



Digitális könyvtárak együttműködési modelljei

Szakirodalmi áttekintés

Bevezetés

A digitális könyvtárak a rend szigetei a web kaotikus információs tengerében. Fontos szerepük van a civilizáció előrehaladásában, mert igen hatékonyan, földrajzi, nyelvi és kulturális határokat átlépve, bármikor és bárhol hozzáférhetővé teszik a felhalmozott emberi tudást. Az még azonban kevés, hogy rájuk bukkanunk egy keresőgép vagy egy tematikus webkatalógus segítségével: ezután be kell lépni valamelyikbe, ott keresgélni vagy böngészni, majd továbbmenni a következőre, ha nem találtuk meg, amire szükségünk van. Felhasználói szempontból fontos lenne tehát, hogy ezek a digitális gyűjtemények „együttműködjenek”, vagyis hogy a különböző számítógépek, hálózatok, operációs rendszerek és alkalmazások „értsék egymást”. Ahogy *Suleman* [1] fogalmazott a diszszertációjában: az „interoperabilitás” azt jelenti, hogy egy e-könyvtár együttműködik más hasonló gyűjteményekkel annak érdekében, hogy minél jobb szolgáltatást tudjon nyújtani felhasználóinak. *Paepcke* és szerzőtársai [2] pedig már 1998-ban arra hívták fel a figyelmet, hogy az együttműködés a digitális gyűjtemények fejlesztésének, illetve továbbfejlesztésének kulcskérdése. Ezt erősítette meg évekkel később *Warren* és *Alsmeyer* [3] is, akik kimutatták, hogy a szakterület egyik legtöbb vitatott témája az egyes szolgáltatások közti kooperáció.

Arms 2002-es cikkében [4] az együttműködés három szintjét vagy modelljét különböztette meg: szövetkezési (*federated*), aratásos (*harvested*) és összegyűjtési (*gathering*). Természetesen a valóságban ezek kombinációi is előfordulnak, vagyis egy digitális könyvtár többféle módon is kapcsolódhat más gyűjteményekhez. Az interoperabilitásnak ezen kívül műszaki, tartalmi és szervezési szintjei is vannak: az elsőhöz tartoznak például a formátumok, a protokollok, a biztonsági megoldások; a másodikhoz a metaadatok és azok

jelentésének értelmezése; a harmadikhoz pedig olyan kérdések, mint például a hozzáférés szabályozása, a felhasználók azonosítása, a fizetés módja. A jelen tanulmányban az iráni szerzők az első kérdéssel, vagyis a digitális könyvtárak együttműködésének technikai megoldásaival foglalkoznak, áttekintve a témában 2008 februárjáig megjelent szakirodalmat és bemutatva a különböző modelleket, illetve az azokat alkalmazó projekteket. A feldolgozandó publikációkat a *digital library*, az *interoperability*, a *meta-searching* és a *cross-searching* kifejezések kombinálásával keresték ki különböző nagy szakirodalmi adatbázisokból.

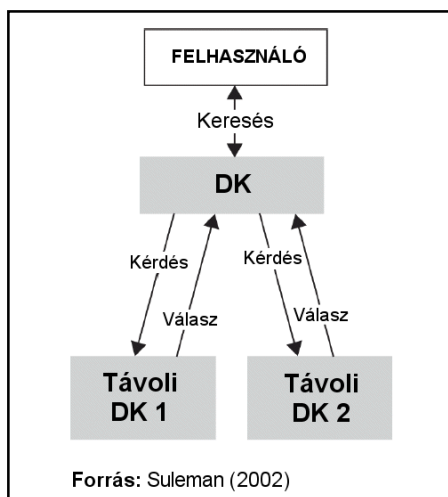
Együttműködési modellek

Több tényező befolyásolja, hogy a különféle technológiákra épített digitális archívumok milyen megoldást választanak az együttműködéshez. Függ például attól, hogy mikor fejlesztették ki a rendszert, mennyi pénz volt erre a célra, és hogy milyen hatékonyságot akartak elérni a fejlesztők. Léteznek általánosan elterjedt egyszerű szabványok, amelyek esetleg kevésbé hatékonyak az adott célra, és vannak újszerű megoldások is (pl. Java-alapú alkalmazások), amelyek célszerűbbek ugyan, de lehet, hogy kevesebben tudják őket használni. Néha elegendő egy gyors, egyszerű, szélesebb kört lefedő, de felületes megoldás, és természetesen vannak olyan digitális gyűjtemények is, amelyeknél megéri egy költséges, de professzionálisabb eredményt adó technológiát választani. Az említett három modell közül a szövetkezési megoldás jelenti a legnagyobb terhet a partnereknek, míg az összegyűjtési módszerrel gyakorlatilag nincs is feladatuk a tartalomgazdáknak – igaz, ez adja a leggyengébb végeredményt. Az aratásos megoldás a kettő között van a megvalósítás nehézsége és a hatékonyság szempontjából. Az sem ritka – ahogy a következő példák mutatják –, hogy egyes projekteknél több modellt is

alkalmaznak, különböző partnerek felé vagy eltérő célokra: például Z39.50 és OAI alapon egyaránt folytatnak adatcserét.

Szövetkezési modell

Ez az együttműködések leghagyományosabb formája, amikor is a partnerek megegyeznek, hogy közös megoldásokat és szabványokat fognak alkalmazni. Ilyen szövetkezés esetén a keresés menete az 1. ábrán látszik: a digitális könyvtár (DK) elküldi a felhasználó keresőkérdését több távoli repozitóriumnak is (DK 1, DK 2), majd ezek válaszait begyűjti, összefésüli és továbbítja a kérdezőnek. A föderált modellnél a leggyakrabban használt szabvány a *Linked Systems Project* keretében kifejlesztett Z39.50 protokoll, melynek első változata még 1984-ben jelent meg, és amelyet régóta használnak a könyvtárak bibliográfiai adatok lekérdezésére. A Z39.50 kommunikáció kliens/szerver elven működik: a kliens program tartja a kapcsolatot a felhasználókkal, illetve a szerverrel, a szabványban meghatározott keresőnyelv és kommunikációs protokoll segítségével. Ezt a modellt köztes program (*middleware*) segítségével is meg lehet valósítani – ilyenkor a közvetítő szoftverréteg kapcsolódik a szerverekhez és a felhasználó ennek köszönhetően hozzáfér más gyűjtemények tartalmához is.



1. ábra A szövetkezési modell folyamata

A föderált keresés néhány konkrét megvalósítása:

- A *Columbia University* PERSIVAL nevű projektje az *SDLIP* (*Simple Digital Library Interoperability Protocol*) és a *STARTS* (*STANford protocol proposal for Internet ReTrieval and Search*) ajánlások kombinálásával valósítja meg különféle do-

kumentumgyűjtemények közös lekérdezését. Az egyesített protokoll: a SDARTS meglehetősen egyszerű, de mégis kifinomult metakeresőket lehet készíteni erre alapozva.

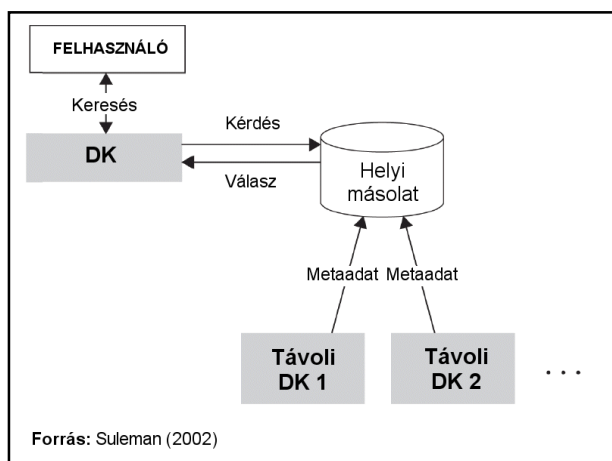
- A *Greenstone* digitális gyűjteménykezelő is támogatja a Z39.50 szabványt. Erre a célra egy új szerveret fejlesztettek ki a rendszer programozói, amely a nyílt forráskódú YAZ programkönyvtárat használja előtérnek (*front end*), vagyis ez fogadja a Z39.50-es kliensektől érkező kéréseket, majd lefordítja őket a Greenstone saját protokolljára. A Greenstone válaszait pedig szintén a YAZ-on keresztül alakítják vissza Z39.50 nyelvű kifejezésekkel és küldik vissza a kliensnek.
- A *Colorado State Library* által 1998-ban indított digitalizálási programban szintén a Z39.50-es protokollt használják a gyűjtemények közös lekérdezésére. Továbbá egy *DC Builder* alkalmazást is létrehoztak azért, hogy a különféle rendszerű és formátumú metaadatokat egységesen Dublin Core-ra konvertálják.
- Illinois állam 12 könyvtára a ZMARKO projektnek köszönhetően teszi elérhetővé a bibliográfiai rekordjait. Egy olyan adatszolgáltatást fejlesztettek ki, amely a Z39.50 szerverrel lekérdezhető MARC rekordokat OAI-PMH protokoll szerint is tudja szolgáltatni (vagyis egy OAI-PMH/Z39.50 gateway). Így ez a rendszer a *federated* és a *harvested* modellt egyaránt támogatja.
- A *Mountain West Digital Library* program Utah állam egyetemi könyvtárainak együttműködése 2002 óta. A metaadatok cseréje itt az Open Archives Initiative protokoll alapján zajlik, de 2003-ban egy ZContent nevű, nyílt forráskódú kiegészítő modult is kifejlesztettek Perl nyelven, ami lefordítja a Z39.50-es kérdéseket, vagyis ez is egy kéttípusú rendszer.

Aratásos modell

Sokszereplős szövetségeket nehéz létrehozni és fenntartani, ezért újabban már inkább lazább együttműködési formákat választanak a digitális könyvtárak. Ezekben az esetekben a résztvevőknek elég csak néhány közös alapfunkcióban megegyezni és nem kell egy egész sor részletkérdésben megállapodniuk. Az Open Archives Initiative is ezt az elvet valósítja meg a metaadatok aratásához. Ennél a modellnél a digitális könyvtárak egyszerű csereformátumban teszik elérhetővé az adataikat, melyeket azután különféle metaadatszolgáltatók begyűjtenek egy szabványos protokollt használva. A metaadatok aratásának ötlete már az 1990-es évek elején felmerült, de akkor még nem terjedt el. 1998-ban azután egy *Univer-*

sal Preprint Server elnevezésű prototípusban valósították meg. Ez az alkalmazás a különféle preprintarchívumok közös lekérdezését célozta, és ebből a kísérletből alakult azután ki az OAI, amely egyszerű és olcsó megoldás azoknak a digitális könyvtáraknak, amelyek meg akarják osztani a metaadataikat másokkal.

A 2. ábrán látható ez a folyamat: az együttműködő könyvtárak felállítanak egy szervert, ahová mindegyik résztvevő egy közönséges HTTP kapcsolaton át elküldi a metaadatait. A felhasználók ezután ezekben a felmásolt metaadatokban tudnak keresni, vagyis egyben látják a különböző gyűjtemények anyagát. Olyan megoldás is elképzelhető, hogy kétirányú kapcsolat van a *harvester* szerver és a partnerkönyvtárak között: ez esetben mindegyik partner letükrözi magának a szerveren levő integrált adatbázist. Az is lehetséges – bár elég komplikált – megvalósítás, hogy a résztvevők szintén kétirányú kapcsolatokon át, közös szerver fenntartása nélkül, közvetlenül egymásnak küldik át a metaadataikat.



2. ábra Az aratásos modell folyamata

Néhány példa az aratásos metaadat-gyűjtésre:

- A *SIGAL* (*Système d'Information Géographique et Agent Logiciel*) projektneél a *C-FOAL* (*Conception par Frameworks Orientés-Agents Logiciels*) módszer segítségével hoztak létre együttműködési környezetet georeferált digitális gyűjteményekhez. A fejlesztésnél három keretrendszerből (szerver, kliens és logikai forrás) alakították ki a rendszer architektúráját.
- Az *NDLTD* (*Networked Digital Library of Theses and Dissertations*) elnevezésű együttműködésnek 13 tagja van a világ különböző pontjain. A közös lekérdezéshez OAI protokollt használnak,
- az alkalmazott metaadatszabvány pedig az *ETDMS* (*Electronic Theses and Dissertation Metadata Set*). A *National Science Digital Library* szintén az OAI-t választotta ahhoz, hogy egy központi adatbázisba gyűjtse a metaadatokat.
- A *University of Michigan* OAIster rendszere OAI aratást végez: a DC metaadatokat XML formátumban gyűjti be, majd ezeket XSLT segítségével a projekt saját bibliográfiai formátumára: DLXS-re alakítja át, leindexeli őket és az XPAT keresőmotorral lekérdezhetővé teszi az adatbázist a felhasználók számára.
- A *NASA* is az OAI-t használja ahhoz, hogy hozzáférést nyújtson a világ legnagyobb űrkutatási és űrtechnológiai információforrásaihoz az USA-ban levő 10 központjában, illetve parancsnokságán, valamint több mint félszáz egyéb országban. A bibliográfiai adatok szabványa a Dublin Core, a teljes szövegű dokumentumok pedig PDF formátumban érhetők el.
- Ugyanezekre a szabványokra alapozták a *CNDLTD* (*China Networked Digital Library of Theses and Dissertations*) projektet, amely még 1996-ban indult és a *CALIS* (*China Academic Library and Information System*) konzorcium irányítja. Az együttműködésben részt vevő kínai felsőoktatási könyvtárak egy központi adatbázisba szolgáltatják be a tudományos fokozatok eléréséhez szükséges értekezések metaadatait.
- A tajvani *National Science Council* 2002 januárjában indította el az *NDAP* (*National Digital Archives Program*) projektet. Ebben az együttműködésben számos egyetem és kutatóintézet vesz részt. Az OAI-PMH segítségével gyűjtik be a DC metaadatokat XML formátumban.
- A *Hong Kong University of Science and Technology* hasonlóan jó példa az OAI-PMH használatára: a DSpace-re alapozott gyűjteményük adatait az OAIster és a Scirus is leartja, így az egész világ számára megtalálhatókká válnak. Ők minősített Dublin Core-t (qDC) használnak a repozitórium metaadataihoz, maguk a dokumentumok pedig PDF-ben tölthetők le.
- A CDS/ISIS-alapú rendszerek számára a CDSOAI projekt teremti meg az interoperabilitás lehetőségét. A CDSOAI egy köztes szoftverréteg az adatforrások és a keresőszolgáltatást biztosítók között, és hasonlóan nyílt forráskódú fejlesztés, mint az ARC nevű, OAI-PMH-t használó Java-alapú programcsomag.
- Bell és Lewis [5] 2006-ban publikált cikkükben az OAI-PMH és a METS használatát vizsgálták a különböző rendszerű repozitóriumok közötti metaadat- és digitális objektumcsere szempontjából. Háromféle alkalmazást elemeztek: a *Uni-*

versity of Wales Aberystwyth könyvtárában működő DSpace-re alapozott repozitóriumot, a *National Library of Wales* Fedora-alapú gyűjteményét és az *ETHOS* (*Electronic Theses Online Service*) által használt UKETD rendszert, s végül az OAI-PMH-t javasolták az elektronikus formában levő disszertációk cseréjére a három repozitórium között. Shen [6] pedig, szintén 2006-os PhD dolgozatában, a digitális gyűjtemények számára kidolgozott 5S keretrendszer alkalmazhatóságát vizsgálta meg abból a szempontból, hogy hogyan lehetne ezen az elven egy integrált digitális könyvtárat kialakítani a régészeti témájú elektronikus archívumokból.

Összegyűjtéses modell

Ha semmilyen együttműködés, közös alap nincs a különféle tartalomszolgáltató szervezetek között, akkor is el lehet érni egy minimális interoperabilitást úgy, hogy egy webes keresőgéppel összegyűjtjük a nyilvánosan elérhető információkat. Tekintve, hogy ebben az esetben a gyűjtemények részéről semmilyen feladat nincs, nagy mennyiségben lehet így összehozni különböző digitális könyvtárak tartalmát, de persze a közös kereső által nyújtott szolgáltatás színvonala alacsonyabb lesz, mint amikor a partnerek közvetlenül együttműködnek. Mindenesetre az új webes fejlesztések egy része éppen az interoperabilitás javítását célozza, úgyhogy, bár a szemantikus web teljes körű megvalósítása még csak szép álom, a jövőben várhatóan egyre jobb szolgáltatásokat lehet majd építeni a keresőrobotokkal begyűjtött adatokra is. A *ResearchIndex* (korábbi nevén: *CiteSeer*) kiváló példa a nyilvános információk teljesen automatizált összegyűjtésére alapozott digitális könyvtári szolgáltatásra. Megemlíthető még, hogy az ingyenes *Greenstone* gyűjteménykezelő rendszerben is megvalósították ezt a *gathering* modellt – a *federated* és a *harvested* mellett – az *Exchange Center* nevű szoftverkomponensben.

Következtetések

A könyvtárak mindig és mindenhol törekednek az együttműködésre, hogy más gyűjteményekhez is hozzáférést tudjanak nyújtani és ezáltal minél jobban kiszolgálhassák a használóikat. Az együttműködés konkrét formái a mindenkori fejlettségi szinttől függenek. A digitális könyvtárak viszonylag új szolgáltatások és a világméretű web virtuális környezetében működnek. Ebben a világban olyan

gyors változások zajlanak, hogy a legtöbb eszköz és protokoll élettartama nagyon rövid. Amelyeknek sikerül hosszabb ideig fennmaradni, azok válnak általánosan elterjedté és a használóik elkezdik támogatni őket. Elképzelhető, hogy vannak az itt ismertetetteken túl még egyéb interoperabilitási technológiák is, de azok vagy nagyon újak, vagy nem használják ezeket a könyvtárak, vagy még nem publikálták őket a fontosabb szaklapokban.

A témába vágó publikációkat áttekintve megállapítható, hogy az első együttműködésekéről szóló beszámolók 1998 körül jelentek meg, tehát nem sokkal a digitális könyvtárak megszületése után. A szakcikkek száma 2003 óta nőtt meg jelentősen, ami azt jelzi, hogy egyre nagyobb jelentőséget kap a kooperáció a digitális gyűjtemények fejlesztésében. Az ilyen jellegű projekteknek mintegy fele az Egyesült Államokhoz kötődik, és itt dolgozták ki az olyan fontos szabványokat is, mint az OAI és a Z39.50. Megfigyelhető az is a szakirodalmat elemezve, hogy bár a Z39.50-nek nagy múltja volt a könyvtári szférában, és széles körben használják hagyományos dokumentumok bibliográfiai adatainak közös lekérdezésére, a digitális könyvtáraknál a több lehetőséget nyújtó és egyszerűbben megvalósítható OAI protokoll terjedt el. Bár a jelen tanulmány a metaadatok szintjén zajló együttműködések vizsgálta, meg kell említeni még, hogy egyre gyakrabban merül fel feladatként maguknak a teljes szövegű dokumentumoknak a cseréje is az egyes digitális gyűjtemények között, amihez szintén kellően szabványosított formátumokra és eljárásokra van szükség.

Hivatkozások

- [1] SULEMAN, H.: Open digital libraries. Dissertation for the degree of Doctor of Philosophy in Computer Science and Applications. Blacksburg, VA., Virginia Polytechnic Institute and State University, 2002.
- [2] PAEPCKE, A. – CHANG, C. – GARCIA-MOLINA, H. – WINOGRAD, T.: Interoperability for digital libraries worldwide. = Communications of the ACM, 41. köt. 4. sz. 1998. p. 33–43.
- [3] WARREN, P. – ALSMEYER, D.: Applying semantic technology to a digital library: a case study. = Library Management, 26. köt. 415. sz. 2005. p. 196–205.
- [4] ARMS, W. Y.: A spectrum of interoperability. = D-Lib Magazine, 8. köt. 1. sz. 2002.
<http://www.dlib.org/dlib/january02/arms/01arms.html> (letöltve: 2008.04.10.).
- [5] BELL, J. – LEWIS, S.: Using OAI-PMH and METS for exporting metadata and digital objects between

repositories. = Program: electronic library and information systems, 40. köt. 3. sz. 2006. p. 268–276.

- [6] SHEN, R.: Applying the 5S framework to integrating digital libraries. Dissertation for the Doctor of Philosophy in Computer Science and Application. Blacksburg, VA., Virginia Polytechnic Institute and State University, 2006.

/ALIPOUR-HAFEZI, Mehdi – HORRI, Abbas – SHIRI, Ali – GHAEBI, Amir: Interoperability models in digital libraries: an overview =. The Electronic Library, 28. köt. 3. sz. 2010. p. 438–452./

(Drótos László)