

Drótos László <drotos.laszlo@oszk.hu>
Németh Márton <nemeth.marton@oszk.hu>
(Országos Széchényi Könyvtár)

A webarchiválás oktatása

Előadásunkban három témakört érintünk: Elsőként beszámolunk a dániai Aarhusi Egyetemen működő NetLab kutatócsoport által webarchiválással foglalkozó szakembereknek és PhD hallgatóknak szervezett online szemináriumról és annak tanulságairól. Az előadás második része az internet archiválásával foglalkozó szervezeteket tömörítő konzorcium, az IIPC újonnan alakult oktatási és képzési munkacsoportjának tevékenységéről nyújt áttekintést. Bemutatjuk annak a kérdőíves felmérésnek az eredményeit, melyet a meglévő tapasztalatok és jó gyakorlatok, ill. a jövőre vonatkozó tervek összegyűjtése céljából végeztek. A harmadik részben pedig a Könyvtári Intézet képzési keretei között a közgyűjteményi szakemberek számára meghirdetésre kerülő, a webarchiválás módszertanát oktató tanfolyam terveit és tematikáját ismertetjük.

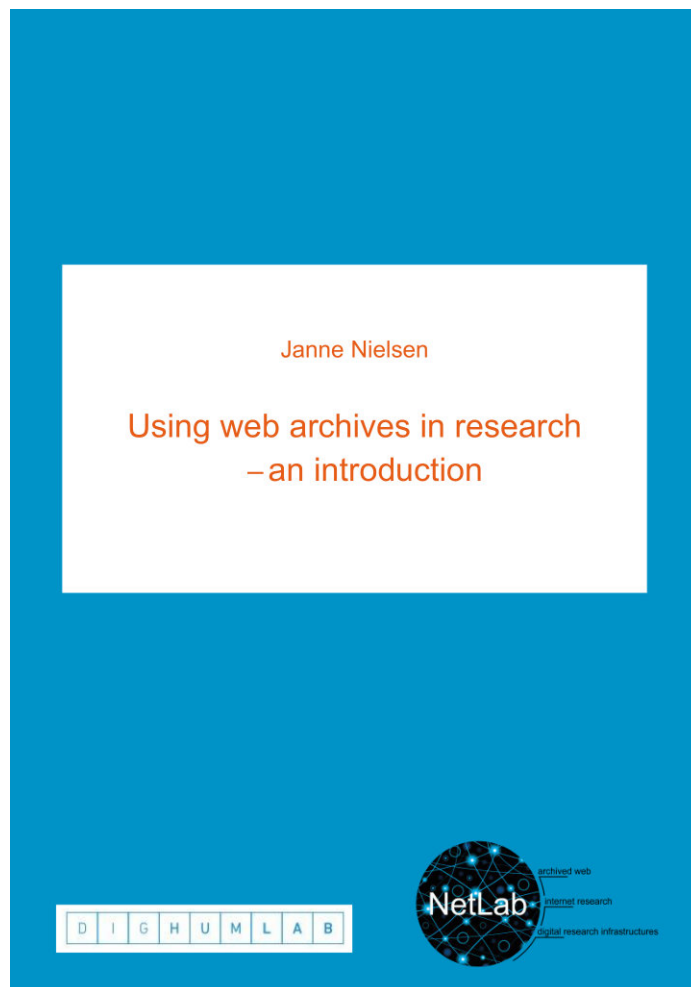
1. A NetLab online kurzusa

A dániai Aarhus egyetemén az országos DIGHUMLAB kutatási infrastruktúra részeként működő, Niels Brügger professzor vezette NetLab kutatócsoport mintegy két éve nyújt online egyetemi kurzusokat¹ PhD hallgatók, ill. a témában érintett szakemberek, kutatók számára a webarchiválás témakörében. 2017 őszén első alkalommal angol nyelven, nemzetközi keretek között is meghirdették a kurzust², ráadásul a részvétel ingyenes volt. Külön növelte az értékét, hogy az OSZK kísérleti projektjéhez hasonlóan nemrég indult belga webarchívum munkatársaival is alkalmunk nyílt megismerkedni, a dán szakemberekkel zajló közös munka mellett. Az online szeminárium egyéni felhasználó név/jelszó párossal védett, Moodle-alapú távoktatási felületen zajlott. Itt publikálták a szervezők a segédanyagait, a feladatok szövegeit, és itt volt lehetőség a válaszok megfogalmazásra és az egymással történő társalgásra is a kurzusfórum keretében. Ez a felület csak a tanfolyam ideje alatt volt elérhető, viszont a kurzus végeztével lehetőségünk volt az összegyűlt anyag lementésére. Az online szeminárium öt fő területre volt felosztva. A kurzus mellé adott, Jan Nielsen által írt kézikönyv³ szolgált a segítségünkre az egyes témák kidolgozása során. (1. ábra)

¹ Bővebben a NETLAB konzorcium oktatási tevékenységeiről:
<http://www.netlab.dk/services/courses/>

² A kurzus ismertetője:
<http://www.netlab.dk/wp-content/uploads/2017/04/NetLab-Web-Archiving-Course-Brochure.pdf>

³ Ingyenesen elérhető itt:
http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf



1. ábra. A kötet címlapja

Először annak felmérése zajlott, hogy a résztvevőknek milyen kapcsolatuk van a webarchiválással és milyen elvárásaik fogalmazódnak meg a kurzussal szemben? Már ebből is következtetni lehetett arra, hogy a konstruktivista pedagógiai alapon szerveződő szeminárium minden egyes félévben más és más profillal bír, a tartalmát erőteljesen befolyásolja a résztvevők érdeklődési köre, szakmai irányultsága. A második téma körül zajló társalgás éppen arra irányult, hogy az általános érdeklődési kör megfogalmazása után mindenki elmondja, hogy miért hasznos számára webarchiválással foglalkozni és felvázoljon valamilyen konkrét munkatervet vagy kutatási tervet ennek kapcsán. A harmadik témakörben keresni kellett három, az egyéni érdeklődési területhez tartozó honlapot, melyek legalább egy éve üzemelnek már, s áttekintést kellett adni ezek archiválásának kihívásairól. Mintaként lehetett használni az Internet Archive⁴ vagy bármelyik nemzeti webarchívum nyilvánosan elérhető mentéseit. A negyedik témakörben meg kellett fogalmazni a kutatási körbe vágó további honlapok összeválogatásának stratégiáját. Ezt követően ki kellett választani a megfelelő szoftveres hátteret, kísérleti aratásokat végezni, majd mini riportban összegezni a tapasztalatokat. Számunkra ez bizonyult a leghasznosabb pontnak, hiszen a belgák is és mi is elő tudtunk állni konkrét tartalmak kísérleti mentéseinek tapasztalataival, s meg tudtuk beszélni egymással az aratáshoz használt egyes szoftverek előnyeit, hátrányait, illetve értékeltük a végeredményt. A társalgásba több ponton bekapcsolódtak a dán nemzeti

⁴ <http://www.archive.org>

webarchívum⁵ munkatársai is, akik saját tapasztalataikkal színesítették a kommunikációt. Az volt a fő cél, hogy mindenki választ kapjon a kérdéseire a konkrét kihívások kapcsán, s ezeket hasznosítva tudja tovább építeni a saját projektjét. Mindezek után pedig az ötödik témakör keretében összesítettük a kurzussal kapcsolatos tapasztalatainkat, s ezzel zárult a szeminárium, melynek elvégzéséről a résztvevők bizonyítványt is kaptak. (2. ábra)

Tanulásként mindenképpen megállapítható, hogy a webarchiválás esetében szerencsés ez a fajta oktatásszervezés, ahol az elméleti háttérként szolgáló ismereteket írásos formában bocsátják rendelkezésre, maga a szeminárium pedig erre az anyagra (is) támaszkodva, konkrét gyakorlati célkitűzések mentén zajlik. Különös figyelmet fordítottak arra, hogy miként lehet a napi munkában, az adott kutatási témakörben hatékonyan hasznosítani a tapasztalatokat. A legtöbbet egymástól tanultunk. A tervezés lépései, a szoftveres problémák, az aratási hibák számbavétele, a konkrét nehézségekkel történő megbirkózás állt a munka középpontjában, Így a kurzus igen jól segítette az itthoni webarchiválási kísérleti projekt szakmai megalapozását.



2. ábra. A kurzus elvégzésével szerzett bizonyítvány

2. Az IIPC Training Working Group

Az IIPC az online tartalmak megőrzésével foglalkozó szervezeteket, intézményeket tömörítő nemzetközi konzorcium⁶. Elsődleges feladata a webarchiváláshoz kötődő technológiák, módszerek, szabványok fejlesztése, az egyes országok jó gyakorlatainak megismertetése, a nemzetközi együttműködés támogatása, valamint a webarchívumokban tárolt tartalomhoz

⁵ <http://www.netarkivet.dk>

⁶ Webportáljuk címe:
<http://www.netpreserve.org>

történő hozzáférés segítése, minél változatosabb célú felhasználásnak előmozdítása. A konzorciumon belül 2017 végén alakult meg az oktatási és képzési munkacsoport⁷. Első projektjük keretében összeállítottak egy kérdőívet⁸. Arra voltak kíváncsiak, hogy az egyes országokban milyen típusú szervezetek, mekkora létszámban foglalkoznak webarchiválással, valamint ezeknek a szakembereknek milyen igényeik vannak az oktatás, a szakmai továbbképzés terén. A kérdőívet 2018 januárjában lehetett kitölteni. E tanulmány írásának időpontjában az elsődleges gyors összegzés készült el⁹.

A kérdőívet 224-en töltötték ki, öt kontinensről. A válaszokból kitűnik, hogy döntően egyetemi, kutatóintézeti, illetve kisebb, de azért jelentős részben nemzeti könyvtári keretek között végeznek webarchiválási tevékenységet. Jelentősnek nevezhető még a levéltári szféra súlya, de jelen vannak a múzeumok, az audiovizuális archívumok, illetve a piaci szereplők is. A webarchiválással foglalkozók létszámát firtató kérdésre adott válaszokból kiderült, hogy az intézmények mintegy felében kevesebb mint 1 teljes státuszú munkatárs foglalkozik e területtel! A válaszadók mintegy negyede 1 és 3 között adta meg a létszámot. A másik végletet, a legalább 10 főt alkalmazókat az intézmények kilenc százaléka jelenti, a többi eloszlik a minimum 3 ill. 5 munkatársat foglalkoztatók között. A harmadik kérdés a webarchiválás során végzett tevékenységek felsorolására irányult. A válaszokból kiderült, hogy legtöbbször az archiválendő anyagok válogatásával, illetve szabályzat-alkotással foglalkoznak. Majdnem azonos a súlya a metaadatszerkesztési, a minőségbiztosítási, a kommunikációs és a konkrét webaratásokat lebonyolító tevékenységeknek, a legkevesebben pedig a szoftverfejlesztők vannak. A kitöltők között többségben voltak a közgyűjteményi háttérű szakemberek az informatikusok rovására. A legtöbbször arról adtak számot, hogy csak nemrég kezdtek munkahelyükön webarchiválással foglalkozni, illetve hogy jelentős kihívást jelent beilleszteni ezt a területet az általános szolgáltatási palettába. Jó néhányan utaltak arra is, hogy egyelőre csak terveznek ezzel a szakterülettel foglalkozni, de konkrét gyakorlati tevékenységet még nem végeznek.

Ebből is kiderül, hogy még nagyon az út elején tartunk, ami az oktatási, képzési tevékenységek súlyát felértékeli. A legtöbbször jelenleg online forrásokra támaszkodnak, ha szakmai kompetenciáik bővítésére vágnak, s ettől jelentősen elmarad bármilyen szervezett oktatási tevékenység hozzáférhetősége, illetve relatív módon jelentős a súlya a semmiféle képzésben részesült munkatársaknak is. Messze a legkevesebben vesznek részt hivatalos tantervi alapú képzésekben. Az oktatási tevékenység, ha elérhető, akkor workshopok keretében, illetve informális keretek között vagy valamely szakmai szervezet által koordinálva zajlik. A képzési igényeket felmérő kérdésre a legtöbbször azt válaszolták, hogy a webarchiválással, a digitális megőrzéssel kapcsolatos szabványok és technológiák terén szeretnék tudásukat elmélyíteni, ill. az egyes szoftvereszközök használatának oktatását tartanák fontosnak. A képzési formák közül pedig a legnépszerűbbnek a webináriumok, illetve a személyes jelenlétre épülő tanfolyamok bizonyultak.

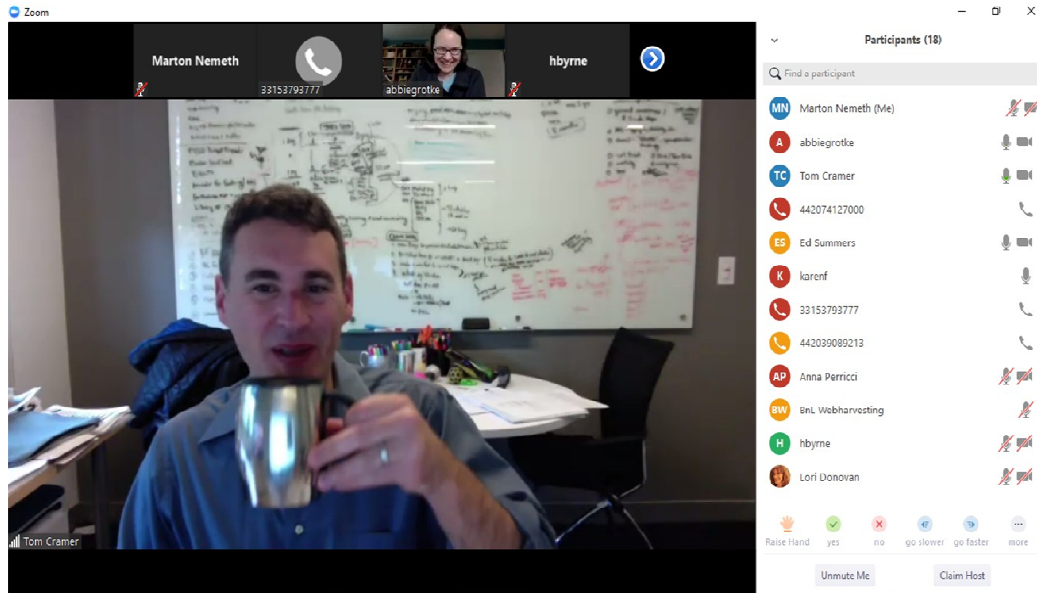
Az IIPC oktatási és képzési munkacsoportja a kérdőív tapasztalatait felhasználva tervezi jelenleg azokat a fajta oktatási tevékenységeket, melyekkel a leghatékonyabb segítséget tudják majd nyújtani tagintézményeik, illetve a lehető legszélesebb körű szakmai célcsoportok

⁷ A Training Working Group tevékenységének ismertetése:
<http://netpreserve.org/about-us/working-groups/training-working-group/>

⁸ A kérdőív elérhető innen:
<https://www.surveymonkey.com/r/V7MVXXW>

⁹ Az összefoglaló alapjául nem nyilvános IIPC munkaanyagok szolgálnak, ezért e publikációban hivatkozható összefoglaló még nem áll rendelkezésre.

sámára. (3. ábra) Fokozottan szeretnék építeni az egyes országokban eddig felhalmozott tapasztalatokra. Most először mérik fel azt is, hogy az egyes országokban milyen oktatási és képzési tevékenységek zajlanak, milyen keretek között, s hányan vesznek azokban részt. Remélhetőleg a közeljövőben olyan oktatási programok és szakmai anyagok kidolgozása kezdődik el, amelyekből a magyar szakmai közönség is sokat profitálhat majd.



3. ábra. Közös munka a virtuális térben az IIPC Training Working Group keretében

3. Hazai tervek

Az Országos Széchényi Könyvtárban zajló, az országos könyvtári rendszert (OKR) megújító program keretében 2017 elején indult kísérleti webarchiválási projektben fogalmazódott meg az igény arra, hogy közgyűjteményi szakemberek számára a Könyvtári Intézetben 30 óra terjedelmű továbbképzést szervezzünk. A webarchiválás ugyanis akkor működik hatékonyan, ha széleskörű intézményi együttműködésen alapul. A képzés fő céljaként ezért az merült fel, hogy a résztvevők ismerjék meg az internetes tartalmak hosszú távú megőrzésére alkalmas technikákat, az ezzel a területtel foglalkozó külföldi projekteket és az OSZK-ban folyó ez irányú tevékenységet. A kurzus elvégzése után pedig legyenek képesek akár magán célra, akár intézményük számára weboldalatokat és webhelyeket tartalmazó gyűjteményeket létrehozni, illetve részt venni a magyar internet archívummal kapcsolatos munkákban. A fő célközönséget a könyvtárosok, a múzeumi és levéltári dolgozók, az archivátorok és az informatikusok jelentik. A képzés oktatói bázisát az OSZK E-könyvtári ill. Infrastruktúra Szolgáltatások Osztályainak a kísérleti projektben érintett munkatársai adják. (4. ábra)

A tananyag az alábbi fő témakörökből áll:

- Az internetes tartalmak archiválásával kapcsolatos szakkifejezések, fogalmak és modellek elméleti ismerete.
- Néhány Windows alatt is használható archiváló, ill. archiválást segítő szoftver, valamint online szolgáltatás működésének alapszintű elsajátítása.
- Egy Linux-alapú webarchívum főbb komponenseinek és működésének felhasználó szintű ismerete.

- Az archiválendő webes tartalmak válogatásával, a mentések utólagos ellenőrzésével és azok metaadatokkal való ellátásával kapcsolatos fontosabb tudnivalók elsajátítása.

- A webarchívumokban megőrzött tartalom különböző (elsősorban tudományos) célú hasznosításával és az archívum fenntarthatóságával kapcsolatos ismeretek.

A finanszírozási feltételek megteremtése és az akkreditációs folyamat lefolytatása után reményeink szerint 2018 őszén már el tudjuk indítani a kurzust. Emellett – a már említett OKR program részeként megvalósuló e-learning rendszert felhasználva – egy „blended” típusú, vagyis részben távoktatással megvalósuló, részben pedig személyes jelenlétet is igénylő tananyag fejlesztése is elkezdődött, amely remélhetőleg szintén még az idén elérhető lesz az internetes tartalmak hosszú távú megőrzése iránt érdeklődők számára.

**„Az internet archiválása mint közgyűjteményi feladat” c.
továbbképzési program oktatói beosztása:**

Oktatók:

Drótos László (DL) <drotos.laszlo@oszk.hu> (E-könyvtári Szolgáltatások Osztály)
Kovács Péter (KP) <kpeter@oszk.hu> (Infrastruktúra Szolgáltatások Osztály)
Moldován István (MI) <moldovan@oszk.hu> (E-könyvtári Szolgáltatások Osztály)
Németh Márton (NM) <nemeth.marton@oszk.hu> (E-könyvtári Szolgáltatások Osztály)
Visky Ákos László (VÁL) <visky.akos.laszlo@oszk.hu> (E-könyvtári Szolgáltatások Osztály)

Időbeosztás: (az óraszámok 50 perces időtartamokat jelentenek!)

1. nap

- 1.1. Bevezető: Miért fontos a digitálisan születő, az interneten terjedő kultúra megőrzése, mi a közgyűjtemények felelőssége és mit tesz a nemzeti könyvtár? A kötelempéldány szabályozás ezen a téren hazánkban és külföldön. (MI 3 óra)
- 1.2. Áttekintés: Archiválási módszerek és archívumfajták. (NM 1 óra)
- 1.3. Külföldi projektek: Az Internet Archive, néhány nemzeti webarchívum, valamint egyéb típusú archívum ismertetése és kipróbálása. Az IIPC bemutatása (NM 4 óra)

2. nap

- 2.1. Internetes tartalmak mentésére és megőrzésére használható ingyenes Windows szoftverek bemutatása és kipróbálása (DL 5 óra)
- 2.2. Weboldal- illetve webhely-archiváló online szolgáltatások bemutatása és az ingyenesek kipróbálása (DL 3 óra)

3. nap

- 3.1. Webarchívum kialakítása Linux szerveren: a Heritrix arató-, az Open Wayback megjelenítő- és a NutchWAX kereső-rendszerek ismertetése és működésük demonstrálása, a WARC/ARC tároló- és a CDX indexformátum rövid bemutatása (KP 2 óra)
- 3.2 A Web Curator Tool keretrendszer bemutatása és kipróbálása, az archiválásra kiválasztott webhelyek metaadatolása (VÁL 3 óra)
- 3.3 A Netarchive Suite keretrendszer bemutatása (VÁL 1 óra)
- 3.4 Válogatási szempontok és utólagos minőségellenőrzés, archiválhatóság, jó és rossz példák bemutatása (VÁL 2 óra)

4. nap

4. ábra. Részlet az előzetes kurzustervből

Zárszó

Előadásunkban beszámoltunk egy webarchiválással foglalkozó dániai továbbképzési program felépítéséről és tanulságairól. Továbbá összefoglaltuk az IIPC Training Working Group első néhány hónapnyi munkáját, a webarchívumok működésének körülményeit és az azokkal foglalkozó szakemberek képzési igényeit felmérő nemzetközi kutatás eredményeit. Végül, de nem utolsósorban itthoni oktatási terveinkről adtunk számot. Nagyon fontosnak tartjuk, hogy a külföldi tapasztalatok hasznosításával egy olyan együttműködési hálózatot alakítsunk ki Magyarországon, melynek révén a webarchiválás üzemszerűen és hatékonyan végezhető. Ennek fontos előfeltétele a folyamatos szakmai fejlődés, a továbbképzés kereteinek szervezett módon történő biztosítása.

Bibliográfia

Az IIPC Training Survey kitöltésére szóló felhívás, Hozzáférés: 2018.05.22,
<https://netpreserveblog.wordpress.com/2017/12/14/iipc-training-survey/>

Az IIPC Training kérdőív, Hozzáférés: 2018.05.22,
<https://www.surveymonkey.com/r/V7MVXXW>

Az IIPC Training Working Group oldala, Hozzáférés: 2018.05.22,
<http://netpreserve.org/about-us/working-groups/training-working-group/>

NIELSEN Janne, *Using the Web archives in Research (a dán NetLab webarchiválás kurzus elméleti segédanyaga)*, Hozzáférés: 2018.05.22,
http://netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf

A dán webarchívum honlapja, Hozzáférés: 2018.05.22,
<http://www.netarkivet.dk>

A NetLab kurzusának ismertető brosúrája, Hozzáférés: 2018.05.22,
<http://netlab.dk/wp-content/uploads/2017/04/NetLab-Web-Archiving-Course-Brochure.pdf>

A NetLab kurzus ismertető oldala, Hozzáférés: 2018.05.22,
<http://netlab.dk/services/courses/>