy dokumentumegységet. A felhasználó jegyzetelés során dokumentumegységeket hoz létre, a jegyzeteit pedig dokumentumegységekhez köti. Minden oldalhoz több dokumentumegység fűzhető, a lapon elfoglalt pozíciójukat tároljuk az adatbázisban. Ugyanakkor minden dokumentumegységhez több meta-adat típus rendelhető és több kulcsszóval írható le.



Ábra 2

Ez az adatbázismodell (Ábra 2) az alapja a keresésnek is, mely történhet kulcsszó alapján, figyelembe véve az összes bejegyzést meta-adat típustól függetlenül, vagy megadva az adattípus-érték párost.

Ugyanakkor az egységesebb leírás érdekében bevezettünk egy, hasonlósági szinteken alapuló kulcsszó-ajánlást. Mi esetünkben a hasonlóságot a jegyzeteket leíró kulcsszavak közös kulcsszavak száma jelenti. Így az n -edik szinten azok a jegyzetek lesznek, melyek n darab közös kulcsszót tartalmaznak az adott jegyzettel.

Felhasználói felület

Egy digitális könyvtár hasznossága nemcsak a digitalizált anyag mennyiségétől függ, hanem attól is ahogyan ezt a képanyagot a felhasználó rendelkezésére bocsátják és ahogy bemutatják a gyűjteményt.

Célunk egy olyan felhasználói felület létrehozása mely segítségével az érdeklődők gyorsan és könnyedén böngészhetik a kéziratokat. Ennek érdekében próbáltunk a papír metaforához igazodni a digitalizált anyag bemutatásakor, valamint a felhasználó más dokumentumnézegető alkalmazással felgyűlt tapasztalatára alapozni. Hasonlóképpen ahogy más alkalmazásokban a felhasználók az egér és gördítősávok segítségével tudnak navigálni az oldalak között, így Armarius-ban is ezeket az eszközöket használhatjuk az oldalak közti eligazodásban. Jegyzeteket is egérrel helyezhetünk el az oldalon, valamint ha ezek fölé helyezzük az egérmutatót, megjelenik egy ablak a jegyzethez csatolt meta-adatokkal.

Fontos szempont volt a kezelőfelület tervezésekor az információ lokalitása, vagyis az hogy egy adott objektumhoz kapcsolódó információk és parancsok a az adott objektumhoz „közel” jelenjenek meg a képernyő síkjában, hiszen ez segít a felhasználónak összekapcsolni az megjelenített adatot vagy parancsot a hozzá illő objektummal. Így a jegyzetekhez kapcsolt meta-adatok szerkesztése is a jegyzet mellett megjelenő kis ablakban történik. Ugyanitt lesznek elhelyezve a felhasználók jegyzethez kapcsolódó megjegyzései.

Szükség lesz egy egyszerű HTML kéziratböngésző oldal kifejlesztésére is, hiszen nem minden böngésző képes JavaScriptet megfelelőképpen futtatni. Felmerül ugyanúgy annak a lehetősége is hogy egy DjVu alapú böngészőoldalt fejlesszünk ki, ez jelentősen lecsökkentene az oldalak betöltési idejét. A hátránya ennek a megközelítésnek az hogy a megjelenítést egy Java applet-ben kell végezni, aminek lefuttatására egy Java virtuális gép szükséges. Sajnos ez nincs minden

számítógépre telepítve, és a telepítés gondot jelenthet egyes felhasználóknak. Fontos az hogy a felhasználó legyen az aki kiválassza hogy a három kezelőfelület közül melyiket használja

A rendszer felépítése

Jelenlegi rendszerünk az AJAX alkalmazásmodell-t követi, így két komponenst különíthetünk el a rendszer struktúrájában: szerver oldali komponenst és kliens oldali komponenst. A szerver oldalon azok a PHP script-ek helyezkednek el, amik felelősek a keretoldal, az oldalakat és jegyzeteket leíró XML állományok kigenerálásával, valamint a felhasználó által módosított jegyzetek elmentésével. A kliens oldalon az a JavaScript kód található ami a felhasználó böngészőjében fut. Ez felelős a böngészőoldal létrehozásáért, az oldalak automatikus betöltéséért, a jegyzetek megjelenítésért és elmentéséért. A két rész egymással a HTTP protokollon keresztül kommunikál. A kliens oldali kód a *XmlHttpRequest* JavaScript objektumon keresztül egy HTTP lekérdezést küld a szervernek, ami meghívja a megfelelő PHP script-et, a script által generált XML-t visszaküldi a kliens böngészőjének. A JavaScript kód a válaszként kapott XML-t az *XmlHttpRequest* objektum *responseXML* mezőjén keresztül éri el, majd létre hozza a megfelelő DOM csomópontokat ezzel megjelenítve az oldalt vagy a jegyzetet. A kéziratböngészőben mindig fix számú oldal van betöltve, szükség eseten újabb oldalak töltődnek be és jelenlegi látható felületen kívül eső oldalak törlődnek.

Tervek a jövőre

Terveink szerint a rendszer regisztrációt és bejelentkezési funkciót is fog biztosítani, ami alapján megvalósulhat a felhasználó-követés. A regisztráció által lehetőség nyílik a személyreszabott rendszer létrehozására, a felhasználók módosíthatják a saját jegyzeteiket, saját nézeteket hozhatnak létre.

Ugyanakkor egy különösen hasznos tervezett funkció a felhasználó-követés, amely a múlt eseményeit figyelembe véve nyújt segítséget a felhasználónak. Így például az előbbi keresések során begyűjtött információ segítségével azon dokumentumokat ajánlja fel a felhasználónak, amelyeket legnagyobb valószínűséggel keres. Ugyanez történik a kulcsszó-ajánlás esetében is, így egységesebbé és könnyebbé válik a kéziratok leírása.

Fontos kérdés az is hogyan osszunk ki jogokat a felhasználóknak, ki mihez férhet majd hozzá illetve mit módosíthat. Itt talán a legmegfelelőbb megoldás az adatbázis kezelőkből átvett jogátruházásos privilégium-modell, amely során egy felhasználó átruházhatja saját jogait más felhasználónak. Ezzel valamennyire elkerülhetjük a wikikben nagyon gyakori vandalizmust, és megszabadítjuk a rendszergazdát a jogosultságok egyenkénti kiosztásától.

A felhasználó-követés mellett fontos szempont a kapcsolatteremtés lehetőségét biztosítani a felhasználóknak. Egy fórumon értékes információkat, tapasztalatot, megfigyeléseket cserélhetnek a világ különböző pontjairól érkező látogatók, ami kutatók esetében különösen jelentős, és reményeink szerint elősegíti a Bolyai kéziratok eddig rejtve maradt titkainak feltárását.

Záró gondolatok

A rendszer első célja hogy egyszerű, könnyen kezelhető felhasználói felülettel lehetővé tegye a régi dokumentumok manuális leírását. Ezáltal nemcsak a kutatók munkája válna könnyebbé, hanem a Teleki Téka felbecsülhetetlen értékű gyűjteményének egy része elérhetővé válna a nagyközönség számára, ezáltal nőne a Téka ismeretség és hozzájárulna a két zseniális matematikus: Bolyai Farkas és Bolyai János munkájának megőrzéséhez és terjesztéséhez.

A rendszer fejlesztése tovább folyik egy „élő digitális könyvtár” fele.

Referenciák

1. Gallica, la bibliotheque numérique, 2007, <http://gallica.bnf.fr/> (online)
2. Bibliotheca Corviniana Digitalis, 2007, <http://www.corvina.oszk.hu/> (online)
3. Early Manuscripts at Oxford University, 2001, <http://image.ox.ac.uk/> (online)
4. Koninklijke Bibliotheek National Library of the Netherlands, 2004, <http://www.kb.nl/>, (online)
5. Digital Image Archive of Medieval Music, 2006, <http://www.diamm.ac.uk/> (online)
6. The William Blake Archive, 2007, <http://www.blakearchive.org/blake/> (online)
7. 12. E. Egyed-Zsigmond, S. Lajmi, Z. Iszai, 2006 : Concurrent use in an image management system : 13th ispe International conference on concurrent engineering: Research and applications, CE2006, Antibes, September 2006, IOS Press, pp. 403-417
8. 16. Z. Iszlai, E. Egyed-Zsigmond. User centered image management system for digital libraries, 2nd IEEE International Conference on Document Image Analysis for Libraries, Lyon, April 2006, pp. 164-171, ISBN: 0-7695-2531-8
9. A. Blandford, Understanding users' experiences: evaluation of digital libraries. DELOS workshop on evaluation of digital libraries, Padova, 2004. pp. 31-34