

Németh Márton

Webarchiválás két szakmai rendezvény tükrében

Bevezető

Írásomban két rendezvényt tekintek át: 2019. március 29-én az Osztrák Nemzeti Könyvtár (ÖNB) rendezett egy félnapos workshopot a webarchiválás kapcsán, június elején pedig a webarchiválással foglalkozó intézmények éves konferenciájára került sor Zágrábban¹. Előbbi esetben rövid áttekintést adok az elhangzottakról, illetve az utóbbiban utalok arra is, miként tudunk itthon a nemzetközi kapcsolatainkból profitálni.

Nemzeti webarchiválási modellek bemutatása, gyakorlati kihívások összegzése Bécsben

A bécsi workshopot az ÖNB főigazgatója nyitotta meg. Ezt követte egy bevezető előadás a web társadalmi hatásáról, a webarchivátorok felelősségéről a megőrzés kapcsán. A kommunikációs szakember szerint a tömegek bölcsességére lehet számítani a közösségi hálózatokon megjelenő tartalmak archiválásánál a válogatás kapcsán. Itt jelenleg még a megfelelő technikai eszköz megtalálása a nagyobb kihívás, amellyel be is lehet gyűjteni ezeket a tartalmakat. A szerzői jogi témákra áttérve, többen is hangot adtak annak a szégyenletesnek tartott ténynek, hogy a webarchívumhoz való szabad hozzáférés nem került be az EU szerzői jogi irányelvébe mint tisztességes felhasználási (fair use) gyakorlat.

A harmadik előadásban a svájci webarchívum került röviden bemutatásra. A webarchiválás 33 intézmény közreműködésével zajlik a nemzeti könyvtár koordinálásával, illetve annak technikai platformján. Egységes űrlapon lehet archiválásra javasolni tartalmakat az egyes intézmények felé, amit bárki megtehet. A válogatás dilemmái az általános trendeket követik: milyen mélységben, milyen reprezentativitással arassák a .ch domain anyagát. A szelektív archiválásnak két fő szempontja van: a tudományos fontosság, illetve a mindennapi kultúra webes lenyomatának archiválása. Általában teljes webhelyeket

igyekeznek archiválni, emellett törekednek a legalább minimális szintű metaadat-leírás mellékelésére is. A hivatalos állami közigazgatási honlapok mentéseihez való hozzáférés korlátlan, egyébként szerzői jogi okok miatt leginkább csak a webarchiválásban részt vevő intézményeken belül lehetséges. A svájci nemzeti könyvtár 33 partnerkönyvtárral együtt alakította ki a gyűjtőköri alapelveit. A szövetségi weblapok, illetve a kantonok hivatalos honlapjainak gyűjtése a 14 ezer helyhatósági honlappal együtt alapvető kötelesség. Ráadásul általában gyakran frissülő webhelyekről van szó. Ezt egészítik ki a különleges tematikai gyűjtemények, gazdasági társadalmi, irodalmi témában. De például az emmentáli sajtó, illetve a svájci csokoládémárkák honlapjainak mentése is kulturális jelentőséggel bír. A tematikus honlapok válogatásában, gyűjtésében igen fontos a helyi tapasztalat, s a webarchiválást segítő társadalmi infrastruktúra megalapozása. Az úgynevezett fair use típusú hozzáférési szabályokat amerikai mintára alakítják ki. A svájci domén mintegy 1%-át gyűjtik tematikusan. A honlaptulajdonosok kérhetik tartalmaik kizárását a tematikus gyűjtésekből. A technikai infrastruktúra döntően a nemzeti könyvtárra támaszkodik. Jogilag nehéz osztott együttműködést létrehozni technikai téren. Viszont a tartalom gyűjtése terén eredményesen működik a konzorcium tematikus, illetve földrajzi alapon elosztott tartalomgyűjtési tevékenysége.

Szó esett még arról is, hogy a német nyelvű, illetve speciálisan osztrák irodalmi élet webes lenyomatait az Innsbrucki Egyetemen próbálják gyűjteni, ideértve könyves blogok és közösségi oldalak anyagait is. Eredetileg az amerikai Internet Archive infrastruktúrájára támaszkodtak, de ma már önálló szerverek és munkatársi gárda áll a rendelkezésükre. A svájci mellett bemutatták a dán webarchívum tevékenységét is. Dániában alapvetően két könyvtár végzi ezt a feladatot, amelyeket nemrégiben vontak össze egymással: az aarhusi Állami és Egyetemi Könyvtár, illetve a dán nemzeti könyvtár. Megállapodást kötöttek a nagy médiaszolgáltatókkal, sajtókonzernekkel a weben megjelenő anyagaik beszolgáltatásáról. A dán kötelespéldány-törvény elő is írja, hogy ha a könyvtár technikai okok miatt nem tud egy adott gyűjtőkörbe eső honlapot archiválni, akkor a honlaptulajdonosnak kell beszolgáltatnia a tartalmat. Dániában a legfőbb kihívást a szerzői jogi problémák jelentik a webarchívumban tárolt anyagok hozzáférése terén. Sikerült kialakítani egy olyan gyakorlatot, hogy a célhoz kötöttség alapelveinek szigorú betartásával kutatók felhasználhassanak, illetve meghatározott időre ki is emelhessenek, el is vihessenek kutatási célra adatkészleteket a webarchívumból. Minden más célra azonban csak külön engedéllyel, könyvtáron belül érhetőek el anyagok. Ausztriában – Dániához képest – jóval szegényesebb a szerverháttér. Évente kétszer gyűjtik a .at domén anyagát, de a mentési folyamat hónapokig tart. A tematikus gyűjtésekre azért is van nagy szükség minden nemzeti webarchívumban, mert például az osztrák webarchívumban tematikusan gyűjtött honlapok 47%-a már el sem érhető az weben. Tapasztalataik szerint a politikai, közéleti tartalmak különösen veszélyeztetettek e téren. De a kulturális tartalmak bizonyos része is dinamikusan változik.

Az esemény jó alkalom volt arra, hogy betekintést nyerjünk egyes nemzeti, illetve intézményi szintű speciális webarchiválási modellek működésébe, másrészt pedig izelítőt kaptunk számos olyan gyakorlati kihívásról, amelyek a mindennapi munka terén keletkeznek.

Mindenki együtt: IIPC-konferencia Zágábban

A webarchiválásban érdekelt intézményeket tömörítő konzorcium, az International Internet Preservation Consortium (IIPC) idei közgyűlését és konferenciáját június 5. és 7. között Zágábban rendezték meg. Ez az egyetlen olyan rendezvény, ahol személyesen is lehetőség nyílik a szakmai tapasztalatcserére, az új szoftverekkel és szolgáltatásokkal történő megismerkedésre, illetve a már jól bevált szolgáltatások fejlesztési terveinek áttekintésére. Az OSZK 2018 óta tagja a konzorciumnak, részt veszünk a tréningmunkacsoport, illetve a nemrég alakult kutatási munkacsoport tevékenységében is. Bár tavaly Új-Zélandon csak egy üdvözlő videóval képviseltettük magunkat, idén a személyes részvételre is lehetőség nyílt. Alkalmunk volt öt percben bemutatni a saját projektünket, illetve húsz percben ismertetni a teljes szövegű visszakereséssel, valamint a metaadat-rögzítéssel kapcsolatos eredményeinket és terveinket.

Ezen a területen a dán nemzeti könyvtár munkatársaival dolgozunk együtt a SolrWayback nevű keresőprogram fejlesztésén és tesztelésén, ami mindkét fél számára kölcsönösen előnyös. A dánok ugyanis – a már fentebb tárgyaltak szerint – szerzői jogi okok miatt nem adhatnak nyilvános hozzáférést a webarchívumuk egyetlen részéhez sem. Így abban a minigyűjteményben tudják tesztelni, kipróbálni, nyilvánosan bemutatni a keresőt, amit a tartalomtulajdonosok engedélyével lementett és nyilvánosan szolgáltatható magyar webhelyekből kialakítottunk.

Aktívan közreműködünk az IIPC tagjai számára formálódó oktatási tevékenységek összehangolásában a tréningmunkacsoport keretein belül. A sok videókonferencia és levelezés után a rendezvényen végre személyesen is tudtunk egyeztetni ebben a témakörben. A konferenciát követően a Digital Preservation Coalition szervezett partnerségben elindult a tartalomfejlesztés, egy webarchiválásról szóló kezdő kurzus megalapozásához. Az oktatási anyagokat az egyes tagok szabadon felhasználhatják saját oktatási szolgáltatásaik bővítéséhez. Továbbá kisebb, csoportos kerekasztal-beszélgetések keretében vitattuk meg, hogy miképpen támogathatná hatékonyabban a konzorcium az egyes tagok webarchiválási törekvéseit. Megújításra kerülnek a stratégiai alapelvek, hogy milyen általános szakmai prioritások mentén működjön a jövőben ez a nemzetközi szervezet.

Kísérleti projektünk keretében 2017 óta részt veszünk a Web Curator Tool (WCT) nevű, eredetileg új-zélandi és brit, jelenleg pedig új-zélandi és holland fejlesztésű programcsomag kipróbálásában, a fejlesztési javaslatok megfogalmazásában. A szoftver segítségével egységes felületen lehet menedzselni a teljes webarchiválási munkafolyamatot, illetve rögzíteni a metaadatokat. Zágábban egy külön workshop keretében vitattuk meg a WCT-t használó könyvtárak munkatársaival a további igényeinket, ötleteinket. A konferencián mindenki megkapta a programcsomag legfrissebb verzióját, tesztelés céljából. A rendezvényt követően pedig több átfogó javaslatcsomagot is eljuttattunk a fejlesztőkhöz. Nagyon hálásak voltak érte, mert így a fejlesztőmunka során nem csupán a munkában részt vevő könyvtárak igényei kerülnek kielégítésre, hanem egy szélesebb kör is fel tudja vázolni a kívánalmait.

Érdekes volt szembesülni azzal a helyzettel, hogy az USA-ban a webarchiválással kapcsolatos technikai fejlesztések terén az Internet Archive szinte monopolhelyzetben van. Az Archive-It nevű, igény szerinti archiválási szolgáltatásokra fizetnek elő a legnagyobb amerikai egyetemek és könyvtárak is. Ezen intézmények számára tehát a webarchiválás pusztán adatkurátori feladatokra korlátozódik: a válogatásra, majd a begyűjtött tartalom

feldolgozására és szolgáltatására. Ezzel szemben a világ egyéb részein, döntően a nagyobb közgyűjteményekben, párhuzamosan zajlik az aratási és tartalomkezelési tevékenység.

Zárszó

Általános és sajnálatos trend, hogy mindenhol, ahol webarchiválással foglalkoznak, igen szűkösen állnak rendelkezésre a személyi és az anyagi erőforrások. Másfelől viszont igen összetartó, lelkes, innovatív szakmai közösségnek lehetünk részesei, a közös munkák eredményei pedig, mint néhány korábban vázolt példában láhattuk, közvetlenül visszahatnak a saját projektünkre is.

Jegyzetek

1. A rendezvényről készült videofelvételek az IIPC YouTube-csatornáján nézhetők vissza: https://www.youtube.com/channel/UCkUsw2Lo1ahekgy_xEb11BA/ (2019.06.10.)

Az általam készített zágrábi előadások prezentációi pedig innen érhetők el: http://mekosztaly.oszk.hu/mia/doc/Zagrab_2019/ (2019.06.10.)



*Michaela Mayr, az Osztrák Webarchívum munkatársa
előadás közben (Fotó: Osztrák Nemzeti Könyvtár)*