

MI A MIA?

JAVASLAT EGY MAGYAR INTERNET-ARCHÍVUM LÉTREHOZÁSÁRA

Drótos László, mekdl@ijf.hu

Országos Széchényi Könyvtár, Magyar Elektronikus Könyvtár Osztály

Hogy miért fontos a digitális kultúra megőrzése (különösen az eleve digitálisan születő alkotásoké), azt – remélhetőleg – már nem nagyon kell magyarázni. Látjuk, érezzük, hogy mennyire meghatározó lett egy évtized alatt nálunk is az Internet, mennyi minden már csak ott érhető el; és látjuk, érezzük azt is, hogy rohamosan növekszik az online tartalom – és rohamosan pusztul is. Nemcsak egy-egy link törik el itt-ott, és vezet egy *404-es hiba* oldalhoz, hanem egész site-ok – oldalak ezrei – tűnnek el nyomtalanul egyik napról a másikra, vagy válnak fokozatosan a digitális entrópia áldozatává. A becslések változnak, de a pesszimistább mérések szerint egy web-lap átlag-életkora nagyjából egy házilégy egy hónapos élettartamával egyezik meg. És miközben országos felháborodást vált ki, ha a rossz tárolási körülmények miatt meg kell semmisíteni a Nemzeti Könyvtárban őrzött kisnyomtatványok egy részét, vagy csőtörés miatt eláznak régi folyóiratok, a digitális “nyomtatványok” megmentését és megőrzését senki nem érzi feladatának. Legalábbis nálunk nem, mert a világ más részein már elég sok munkát elvégeztek ezen a téren, és – bár a feladat kétségtelenül nagyon nehéz – vannak már biztató eredményeik is.

Az UNESCO Közgyűlésének 32. ülészakán, 2003. október 17-én elfogadott “Charta a digitális örökség védelméről”¹ világosan megfogalmazza az ezzel kapcsolatos feladatokat, és külön kiemeli, hogy a válogatott archiválás esetén a “born digital” anyagoknak prioritást kell adni. Ez a dokumentum meghatározza a felelősöket is: “*A digitális örökség megőrzése a kormányok, alkotók, kiadók, releváns iparágak és az örökségvédelmi intézmények kitartó erőfeszítéseit igényli.*”

A nemzeti könyvtáraknak természetesen kiemelt felelősségük van, bár az nagy kérdés, hogy a hálózaton levő rengeteg féle digitális objektumból mi tekinthető a gyűjtőkörük részének. Ezen a téren nagyon eltérő a nemzetközi gyakorlat is: van ahol csak a “kiadványnak” minősíthető dokumentumokat gyűjtik ezek a könyvtárak (azokat is önkéntes letét vagy kötelezpéldány alapon), van ahol meg az egész nemzeti web-teret learatják időről időre és elteszik valami tartósabb tárolóra. Magyarországon 1994-ben egy-két felsőoktatási könyvtárban kezdődött meg a digitális dokumentumok gyűjtése és archiválása: a Magyar Elektronikus Könyvtár² nevű kezdeményezéssel. A MEK projektet 1999-ben az Országos Széchényi Könyvtár vette át és azóta egy néhány fős önálló osztály is kialakult mögötte. A MEK a monografikus jellegű, lehetőleg terjedelmesebb, online vagy offline publikált, nyilvánosan szolgáltatható, alapvetően szöveges digitális hungarikumok gyűjtését vállalta fel, vagyis lényegében egy

¹ <http://www.unesco.hu/index.php?type=node&id=508>

² <http://mek.oszk.hu>

erősen válogatott könyvgyűjtemény (beleértve most már az MP3 hangoskönyveket is). Az Osztály munkatársai 2003-ban elindították az Elektronikus Periodika Archívum és Adatbázis³ nevű szolgáltatást is, melynek adatbázis része teljességre törekedve tartja nyilván a hungarikumnak tekinthető, online vagy offline formában létező, elektronikus időszaki kiadványokat; archívum része pedig elsősorban folyóiratokat és hírleveleket ment le, őriz és szolgáltat. A MEK nyitólapjáról elérhető “kiállítóteremben”⁴ pedig már arra is akad néhány példa, hogy teljes site-ok, honlapok vannak archiválva (természetesen a készítőik beleegyezésével, sőt kifejezett kérésére), például: “Erdélyi és csángó költészet”, “Váli Dezső oeuvre”, “Vészi Endre honlapja”, “Lénárd Sándor honlapja”. Van tehát már némi gyakorlatunk abban, hogy hogyan lehetne a magyar web-nek legalább a legértékesebb részét elmenteni és használhatóvá tenni mind a jelen, mind a jövő számára. Ez a tapasztalat arra már elegendő, hogy lássuk, milyen nagy és komplex ez a feladat, mennyi katalogizálási, technikai és jogi problémát vet fel, és hogy egyetlen intézmény vagy intézménytípus önmagában nem tudja megoldani egy Magyar Internet Archívum (rövidítsük egyelőre MIA-ként) létrehozását és fenntartását. Ezért az OSZK MEK Osztálya javasolja egy konzorcium alakítását, melyben a közgyűjtemények mellett informatikai intézmények és cégek is társulnak a MIA megvalósítása érdekében.

A továbbiakban nézzük át, hogy miből áll és máshol hol tart ez a munka, és hogy mi – az eddigi tapasztalataink alapján – mit tartunk reálisan járható útnak itt és most:

Technikai vonatkozások:

Az “Internet archívum” vagy “web-archívum” kifejezés többféle dolgot is jelent a szak- és köznyelvben. Az alábbiakban a komplett site-ok, szolgáltatások időszakosan ismétlődő lementésével létrejövő másolatot értem ez alatt. Nem tekintem ide tartozónak a MEK-hez vagy az EPA-hoz hasonló, egyedi dokumentumok vagy egyedi kiadványok egyszeri vagy folyamatosan bővülő mentéseit.

De a teljes szolgáltatások mentése is kétféle archívumot eredményezhet:

1. Fájlrendszerbe való mentés

Ennél a megoldásnál valamilyen letöltőprogram segítségével (HTTP vagy FTP protokollon át) egy másolat készül egy adott web-helyről, amennyire csak lehet megőrizve annak eredeti arculatát és struktúráját. A mentés során a belső linkeket relatív útvonalra kell konvertálni, így a másolat ugyanúgy navigálható maradhat, mint az eredeti. Természetesen a szerver oldalon futó scriptek nem biztos, hogy működnek az archív szerveren is (ezek letöltéséhez amúgy is a web-mester engedélye kell), és ha adatbázis van az eredeti szolgáltatás mögött, akkor a teljes szoftver és operációs rendszer környezetet meg kell teremteni az archiváló gépen is a funkcionalitás megőrzéséhez. A csak statikus HTML dokumentumokból álló site-ok könnyen és jó minőségben elmenthetők így, de ezeknél is szükség van egy fájl-menedzsment rendszer kialakítására, a rohamosan szaporodó állományok nyilván- és karbantartása érdekében; valamint a tárhelykímélés céljából hasznos duplikátum-szűrés is komoly feladat, ez szintén a belső linkek átalakítását igényli. Ez a technika rosszul skálázható, tömeges és gyakori mentésre nem alkalmas, viszont a felhasználók számára jól böngészhető, valódi “web-múzeum” érzést kelt.

³ <http://epa.oszk.hu/>

⁴ <http://mek.oszk.hu/html/kiallitas.html>

2. Adatstruktúrába való mentés

Ilyenkor egy harvester vagy crawler robotot alkalmaznak, amely akár egy teljes top-level domain tartalmát is le tudja szüretelni, A begyűjtött anyagot egységes szerkezetű (pl. XML-re konvertált, metaadatokkal ellátott és tömörített) archív állományokba teszik, majd adatbázist és indexeket készítenek hozzájuk. Az archivált anyag egyes részeinek elérése vagy valamilyen URI azonosító alapján vagy teljes szövegű kereséssel történhet. A kikeresett web-lapról a további navigálás vagy csak nagyon korlátozottan lehetséges az archívumon belül, vagy a belső linkek az eredeti forrásra visznek tovább (amennyiben az még létezik). Egy ilyen archívum felépítése komolyabb technikai feladat, viszont jól skálázható, hatalmas mennyiség gyűjthető így be, szinte teljesen automatizáltan. Azoknak a felhasználóknak jó, akik böngészés helyett célzottan keresni szeretnének egy nagy archívumban.

A fentén kívül természetesen másféle felosztások is lehetségesek, például besorolhatjuk az internet-archívumokat médiatípus szerint (web-, newsgroup-, sugárzott multimédia-gyűjtemények stb.), vagy válogatási szempontok szerint (teljes nemzeti web-teret vagy egyéb nagyobb domaint gyűjtő, vagy csak minőségi web-helyeket, vagy nagyobb témaköröket illetve csupán néhány témát vagy egy-egy eseményt gyűjtő archívumok), vagy pedig a gyűjtés iránya szerint (pull-típusú lementés, illetve push-típusú önkéntes vagy kötelező feltöltés). És természetesen ezeknek különböző variációi is elképzelhetők, s vannak is már működő példák rájuk.

Egy komoly archívum több szoftverkomponensből áll össze, “kulcsrakész” rendszerek nem nagyon vannak ezen a téren; a már működő rendszerek részben saját fejlesztésű, részben kész – többnyire nyílt és ingyenes – elemekből lettek összerakva. A letöltő illetve szüretelő programok mellé kell egy adatbáziskezelő a metaadatok tárolására, egy teljes szövegű indexelő és kereső, valamint egy szolgáltatási felület. Az egész munkafolyamat kézbentartására és a minőségbiztosítás céljára pedig ki kell alakítani egy menedzsment keretrendszert. Az archiválás során rengeteg technikai problémával szembesül az üzemeltető: az Internet – és azon belül a web önmagában is – egy nagyon bonyolult, gyorsan változó, nehezen megőrizhető médium. Hosszú távú és mindenre kiterjedő megoldás – a papírra nyomtatáson kívül :-) – még sehol sem létezik. Nagyon nagy szükség volna egy olyan W3C ajánlásra, amely – az online szolgáltatások akadálymentesítéséhez hasonlóan – javaslatokat tartalmazna arra vonatkozóan, hogy hogyan kell úgy kialakítani egy web-helyet, hogy az könnyen és hosszú távon is archiválható legyen. Ez a fajta akadálymentesítés úgy is megoldható lehet egyébként, hogy az archiváló robotok számára egy alternatív verziót “exportál ki” a site üzemeltetője (pl. statikus HTML lapok formájában), nem muszáj a nyilvános szolgáltatást átalakítani emiatt.

Külföldi példák:

A világban 1-2 tucat ilyen, nagy nyilvános szolgáltatás illetve projekt létezik, ezek jellegükben és méretükben is erősen különbözőek. Jellemző, hogy sok köztük a még csak pilot-jellegű próbálkozás, amelyek csupán egy előre lehatárolt, néhány éves időszakra terjednek ki, kevés a már üzemszerűen működő, valóban nagy archívum. Ugyancsak jellemző mostanában az archívumok integrálódása, a közös szabványokra, technológiákra és a munkamegosztásra való törekvés.

A legrégebb és legismertebb kezdeményezés természetesen az *Internet Archive*⁵ nevű non-profit szervezet San Franciscóban, amely a web-archiváláson, mint alapcélon túlmenve, a

⁵ <http://www.archive.org>

“könyvtár” fogalmát a legszélesebb értelemre kiterjesztve, a digitális objektumok Alexandriai Könyvtárává szeretne válni. De legismertebb szolgáltatásuk azért az Alexa Internet cég (jelenleg az Amazon.com tulajdona) által összegyűjtött anyagra épülő Wayback Machine. Ez 2006 februárjában mintegy 55 milliárd web-oldalt tud előkeresni, URL alapján. Most folyik a Nutch nevű teljes szövegű kereső beépítése. A szüretelés 1996-ban indult és elvileg a teljes nyilvános web-re kiterjed, de a népszerűbb site-okat gyakrabban begyűjtik (az átlag gyakoriság 2 hónap, az átlag növekedés havi 20 terabyte), az anyag 6 hónap késéssel válik nyilvánossá, de az archívum egyes részei csak kutatóknak érhetőek el. Tipikusan a 2. csoportba tartozó rendszer: a mintegy 1 petabyte-nyi anyag megfelelő részének előkeresése az archív fájljából meglehetősen hosszú válaszidőket eredményez, sok a hiányzó objektum, úgyhogy a pontos URL ismerete és némi szerencse is kell ahhoz, hogy hiánytalanul megtaláljunk egy régi web-oldalt. Az IA együttműködő partnerei közt van az amerikai nemzeti könyvtár is, és egyik alapító tagja a 2003 nyarán létrejött IIPC-nek (International Internet Preservation Consortium)⁶, melyet a Bibliotheque Nationale de France vezet. A jelenleg 12 tagú IIPC az Internet archiválás módszertanának kidolgozását koordinálja.

A skandináv országok nemzeti könyvtárai 2000 szeptemberében indították a *Nordic Web Archive*⁷ nevű projektjüket, amely 2002 júniusában zárult és a Nordunet2 informatikai programból finanszírozták. Az egyes országok web-tereinek kísérleti jellegű archiválása mellett több nyílt forráskódú eszközt is kifejlesztettek a web-archiválás céljára, melyeket átadtak más országoknak is (pl. Csehország, Észtország). Az NWA az Internet Archive-hoz hasonló 2. típusú technológia: előbb a NEDLIB harvesterrel kísérleteztek, majd áttértek az Internet Archive által is használt Heritrix programra. A begyűjtött objektumok automatikusan készülő metaadatokkal együtt XML fájljokba kerülnek, ezeket indexelik és egy WERA (Web aRchive Access) nevű felületen lehet bennük keresni – URL cím vagy teljes szöveg alapján. Svédországban a Royal Library már 1996-ban foglalkozni kezdett a web-archiválással. Az első próbálkozás 1997-ben történt, ezt 2003-ig 10 alkalommal ismételték meg és így 185 millió fájl (több mint 5.5 terabyte) gyűlt össze. Jelenleg már évi 2-3 alkalommal aratják le a svéd szervereket. 2003 óta nyilvános az archívum, a Wayback Machine-hoz hasonló hozzáférést biztosít. A finn web-tér első archiválása 11.7 millió fájl eredményezett 2002-ben (kb. 500 gigabyte), ezt a szám 2003 októberére 15 millióra nőtt. Az északi országok 2003-ban szintén csatlakoztak az IIPC-hez, “apportként” felajánlva az NWA-hoz kifejlesztett szoftvereket és az elmúlt években szerzett tapasztalataikat.

A National Library of Australia által irányított konzorcium *PANDORA* projektjének (Preserving and Accessing Networked Documentary Resources of Australia)⁸ kezdetei 1996-ig nyúlnak vissza és az 1. típusú archívumok közé tartozik. Válogatott site-okat mentenek le egy jól szervezett, munkamegosztáson alapuló rendszerben, amelyek 15 nagyobb témacsoport szerint böngészhetők és teljes szöveggel is kereshetők. Saját fejlesztésű PANDAS (PANDORA Digital Archiving System) nevű rendszerük a félautomatikus munkafolyamat minden fázisát támogatja: az archiválásra kiválasztott URL-ek nyilvántartása, a lementés időzítése és indítása, a lementett anyag minőségellenőrzése és hibajavítás, metaadatok hozzárendelése, előkészítés a nyilvános szolgáltatásra, a hozzáférési korlátozások, statisztikák és jelentések összeállítása. Az első verzióban 2001 júniusában elkészült PANDAS-t azóta kétszer is továbbfejlesztették, a 3. változat megjelenését ez év elejére ígérik. 2006 januárjában mintegy 11 ezer site vagy dokumentum mentése volt az archívumban (a különböző

⁶ <http://netpreserve.org>

⁷ <http://nwa.nb.no>

⁸ <http://pandora.nla.gov.au>

időpontokban történt ismételt mentéseket is beleszámítva összesen 21.5 ezer tétel), ezek 29 millió fájl jelentettek (kb. 1 terabyte). A válogatás szempontja: ausztrál témájú vagy ausztrál szerzőjű, társadalmi, politikai, kulturális, vallási, tudományos vagy gazdasági témájú, hosszú távon is kutatásra érdemes anyagok, melyeknek az archiválására a készítőjük engedélyt ad. Fontos része a szolgáltatásnak, hogy minden dokumentumot stabil URI-val látnak el, így azok megbízhatóan hivatkozhatók maradnak, akkor is, amikor az eredeti helyükről már eltűntek. Az archivált anyagok katalógustételei a nemzeti bibliográfiai adatbázisba (Kinetica) is belekerülnek. Az ausztrál nemzeti könyvtár szintén az IIPC tagja, és létrehozta egy tematikus honlapot is PADI (Preserving Access to Digital Information)⁹ néven a nemzetközi tapasztalatok összegyűjtése céljából.

Az Egyesült Királyság nemzeti könyvtára az első web-archiválási kísérletet 2001-ben tette, ekkor 100 db. brit történelmi és kulturális site-ot mentettek le, de a gyűjtemény nem lett nyilvános. Az *UK Web Archiving Consortium*¹⁰ 2004 júniusában alakult hat nagy intézmény (köztük a British Library, a National Archives, a Joint Information Systems Committee of the Higher and Further Education Councils) összefogásával. Egy 2 éves projektet indítottak, mely során mintegy 6 ezer web-hely archiválását tervezik (2006 elején ebből kb. ezer már elérhető), a fontosabb tudományos, oktatási, kulturális és közéleti honlapok közül – a szolgáltatók engedélyével. A tagok felosztották egymás közt a begyűjtendő anyagokat, hagyományos gyűjtőkörük és szakértelmük alapján. Az ausztrál PANDAS rendszert vették át és fejlesztik tovább saját igényeiknek megfelelően, letöltő programnak pedig a HTTrack-et használják. 2003-ban a British Library is csatlakozott az IIPC-hez.

A Library of Congress 2000-ben indította a *Minerva* (Mapping the Internet the Electronic Resources Virtual Archive)¹¹ projektet, első lépésben 35 web-site HTTrack-kel való lementésével. Az Internet Archive-val és egyetemi intézményekkel együttműködve a 2000-es elnökválasztási kampány alatt már 200 site-ról csináltak napi mentéseket. Azóta már több mint 35 ezer site-ot mentettek le, általában valamilyen eseményhez kapcsolódva: pl. a szeptember 11-i terrortámadás, a 2002-es téli olimpia, az iraki háború stb. 2002 második félévében 1.3 terabyte-nyi anyagot gyűjtöttek össze. Az archívumnak csak egy része nyilvános, nagy hangsúlyt fektetnek a copyright szabályok betartására. A metaadatok leírása a saját fejlesztésű, MARC-elemeket is tartalmazó, XML-alapú, MODS (Metadata Object Description Schema) segítségével történik, az adatok a könyvtár katalógusába is bekerülnek. A Library of Congress a vezetője a NDIIPP (National Digital Information Infrastructure and Preservation Program)¹² együttműködésnek, amely az Egyesült Államokban folyó digitális archiválási tevékenységet koordinálja. Természetesen tagjai a nemzetközi IIPC-nek is.

A Cseh Nemzeti Könyvtár egy egyetemi partnerrel együttműködve 2000-ben kezdett egy 2 éves pilot projektbe és hozta létre *WebArchive*¹³ nevű szolgáltatását. A teljes nemzeti web-tér begyűjtését célozták meg, a NEDLIB harvesterrel. 2002-ben megismételték a szüretelést és elindult a begyűjtött anyag integrálása a könyvtár online szolgáltatási felületébe valamint a Cseh Nemzeti Bibliográfiába. A projekt melléktermékeként URN szerver, DC- és MD5 checksum generátort is beüzemeltek.

⁹ <http://www.nla.gov.au/padi>

¹⁰ <http://www.webarchive.org.uk>

¹¹ <http://www.loc.gov/minerva>

¹² <http://www.digitalpreservation.gov>

¹³ <http://www.webarchiv.cz>

Javaslatok:

Mint a fenti példákból is látszik, a nemzeti könyvtárak minden országban vezető vagy legalábbis kezdeményező szerepet játszottak az Internet-archiválás elindításában. De az is látszik, hogy szinte mindenhol partnereket kerestek maguknak ehhez a munkához: elsősorban informatikai intézményeket, egyetemi tanszékeket és/vagy cégeket. Nálunk is ez látszik a leginkább járható útnak, ezért kellene egy konzorciumot létrehozni mindazoknak a szervezeteknek, amelyek érdekeltek, érintettek ebben a kérdésben. A technikai feltételek lényegében nálunk is adottak a feladat elvégzéséhez: vannak nagysebességű vonalaink a letöltéshez, a terabyte-os tárolók is egyre elterjedtebbek, van URN szerverünk¹⁴, rövidesen elkészül a magyar DC generátor¹⁵, van országos metaadatgyűjtő rendszerünk¹⁶ és saját fejlesztésű keresőnk¹⁷... A szükséges további szoftverek részben freeware-ként hozzáférhetők, részben megkaphatók az IIPC-től, ha mi is csatlakozunk. Ami hiányzik, az egyrészt az információs és kulturális kormányzati akarat és költségvetési támogatás, másrészt a szükséges jogi környezet – ezeket a konzorciumnak ki kell lobbizni. Kell továbbá egy realisan megvalósítható közös vízió a középtávon elérni kívánt célról vagy célokról.

Mi egy 1. típusú, az ausztrál PANDORA rendszerhez hasonló, válogatás és engedélykérés után egyedi web-helyek mentéseiből álló fájlrendszert javasunk, mert az kisebb fejlesztést és tárhelykapacitást igényel; így viszonylag gyorsan létrehozható lenne egy néhány száz darabos, elsősorban kulturális, oktatási, tudományos és kormányzati honlapokból álló gyűjtemény. A metaadatok részben automatikus, részben pedig emberi közreműködéssel készülnének Dublin Core alapokon, és bekerülhetnének az NDA-ba valamint az OSZK katalógusába is. De emellett kísérleti jelleggel elindulhatna egy 2. típusú, a .hu domainra valamint az ezen kívüli magyar tulajdonú vagy tartalmú szerverekre kiterjedő szűretelés is, ami valószínűleg jogi problémák miatt csak a könyvtárakból lenne hozzáférhető. Akár az egyik, akár a másik változatot választjuk végül, vagy esetleg mindkettőt, mindenképpen a lépcsőzetes, pragmatikus építkezés a célravezető, mert egy maximalista hozzáállással túl sok feltételnek kellene megfelelni és túl sok problémát megoldani, ami valószínűleg zsákutcába vezet.

A projekt hosszú távú fennmaradásához azt is végig kell gondolni, hogy az UNESCO által ránk rótt kötelezettség teljesítésén kívül milyen előnyei vannak egy Internet-archívumnak? Nem csak azért hasznos, mert a a 404-es hibák egy részére megoldást ad, ha van egy másolat az illető web-lapról egy másik szerveren. Egy ilyen archívum új – időbeli – dimenziót ad az amúgy jelenidejű Internetnek. Mivel egységes szerkezetben, stabil szerveren, metaadatokkal és állandó URI azonosítóval ellátva, több időbeli állapotot rögzítve vannak benne a web-oldalak, informatikus szempontból sokkal "jobban viselkedik", mint az eredeti, kaotikus és efemer Internet. Egy ilyen gyűjteményre közhasznú és üzleti célú szolgáltatások sora építhető rá: pl. tematikus összeállítások készíthetők évfordulókra, eseményekhez; részhalmozok képezhetők és azokhoz speciális keresők rendelhetők médiatípus, témakör, célközönség vagy egyéb szempontok alapján; idődimenziót is tartalmazó nyelvi elemzők és egyéb statisztikai programok futathatók rajta; szöveg- és adatbányászati rendszerek, tématerképek építhetők rá; szakirodalmi hivatkozásoknál és linkelésnél jól használható stabil és rövid URN vagy URL címek rendelhetők nemcsak az egyes web-lapokhoz, hanem akár azok minden eleméhez (pl.

¹⁴ <http://nbn.urn.hu>

¹⁵ <http://mek.oszk.hu/dc>

¹⁶ <http://www.nda.hu>

¹⁷ <http://keres.sztaki.hu>

fejezetcímekhez, táblázatokhoz) külön is... Egyszóval egy ilyen gyűjtemény hatalmas értéket képvisel, amelyet jól kihasználva a projekt idővel önfenntartóvá válhat.

A legelső dolog, amit már akár holnap is el kellene kezdenünk: a 90-es évek első felében megszületett magyar online szolgáltatások maradványainak összegyűjtése és egy kis web-múzeum kialakítása belőlük, mielőtt a hazai Internet legizgalmasabb korszakának emlékei végképp eltűnnek a digitális Nirvánában.

Irodalom:

Dr. Magyar Gábor: Internetarchiválás, illeszkedés az NDA-hoz

NDA konferencia, 2004. december 14.

http://www.nda.hu/resource.aspx?ResourceID=magyarg_intenetarchivalas_041214_V1

Moldován István: Archiválás a digitalizáció korszakában

Informatikai és Könyvtári Szövetség, OSZK, Budapest, 2002. szeptember 17.

<http://mek.oszk.hu/html/irattar/eloadas/2002/iksz-oszk.ppt>

Juha Hakala: Archiving the Web: European experiences

Presentation in CONSAL XII, 20-23 October 2003, Brunei, URN:NBN:fi-fe20031951

<http://www.lib.helsinki.fi/tietolinja/0203/webarchive.html>