

„404 Not Found – Ki őrzi meg az internetet?”

Webarchiválás workshop az Országos Széchényi Könyvtárban

Bevezetés

2017. október 13-án első alkalommal került sor kifejezetten a számítógépes világháló archiválásával foglalkozó rendezvényre az *Országos Széchényi Könyvtárban* (OSZK). Az intézmény és a *Kormányzati Informatikai Ügynökség* (KIFÜ) keretei között zajló *Országos Könyvtári Rendszer* (OKR) projekt egyik munkacsoportjaként idén tavasztól kezdhettünk el egy kísérleti projekt keretében foglalkozni a webarchiválással. (Bővebb információt a <http://mekosztaly.oszk.hu/mia> oldalon lehet találni erről.) Célunk az, hogy a projektidőszak végére egy olyan koncepcióval álljunk elő, mely lehetővé teszi, számos európai nemzeti könyvtár mintájára, az üzemszerű munkafolyamatként zajló webarchiválási tevékenység ellátását, illetve szervezését az OSZK részéről. Egy olyan rendszert kívánunk létrehozni, amely a kulturális örökség hosszú távú megőrzésének feladata mellett képes kiszolgálni az oktatás, a tudományos kutatás, az állami szervek, az üzleti szféra és az egyes internethasználók igényeit is. Az archívum megvalósulásával a most csak jelen időben létező magyar internetnek „múltja” is lenne, és olyan lehetőségek nyílnak meg a mai és a jövőbeli felhasználói számára, amelyek jelenleg nem, vagy csak nehezen valósíthatók meg (pl. megszűnt weboldalak megtalálása, webhelyek időbeli változásának elemzése és vizualizálása, stabil hivatkozhatóság, idődimenziót is tartalmazó szöveg- és adatbányászati alkalmazások futtatása, internettörténeti kutatások, hiteles másolatok szolgáltatása). A projekt első fél évét mintegy lezárva került sor rendezvényünkre. A program összeállításakor különös gondot fordítottunk a meglévő külföldi szakmai tapasztalatok, illetve az itthoni előzmények bemutatására. A workshop hangsúlyos céljaként szerepelt továbbá a teljes közgyűteményi szféra (könyvtárak, levéltárak, múzeumok) megszólítása.

Lendvay Miklós az OSZK E-szolgáltatási igazgatója bevezető előadásában tágabb információtudományi, illetve informatikai perspektívában mutatta be a webarchiválás fontosságát, illetve elhelyezte azt az OSZK átformalódó szolgáltatási struktúrájában.

Kees Teszelszky, aki Magyarországon *R. Várkonyi Ágnes* tanítványaként doktorált történészként az *ELTE*-n, most a *Holland Nemzeti Könyvtár* munkatársaként foglalkozik webarchiválással. Bemutatta a holland, illetve a belga webarchiválás jelenlegi helyzetét, szakmai, technológiai háttérét. Szót ejtett e tevékenység és a történetiség kapcsolódási pontjairól, illetve a webarchiválás szerepéről a hamis hírforrások elleni küzdelemben.

Peternák Miklós, a *C3 Kulturális és Kommunikációs Központ Alapítvány* vezetője a webművészetnek mint önálló képzőművészeti ágának az elmúlt 20 évben bekövetkezett fejlődéséről adott nagyívű áttekintést. Mindeközben bemutatta azokat a fő kihívásokat, melyek a webes művészeti alkotások megőrzése, szolgáltatása során vetődnek fel. Rácsodálkozhattunk arra, hogy mekkora küzdelmet kell folytatni a virtuális térben is a szolgáltatások, alkotások megőrzéséért a művészeti, tartalom-szolgáltatói szándékok szerint, vagy épp a rombolási vágy ellenében.

Kampis György egy néhány éves webaratási projekt tapasztalatairól beszélt, melyet egyetemi oktatók és informatikusok közösen végeztek a *Petabyte Kft.* informatikai bázisán, egyetemi TÁMOP projektek felhasználásával az akadémiai szféra webes tartalmainak feltérképezésére, archiválására, valamint híroldalak rendszeres mentésére. Mára már csupán ez utóbbi tevékenység zajlik, az is eléggé takaréklángon. Az elért eredmények, tanulságok viszont remélhetőleg hasznosíthatók lesznek az OSZK kísérleti projektjének keretei között is.

Végezetül pedig mi is lehetőséget kaptunk a bemutatkozásra, az eddig végzett munka áttekintésére Drótos László részéről.

A konferencia prezentációi elérhetők a http://mekosztaly.oszk.hu/mia/404_workshop.html oldalon. Itt a rendezvényről készült fényképanyag, illetve a Videotoriumra felkerülő felvétel linkje is szerepel már. E rövid tanulmány célja egy rövid szubjektív áttekintés felvázolása lehet csupán az adott témák kapcsán. Mindenkinek lehetősége lesz rá, hogy a prezentációk és a videofelvétel segítségével a saját értelmezéseivel egészítse ki az itt felvázolt összképet.

A webarchiválás tágabb összefüggésrendszere

Lendvay Miklós előadásában voltaképpen az internet fejlődésének legújabb fejleményeihez kötötte hozzá a webarchiválás témakörét. Leszögezte, hogy mára a web az egyik legnagyobb publikációs adatforrássá vált, szövegben, hangban, képben hatva minden érzékszervünkre. A hatalmas méretű információrobbanás és az ahhoz kötődő folyamatos technológiai átalakulás közepette illúzió azt gondolni, hogy minden világhálós információforrás tételesen menthető és leírható, ahogyan a hagyományos könyvtárosi szemlélet sugallná. A könyvtáros felelőssége a válogatásban van, illetve az általános pillanatképek készítésében, amiben számot tudunk adni arról, hogy például a web magyar szellete jellemzően hogyan nézett ki egy adott időpillanatban. A helyzetet még bonyolultabbá teszi a szemantikus web fejlődése, ami azzal a Horváth Ádám és Dudás Anikó által többször emlegetett jelszóval ragadható meg, hogy nem a weben kell lenni, hanem webből kell lenni. A szemantikus web egyik fő célja, hogy a webet úgy lehessen kezelni, mint egy adatbázist. Strukturált metaadatok, kontrollált szótárak, s kapcsolt adatok kelljenek (lehetőleg nyílt formában) a szemantikus web megalapozásához. A források egyedi megnevezését pedig szabványos azonosítórendszerrel (pl. URI) lehet biztosítani. Az újfajta szemantikus tartalomszervezési, tartalomfeltárási paradigma egészen másfajta dimenzióba helyezheti a webarchiválást is, elég ha csak arra gondolunk, hogy az egyes honlapok egyedi tartalomfeltárása az azt körülvevő tartalmi környezet figyelembe vételével, a fő kapcsolódási pontok feltáráásával valósulhat majd meg. Ennek a fejlődési útnak szinte az első lépéseit sem tettük még meg. A legújabb *IFLA Reference Model* már magas elméleti szinten összegzi mindazokat a követelményeket, amelyek bibliográfiai adatok logikai struktúráját igyekeznek leírni függetlenül

attól, hogy az adatforrás milyen különféle rendszerekben vagy alkalmazásokban van tárolva. Ezen az általános szinten nem tesz különbséget a hagyományos értelemben vett bibliográfiai, illetve állományrekordok, valamint a besorolási rekordok között, az összes rekordot a bibliográfiai információ fogalma alatt értelmezi. Formálódnak azok a szabványok és eljárásmodok, melyek a MARC alapú katalogizálás helyébe léphetnek. A rekordokat leváltják az entitások, mint adatelemek, illetve a közöttük lévő kapcsolat feltérképezése.

A nemzeti könyvtár szempontjából a webarchiválás szervesen illeszkedik a digitálisan született, illetve digitalizált anyagok gyűjtéséhez, szolgáltatásához, s az ezek hosszú távú megőrzését célzó erőfeszítésekhez. A digitális anyagok feltárását és visszakeresését a formálódó nemzeti névtér is sokban tudja segíteni majd. A könyvtárak jövője tehát számos nyitott kérdéssel áll előttünk, s jórészt rajtunk, szakembereken áll, hogy milyen irányban indulunk el. A gyors módon alakuló környezet pedig azt sugallja, hogy időről-időre úgyszólván át kell tekinteni céljainkat és eszköztárszereket, az a legfontosabb tehát, hogy valamilyen irányban előrelépünk, amely megteremti majd egy esetleges újabb jövőbeni irányváltás alapjait.

A webarchiválás modelljei Hollandiában és Belgiumban

Kees Tszelszky előadásában először áttekintette a holland webarchiválási tevékenységeket, majd rámutatott arra, hogy milyen feladatai, nehézségei vannak e téren a *Holland Nemzeti Könyvtárnak*. Kitért a holland tapasztalatok kapcsán a belgák most formálódó webarchiválási projektjére, illetve rövid összefoglaló áttekintésben helyezte el a két országot az európai szakmai térképen. Előadása utolsó részében arra kereste a választ, hogy a holland gyakorlat milyen tanulságokkal szolgál a webarchivátor szakemberek mellett a világhálón foglalkozó tudósok számára is.

Kees felidézte, hogy a holland webteret már korai időszakában is a gyors fejlődés, illetve az innovativitás, a magas fokú decentralizáltság jellemezte. A közgyűjtemények szempontjából azonban a helyzet távolról sem ideális. A sajátos történeti fejlődés eredményeként nincs kötelesempéldánytörvény sem a papíralapú dokumentumokra, sem a digitális anyagokra. Ennek következtében nem lehetséges általános mentéseket csinálni a holland webtérrel. Csupán egyes témakörökre támaszkodó tematikus mentések vannak a Holland Nemzeti

Könyvtár részéről, azok is csak olyan honlapokról, melyek mentésére tulajdonosaik egyedi engedélyt adtak. Ehhez hasonló, abszurdnak nevezhető jogi helyzet csak Svájcban fordul elő Európán belül. Ennek eredménye azután az, hogy a Holland Nemzeti Könyvtár a hollandiai webtér teljes anyagának csupán 0,14%-át tudja begyűjteni, illetve feldolgozni. A webarchiválási tevékenység elindult ugyan a kilencvenes évek közepén, de azután leállt, szinte minden erőforrást a papíralapú dokumentumok kezelésére csoportosították át, egészen a legutóbbi időkig, amikor újra tudott indulni ez a tevékenység. További probléma, hogy webarchiválást számos egyéb intézmény is végez (pl. a *Rotterdami Városi Levéltár*, a *Holland Nemzeti Levéltár*, kutatóintézetek, a fríz kisebbség kultúráját ápoló szervezet), ezek munkája, a tevékenység szakmai, technológiai keretei nincsenek koordinálva. Az állami szféra webes anyagait a *Nemzeti Levéltár* a Nemzeti Könyvtártól teljesen függetlenül archiválja. A Nemzeti Könyvtárban a webarchiválás kezdete, 2007 óta 13 000 webhelyet, 30 TB terjedelemben arattak le szelektív módon, évente mintegy 1000 honlappal nő a mennyiség. Külön kutatási projekt foglalkozik a tartalmak kiválasztásának, learatásának és tárolásának szakmai kérdéseivel. Az *Új-Zélandi Nemzeti Könyvtárral* együtt a hollandok fejlesztik tovább az archiválás adminisztrációs funkcióit biztosító Web Curator Tool szoftvert. Amellett, hogy másfél ember foglalkozik főállásban a webarchiválással, ez azért nem kis teljesítmény. Korábban egy olyan kutatási programot is indítottak, mely azt vizsgálta, hogy miként lehet esetleg a nyomára akadni már elveszett internetes tartalmaknak. Kees talált olyan nyilvántartásokat, melyeket a Nemzeti Könyvtár készített, s a holland web hajnalán még a nemzeti katalógusban írták le a születő honlapok adatait. Persze az első pár év után már nem bírtak az áradattal, s feladták ezt a tevékenységet. Az Euronet holland internetszolgáltató statisztikai kimutatásait is felhasználták arra, hogy nyomon kövessék a holland webtér szerkezetének alakulását, a felhasználásra, tartalomra vonatkozó tételes és statisztikai adatokat, a honlapokat bejegyzők nevét és címét. Hollandiában az ilyen fajta webes régészet előtt azért számos akadály tornyosul. Sokszor, ha létezik is még az oldal, nehéz megtalálni a bejegyzett tulajdonost, aki engedélyezhetné a tartalom learatását. Egyes esetekben pedig a „digitális demencia” jelensége lép fel, amikor a tartalom tulajdonosa nem akar azonosulni azzal, amit a múltban közzétett, illetve kiadott a honlapján. Sokszor sem a szerző, sem a tartalomtulajdonos nem érdekelt abban, hogy mentsék a tartalmait, nincs jogi eszköz arra, hogy

a beleegyezésük nélkül ezt megtegyék. A webarchiválás további fontos területe a hamis hírek forrásainak vizsgálata, illetve archiválása. Többféle vizsgálatot lehet ennek segítségével végezni a hamis hírszolgáltatás, a propaganda természetére vonatkozóan. Sok európai országban előfordul, hogy a választási kampányok során megjelentetett anyagok egy része hirtelen eltűnik utána, sőt a valaha volt létezését is tagadni próbálják az érintettek. Történelmi szemmel a múlt aktuális hírszolgáltatási trendjeinek elemzése, a rejtett összefüggések kimutatása is elsőrangú fontosságú a híroldalak mentése kapcsán.

Belgiumban idén nyáron indult az a kísérleti projekt, amelynek ott is az üzemszerű webarchiválás megteremtése a célja. A holland hibákból tanulva koordinált nemzeti modellben gondolkodnak, elektronikus kötelempéldány-rendelettel, mely a webarchiválást is magában foglalja, bevonva a tevékenységbe a nemzeti könyvtár mellé a múzeumokat, illetve a levéltárakat is. A belga helyzet sajátossága, hogy a .be domén mellett számos városi, illetve regionális domén is létezik (pl. .brussels, .vlaanderen), ezekre is figyelemmel kell lenni a webarchiválás során. Ráadásul a holland .nl domén nevek 2,5%-a is belga tulajdonos nevére van bejegyezve, átfedést eredményezve így a két ország webes tartománya közt. Általában is az jellemző, hogy a flandriai vidékeken majdnem olyan sűrű a bejegyzett internetes honlapok aránya, mint a szomszédos Hollandiában, ellentétben Vallóniával, vagy a németlakta vidékekkel, ahol sokkal alacsonyabb ez az arány.

Az előadás fő tanulsága tehát, hogy a webarchiválás szakmai és jogi keretei nagyban függenek a jogi környezettől, a helyi közgyűjteményi tradícióktól. Ötévente felül kell vizsgálni a gyűjteményszervezési gyakorlat alapelveit, s azokat amennyire lehet, visszamenőleg is érvényesíteni (ez a webes régészet területe). A helyi webarchiválási gyakorlatot, eredményeket és tapasztalatokat mindenképpen helyes nemzeti keretek között koordináltan összegezni az érintettek bevonásával.

A médiaművészet és a webarchiválás múzeumi kontextusa. A webarchiválás tartalmi hitelességének dilemmái

Peternák Miklós, a C3 központ igazgatója a magyar webművészeti, webmúzeumi tartalmak 20 éves történetére tekintett vissza, erős nemzetközi kitekintéssel. Példákat villantott fel arról, hogy az eltűnt tartalom jelzéseként szolgáló 404-es hiba-

üzeneteket is lehet művészeti installációként felhasználni, a mondanivalót érdekesen tálalva, bár napjainkban már ez is egyre inkább uniformizálódik. Beszél arról is, hogy a tartalmak sokféle módon eltűnhetnek a webről. Feladhatja, kivonulhat a tartalomtulajdonos, megszűnhet a tartalomszolgáltató. Az is gond, hogy a hosszú ideje üzemelő oldalak életútjuk során általában számos formai s ezzel összefüggő tartalmi, szerkezeti megújulás esnek át. A külső linkek, amelyek rájuk mutatnak, ezt nem mindig tudják lekövetni. Arról nem is beszélve, hogy az aktuálisnál korábbi állapotot szervezett webarchiválás hiányában senki sem menti el, az elvész a benne foglalt információforrásokkal együtt. Nem lehet tanulmányozni az adott honlap, webhely fejlődését, alakulását. Így következhet el az örök jelen káprázatában élő korszak után a digitális középkor. Még az a szerencse, hogy az *Internet Archive* szolgáltatásának segítségével számos magyar közgyűjteményi és webművészeti oldal pályafutását figyelemmel lehet kísérni.

A fentebb felsoroltak mellett különleges problémát jelent a technológiai elavulás. Még ha a tartalom nem is változik, egy idő után a mögötte álló webtechnológiai környezet már olyan elavulttá válhat, hogy a megjelenítéshez szükséges eszközöket (böngészőket) is emulálni kell, hogy abban a környezetben lehessen megjeleníteni őket, ahogyan eredetileg megalkották. Művészetelméleti szempontból jelentős kérdéseket vet fel az, hogy elég-e ragaszkodni a művészi mondanivalóhoz, az installáció művészi hitelességéhez, az eredeti webes megjelenítés aprólékos reprodukciója helyett? Művésze és szolgáltatási helyzete válogatja, hogy mikor lehet a technikai, illetve dizájn háttérrel hozzáalakítani a jelenkori követelményekhez, s mikor marad a technikai környezet emulációja mint egyetlen eszköz az adott tartalom hiteles megjelenítéséhez. A csak digitálisan létező alkotások megőrzésénél tehát a megjelenítési környezet megőrzésének is központi szerepe van, mely a fizikai valójában létező műalkotások esetében szinte alig bír jelentőséggel. Ez komoly mértékben összetettebbé teszi a webarchiválás tevékenységét, különösen a hosszú távú megőrzés szempontjai szerint. S előbb-utóbb szinte minden típusú weboldalnál megjelenik a jövőben a megjelenítési környezet emulálásának szükségessége, mert olyan dinamikusan fejlődik a tartalomszolgáltatási eszközrendszer. Peternák előadásában olyan bravúros nemzetközi vállalkozásokra is kitért, amikor a tulajdonosa, *Reinhold Greith* által halálra ítélt újmédia művészeti weblap- és linkgyűjteményt sikerült külső másolatok segítségével újra életre kelteni más

domén alatt, döntően az Internet Archive mentésének segítségével. Nyilván ez nem lesz pontos replikája az eredeti honlap bármilyen időpillanatban létezett verziójának, mert az Internet Archive az adott honlap egyes elemeit is hajlamos más-más időpontban lementeni, majd összetenni a megjelenítéskor, de művészetileg mégiscsak egy megőrzött, hiteles másolatot sikerült közzétenni így.

E tanulmány szerzője ehhez még hozzáteszi, hogy különféle szempontok csapnak itt össze. Levéltári, illetve könyvtári szempontból fontos az, hogy időbelileg is hiteles legyen a mentés (egy adott mentést egy adott időpillanathoz köthessünk). Másrészt viszont ez a gyakorlatban sokszor megvalósíthatatlan, mert például gyorsan és dinamikusan változik a teljes tartalom, vagy annak egy szelete a mentési művelet közben. A múzeumi megközelítés szerint viszont az az elsődleges, hogy tartalmilag, illetve művészileg hiteles másolat készüljön el, az másodlagos, hogy a mentett honlap egyes elemeit pontosan mikor mentették el. Egyre inkább úgy tűnik, hogy el kell fogadnunk alapaxiómaként, hogy a lementett weblap nem lehet azonos minden tartalmi, megjelenítési részletében az egy adott időpillanatban létező élő honlappal.

Az első magyar internetarchívum-kísérlet az egyetemi szférában

Kampis György egy érdekes kísérleti projektet mutatott be, mely egyfajta közvetlen előzményeként is felfogható az OSZK jelen webarchiválási erőfeszítéseinek. A *PetaByte Nonprofit Kutatási Kft.* 2013-tól kezdődően a MIA nemzeti archívum kialakításának feltételeit egy pilotprojekt keretében kezdte el vizsgálni, ennek érdekében felvették a kapcsolatot a MIA eredeti kezdeményezőivel és az akkori NIIF-fel, számos fejlesztést és előkísérletet végeztek, egyeztetéseket folytattak. Műszaki oldalról megpróbálták finomhangolni, s amennyire lehetett, kissé továbbfejleszteni az aratásokat végző Heritrix szoftvert. A tesztalanyok az NIIF HBONE hálózat akkori kb. 500 tagjának honlapjai voltak. A begyűjtött adatokat longitudinális módon elemezték. Mintegy 17 terabájtnyi anyag begyűjtése után további tárhely hiányában a pilotprojektet lényegében le kellett állítani 2015–2016-ban. Csak a kapcsolódó részprojekt, a híroldalak tartalmának bizonyos mértékű mentése élt tovább 2016 után is. A begyűjtött anyag annak ellenére nőtt ilyen nagyra, hogy az ISO és ZIP formátumokban tömörített fájlokat, a hangalapú médiaállományokat, illetve a Microsoft Office xls, xlsx fájlformátumaiban levő

tartalmakat kizárták a mentésből. A videofájlokat viszont learatták. Az egyetemi, akadémiai oldalakról a Heritrix segítségével végzett mentéseket kiegészítették a wget szoftver segítségével készült híroldalmentések (a híroldalak száma 139-től 4965-ig bővült dinamikusan). A felsőoktatási intézmények oldalairól 53 GB tartalmat gyűjtöttek be, ebből 36 GB volt a szöveges tartalom. Az MTA kutatóintézetek oldalairól 33 GB adatot arattak le, ebből 5 GB volt a szöveges anyag.

A híroldalakról gyűjtött anyagot különféle tartalom-elemzésekre is felhasználták, például hogy mely szavakkal írták le leginkább a Brexittel kapcsolatos eseményeket a magyar hírportálok. Az eredményeket szófelhőben, illetve szógyakoriságot jelző grafikonokon prezentálták.

A projektnek rejtett haszna is lett, mert ráébredt ez a nem hivatalos archivátor közösség arra, hogy milyen hihetetlenül sokszínű kihívások rejlenek a webes tartalmak learatásával kapcsolatban, mekkora adattartalmat kell tárolni, milyen aprólékosan kell szoftvereket beállítani, milyen tartalmakat lehetséges learatni, s melyeket nem. Nem győzték már kézzel tisztítani a begyűjtött anyagokat a szükségtelen, illetve a tartalom szolgáltatását akadályozó részletektől. A tárhely betelése, a megszaladó erőforrásigények mellett ezek a nehézségek a pilot lezárását segítették elő. Az aratások ugyan leálltak, de az adatokat megőrzik. Jelenleg annak lehetőségét vizsgáljuk, hogy az elért eredményeket miképpen lehetne hasznosítani az OSZK most futó pilotprojektje számára.

Kampis György is rámutatott a projekt tanulságaként arra, hogy a megfelelő jogi háttér megléte mennyire elengedhetetlen a webarchiválási tevékenységhez. A műszaki problémák jelentős része a jövőben megoldódhatna olyan ajánlások közzétételével, melyek az archiválandó tartalmak technikai jellemzőit alakítanák archivumbaráttá. Az előadó hangsúlyozta az adattípusok előzetes kizárásának fontosságát is, mely egyfajta komoly szűrőként is felfogható még az aratás elkezdése előtt. Végezetül megállapította, hogy meglátása szerint a longitudinális letöltéseknek illetve elemzéseknek is van létjogosultsága. A vonatkozó publikációk adatait az előadás prezentációja tartalmazza.

Az OSZK webarchívum-kísérleti projekt első fél éve

A pilotprojekt alapcélkitűzéseit már a bevezetőben felsoroltuk. Első lépésként kialakítjuk a koncepciót,

az ütemtervet, a forrásigényt, a gyűjtőkör ideiglenes lehatárolását a pilotprojekt igényei szerint. Jelenleg a témafelelős, Drótos László mellett egy főállású webkönyvtáros, egy adminisztrátor és egy informatikus alkotja a stábot, s fel szeretnénk venni még egy rendszergazdát is egyes informatikai feladatok elvégzésének céljából. Alapvetően nyílt forráskódú szoftverekkel dolgozunk. Több szoftvert is tesztelünk folyamatosan az általános webaratósi feladatok mellett, kifejezetten a közösségi hálózatok tartalmainak mentésére (pl. Webrecorder). Számos tesztaratóst végeztünk, könyvtárak, levéltárak, múzeumok, egyetemek, az OSZK saját felületeinek, illetve e-periodikák honlapjainak mentésével. Emellett nagyobb léptékű terheléses tesztekre is sor került a szükséges hardver infrastruktúra-követelményeinek meghatározására. Az eredmények kiértékelése folyamatosan zajlik. Meghatározzuk a típushibákat, külön listába gyűjtjük a közel tökéletes mentéseket. Elkezdődött egyes intézményektől a hivatalos engedélykérés, hogy honlapjaik tartalmát folyamatosan learathassuk és szolgáltatathassuk. Tervezzük a jól sikerült mentésekből egy demófelület kialakítását, ahol a nagyközönség is meg tud majd győződni a webarchiválás hasznosságáról és a szolgáltatási lehetőségekről.

Az ismeretszerzés is nagyon fontos eleme lett munkánknak. A döntően angol nyelvű szakirodalom áttekintése, folyamatosan frissülő bibliográfia összeállítása, szabványok és szakkönyvek beszerzése a *Könyvtártudományi Szakkönyvtár* segítségével, illetve az *Aarhusi Egyetem* internetkutató csoportja által szervezett Phd-szemináriumon való részvétel jellemzik ezt a tevékenységformát.

Ismeretterjesztési törekvéseink körébe illeszkedik a projekt ideiglenes honlapjának létrehozása, levelezőlista az érdeklődő szakemberek részére, wiki publikálása a honlapon, belső előadás a *Könyvtári Intézet* munkatársai részére, különféle külső szakmai rendezvényeken előadások tartása (pl. MKE Vándorgyűlés, CogInfoCom konferencia). Tanulmányainkat és szakirodalmi szemléinket a hazai könyvtári szaksajtóban, illetve a digitális bölcsészet körébe tartozó publikációs fórumokon tesszük közzé.

Nagyon fontos eleme a projektnek a hazai és nemzetközi kapcsolatépítés. Itthon közgyűjteményekkel, egyetemi tanszékekkel vettünk fel már kapcsolatot. Hallgatóként részt vettünk az idei IFLA konferencia webarchiválással foglalkozó szekciójában, értékes kapcsolatokat alakítottunk ki

az osztrák, szlovén, szlovák, holland, belga, dán webarchívumok munkatársaival. Arra törekszünk, hogy a külföldi tapasztalatokat beépítsük a munkánkba, időt és energiát spórolva ezzel a projekt fejlesztése során. Sikeres csatlakozási kérelmeket nyújtottunk be a webarchiválással foglalkozó intézményeket tömörítő IIPC konzorciumban való részvételre, mely a tapasztalatcsere további lehetőségeit nyitja meg.

A jövőben a már említett demófelület kialakítása mellett újabb szoftvereket szeretnénk kipróbálni. Szeretnénk megtervezni a hardverkonfigurációt, s elindítani a beszerzést. A szelektív webarchiválások (tematikus, eseményalapú mentések) metaadatkezelésébe partnerintézményeket és a felhasználókat is bevonnánk. Doménszintű aratást is tervezünk a magyar webtér tartalmáról évente kétszer (a .hu domén alatt több mint 700 ezer bejegyzett név van jelenleg). A begyűjtött tartalom összetételéről statisztikákat is fogunk készíteni.

Következtetések

A webarchiválás alapvető szakmai kérdéseivel foglalkozó önálló szakmai rendezvényre Magyarországon korábban még nem került sor. Reményeink szerint ez a mostani workshop magában rejti a sikeres folytatás lehetőségét. A webarchiválással foglalkozó szakmai diskurzusnak minél szélesebb közgyűjteményi körre kellene kiterjednie. A kísérleti projekt előrehaladásával, majd remélhetőleg az üzemszerű webarchiválás elindulásával összhangban a nemzeti könyvtár fontos feladata, hogy az e területen zajló kommunikációnak workshopok, szakmai napok formájában teret adjon a jövőben is.

Németh Márton

(Webkönyvtáros, OSZK,
E-könyvtári Szolgáltatások Osztálya)