

Ki őrzi meg a helyi webet?

Helyismereti vonatkozású internetes tartalmak archiválása webaratással

Moldován István, az OSZK E-könyvtári Szolgáltatások Osztályának vezetője
Drótos László, az OSZK kísérleti webarchiválási projektjének témafelelőse

A digitális kultúra megőrzése

Majd tizenöt éve, 2003 októberében, az UNESCO egy chartát adott ki a digitális örökség védelméről. A charta röviden, érvekkel alátámasztva hívta fel a világ kormányainak, az országok közgyűjteményeinek és kiadóinak, valamint a releváns iparágaknak a figyelmét arra, hogy:

- A digitális örökség ugyanolyan fontos és értékes tartalom lehet, mint a hagyományos információhordozókon lévő kulturális örökség.
- A digitális örökség veszélyeztetettebb, mint a hagyományos hordozókon lévő tartalom. A digitális eszközök, adathordozók avulása sokkal gyorsabb, ezért a rajtuk lévő információk gyorsabban tűnhetnek és tűnnek is el a világ számára, mint minden korábbi esetben.
- A digitális örökség megmentése és a hozzáférés biztosítása érdekében mielőbb cselekedni kell minden felelős résztvevőnek. Cselekvési terveket kell kidolgozni és megoldani a digitális örökség megőrzését a jövőnek.

A charta tehát kihangsúlyozta, nyomatékosította, hogy a digitális örökség nem másodlagos az emberi kultúrát őrző és továbbító egyéb információhordozók mellett. Ami kiemelten fontos, hogy ez az örökség természetéből következően különösen veszélyeztetett. A ókori ékírások, a középkori pergamenek egy része, a papírra nyomtatott írások többsége régóta megtalálható a közgyűjteményekben. Ezzel szemben a digitális adathordozók néhány évtized alatt is elveszítethetik a tartalmukat, visszaolvashatatlaná, megfejthetetlenné válnak¹. Ezek nagyságrenddel gyorsabban avulnak az eddigieknél, például a különböző méretű floppy lemezek villámgyorsan mentek ki a divatból a pendrive-ok, később pedig az egyre nagyobb kapacitású memóriakártyák megjelenésével. De a fizikai hordozók mellett az információt megjelenítő szoftverkörnyezet elavulása is egyre nagyobb problémát jelent.

A magyar könyvtárak gyűjtőköre

Érdekes megvizsgálni ebből a szempontból a hazai könyvtárak gyűjtőkörét. Ha megnézzük közelebbről néhány egyetemi, szak- vagy megyei könyvtárunk gyűjtőköri leírását, ott a tartalmi lehatárolás mellett főként a papíralapú dokumentumtípusokat találjuk. Ezek mellett még jellemzően az offline elektronikus hordozók jelentek meg (CD-ROM-ok, DVD-k), de a nem fix tárolóeszközhöz kötött online dokumentumokat nemigen említik. Még leginkább a felsőoktatási könyvtárak által működtetett repozitóriumokban található digitálisan született tartalmat: tudományos publikációkat, szak- és PhD dolgozatokat, egyéb intézményi kiadványokat. A könyvtárak egyre nagyobb számban létrejövő digitális gyűjteményei főként a nyomtatott állomány digitalizált változatait tartalmazzák. A digitálisan született és online terjesztett tartalom gyűjtésére, megőrzésére még kevés helyen találunk példát.

¹ Az elpusztíthatatlan adatok legendájáról, 2011.12.06
https://cydonia.blog.hu/2011/12/06/az_elpusztithatlan_adatok_legendajarol

Az OSZK alapító okiratában² (az új változat 2016-os!) elsőként, legfontosabb tevékenységként a Magyarországon keletkezett sajtótermékek gyűjtése van megnevezve, a kötelezpéldányra alapozva. A 2015-ben megújult gyűjtőköri szabályzat³ 2.4-es pontja a „nemzeti kulturális örökség részét képező dokumentumok digitalizálását és a digitalizált változatok gyűjtését” írja elő feladatul. Azonban egy 30/2014-es EMMI rendelet alapján a digitalizált változatok gyűjtése is csak a könyvtárak által digitalizált dokumentumok körére vonatkozik. Ennek végrehajtására évek óta készül az OSZK, a jelenlegi tervek alapján a közeljövőben létrejövő OKR rendszer fogja majd megoldani ezt a feladatot is. A 10.12-es pont nevezi meg a Magyar Elektronikus Könyvtár Osztály feladatait, ebben már az OSZK nemzeti könyvtári gyűjtőkörébe tartozó távoli hozzáférésű tartalomról is szó esik. Itt a statikus dokumentumok mellett az online források automatikus, webaratással való gyarapítása is szerepel, konkrétan honlapok gyűjtése.

A kötelezpéldányok gyűjtéséről azonban még mindig a 60/1998. (III. 27.) Korm. rendelet van hatályban, amely természetesen igencsak szűkszavúan utal az elektronikus formában is megjelent sajtótermékekre. A jogszabály aktualizálása már több mint 10 éve elkezdődött. A társadalmi vitára bocsátott 2013-as jogszabálytervezet⁴ a Könyvtári Intézet honlapján⁵ is megjelent, igaz, már csak célzott kereséssel található meg. Egy későbbi tervezetet már a parlament is tárgyalta, sőt megvitatták⁶. Azóta ez a tervezet is módosult, egy tavalyi változatát 2017 tavaszán az MKE is véleményezte⁷, de ez már nem kapott olyan nagy visszhangot, mint a 2013-as tervezet. A legújabb hírek szerint a megújult kormányzat a közeljövőben most már tényleg tervezi elfogadni és életbe léptetni az új szabályozást, amely már az online dokumentumokra is kiterjed. Tanulság azonban, hogy ha a jogi szabályozásra várnak a könyvtárak és a könyvtárosok, akkor sok év elmehet mellettünk, és sok dokumentum tűnhet el a digitális térben, ezért előbb kell cselekednünk.

A digitális tartalomhoz viszonyulást jól jelzi, hogy az EMMI 2017-ben kidolgozta és meghirdette a Közgyűjteményi Digitalizálási Stratégiát⁸ (KDS), amelynek célja, hogy „biztosított legyen a nemzeti kulturális kincseink, a közgyűjteményi tartalmak minél szélesebb körű, akadálytalan hozzáférése a kultúrafogyasztók számára”. Ennek alapján született meg 2017 őszén az OSZK 2017-2025 közötti digitalizálási stratégiája és még készül az OSZK által az Országos Könyvtári Digitalizálási Stratégia. A KDS mindösszesen egy oldalon, egy pontban utal a digitális tartalmakra: „2. A magyar digitális örökség megőrzésének egy rendezetlen területe: Webaratás”, amelyben leírják az online tartalmak, értsd honlapok begyűjtésének technikáját, utalnak néhány nemzetközi példára, de az eddigi hazai törekvésekre sajnos egy szóval sem. A digitalizálás feltétlen prioritása mellett a kormányzati, ágazati stratégiákban az eredeti digitális tartalmakra szinte utalás sem történik.

Digitális megőrzési programok külföldön

Érdeemes megnézni egy-két példát a digitális megőrzésre külföldön. Az USA-ban 2000 decemberében a Kongresszus támogatásával indult a Nemzeti Digitális Infrastruktúra és Megőrzési Program (NDIIPP)⁹. A program főként a digitálisan létrejövő fontos tartalmak gyűjtésére, megőrzésére és hozzáférhetővé tételére kezdett el kidolgozni egy nemzeti stratégiát.

² http://www.oszk.hu/sites/default/files/alapito_okirat.pdf

³ <http://www.oszk.hu/szabalyzatok>

⁴ <http://2010-2014.kormany.hu/download/9/bd/c0000/K%C3%B6teles%20p%C3%A9ld%C3%A1ny.zip>

⁵ <https://ki.oszk.hu/hir/konyvtari-intezet/kotelespeldanyaszolgaltatas-rendellettervezete>

⁶ <https://goo.gl/9EkDWt>

⁷ <http://mke.info.hu/blog/2017/05/a-kiadvanyok-kotelespeldanyainak-szolgaltatasarol-megorzeserol-es-hasznalatarol-szolo-eloterjesztes-velemenyezese-2017-05-23/>

⁸ <https://digitalisjoletprogram.hu/hu/tartalom/kds-kozgyujtemenyi-digitalizalasi-strategia>

⁹ <http://digitalpreservation.gov>

Önmagában azonban még az USA nemzeti könyvtára is kevés egy ilyen mérvű feladathoz, ezért több száz partnerrel több országból dolgozik ezen a programon.

Németországban a NESTOR¹⁰, egy, a német nemzeti könyvtár által vezetett intézményi szövetség foglalkozik a digitális források hosszú távú megőrzésével. A könyvtárak mellett múzeumok, intézetek is részt vesznek a 2003-ban indult projektben, amely többek között kiterjed a formátumokra, a kutatási adatokra, az emulációra, a személyes digitális megőrzésre és számos egyéb területre is.

Digitális megőrzés az OSZK-ban

Az OSZK-ban a digitális megőrzéssel 1999 óta foglalkozunk, amióta a Magyar Elektronikus Könyvtár a nemzeti könyvtárba integrálódott. Digitalizálni később és külön pályázatok mentén kezdett el az OSZK (pl. Corvina projekt¹¹, Képidő¹²), és jelenleg is a könyvtár állományának a digitalizálását – főként megrendelésre – a Digitalizáló Osztály végzi. A MEK Osztály, jelenlegi nevén az E-könyvtári Szolgáltatások Osztálya kezdte el (azaz folytatta) az eredeti digitális dokumentumok gyűjtését. Alapvetően elkülönült, önálló tartalmi egységekkel foglalkozunk. A MEK¹³ könyv jellegű digitális dokumentumokat gyűjt és szolgáltat. 2004-től az Elektronikus Periodika Adatbázis és Archívum¹⁴ (röviden EPA) keretében már az elektronikus időszaki kiadványokat is elkezdtek gyűjteni. 2007-től pedig a képeket is megőrizték a Digitális Képarchívum¹⁵ (DKA) nevű gyűjteményben. A kiadóktól egyre nagyobb számban beérkező, kereskedelmi forgalomba kerülő e-könyveket pedig az OSZK Digitális Könyvtára¹⁶ (OSZKDK) fogadja 2008-tól.

A fentiek közös jellemzője, hogy lépésről lépésre, egyre többféle típust bevonva, önálló, elkülönült dokumentumok szintjén kezdődött el a digitális archiválás és megőrzés. Az interneten azonban az 1990-es évek közepétől a webhelyek váltak az egyik, ha nem a legfontosabb információhordozóvá. Ezek már nem egy-egy lehatárolt dokumentumot jelentenek, hanem egy összetett hipertext rendszert alkotnak számos, különböző formátumú és típusú digitális dokumentumból, vagyis szöveget, képet, hangot és videót is tartalmazhatnak. A honlapok élettartama egyes becslések szerint két év körül van, miközben az első magyar honlapot¹⁷ 1994-ben hozták létre a BME-n. A fent említett gyűjtemények megtervezése és elindítása mellett 2006-ban egy Networkshop előadásban fogalmazódott meg először nyilvánosan egy Magyar Internet Archívum (MIA) létrehozásának a gondolata és javaslata¹⁸. A javaslatot elméletben többen is támogatták, az OSZK és a kulturális minisztérium vezetése is, de plusz támogatás nélkül a megvalósítás javasolt lépéseit nemigen lehetett elkezdni. Az első óvatos próbálkozások 2009-ben kezdődtek el az OSZK-ban a webaratás technológiájának megismerésével és tesztelésével, amelyről Kovács Péter (aki a jelenlegi projektnek is az egyik informatikusa) a 2010-es Networkshop konferencián be is számolt „Web archiválás automatikus módszerekkel” című előadásában¹⁹. A későbbi években a Szegedi Tudományegyetem Könyvtára és az NIIF is sikertelenül vették be egy-egy pályázatukba a hazai webarchiválás elindítását, mert sajnos egyik pályázat sem nyert.

¹⁰ <http://www.langzeitarchivierung.de>

¹¹ <http://www.corvina.oszk.hu>

¹² <http://www.kepido.oszk.hu>

¹³ <http://mek.oszk.hu>

¹⁴ <http://epa.oszk.hu>

¹⁵ <http://dka.oszk.hu>

¹⁶ <http://oszkdk.oszk.hu>

¹⁷ http://wow.iit.bme.hu/hungary/homepage_h.html

¹⁸ Drótos László: Mi a MIA? – Javaslát egy magyar internet-archívum létrehozására

<http://mek.oszk.hu/html/irattar/eloadas/2006/mia.htm>

¹⁹ <http://kifu.videotorium.hu/hu/recordings/679>

Végül az OSZK átfogó informatikai fejlesztését biztosító Országos Könyvtári Rendszer (OKR) programban nyílt lehetőség arra, hogy egy alprojekt keretében elkezdődjön végre a magyar web teljes és szelektív archiválása²⁰. Az egész projektet a Kormányzati Informatikai Fejlesztési Ügynökség (KIFÜ) és az OSZK közös konzorciumban valósítja meg. A web-archiválási alprojekt célja, hogy:

- kidolgozzuk a webarchiválás technológiai munkafolyamatait, teszteljük a szóba jöhető szoftvereket, felmérjük egy folyamatos, üzemszerű webarchiváláshoz szükséges infrastruktúra igényeit;
- kidolgozzuk a folyamatos archiválás munkafolyamatait, beleértve az ehhez a feladathoz szükséges tudás megszerzését, rögzítését és továbbadását dokumentumok, oktatások, előadások keretében.

A munka²¹ 2017 tavaszán kezdődött el és a projekt ütemezése alapján az első szakasz 2018 december 31-ig tart. Sikerült a tesztfeladatokhoz szükséges szervereket és tárhelyeket üzembe állítani. És ami elengedhetetlenül szükséges volt: sikerült két-három munkatársat is felvenni, akik csak a webarchiválással foglalkoznak.

A kísérleti webarchiválási projekt

Mint az előzőekből kiderült, bár az egyedi dokumentumok digitális megőrzését viszonylag korán el tudtuk kezdeni, a webhelyek archiválása esetében tíz-húsz éves lemaradást kellene behoznunk más országokhoz képest. A hátrányunk egyetlen előnye, hogy hasznosíthatjuk azt a tudást, amit a külföldi kollégák már felhalmoztak, és használhatjuk a mások által kifejlesztett – nagyrészt ingyenes – szoftvereket és szolgáltatásokat. Emiatt is volt az egyik első fontos cél, hogy az OSZK is belépjen az internet-archiválással foglalkozó intézményeket tömörítő szervezet, az International Internet Preservation Consortium²² tagjai közé, ami az idén meg is történt. Az elmúlt kb. másfél évet néhány informatikus és könyvtári kollégával együtt tanulással, teszteléssel, alternatív megoldások kipróbálásával, és a majdani – remélhetőleg – üzemszerű működéshez szükséges munkafolyamatok egyes elemeinek kidolgozásával töltöttük.

Lementettünk például több mint 5 ezer magyar webhelyet egy-két alkalommal, több-kevesebb sikerrel és nem teljes mélységben. Csináltunk tematikus aratásokat könyvtárak, levéltárak, múzeumok, egyetemek, kutatóintézetek, önkormányzatok honlapjairól, és esemény-alapú gyűjtéseket az idei téli olimpiával, ill. országgyűlési választásokkal kapcsolatos híreket, információkat közlő weboldalokról. A honlapok, blogok és elektronikus periodikák közül kiválogattunk néhányat, amelyek mentései jobban sikerültek és ezek tulajdonosaitól – köztük a győri megyei, ill. egyetemi könyvtártól (*1. és 2. ábra*) és az egyházmegyei levéltártól – engedélyt kértünk arra, hogy a lementett mementókat nyilvánosan is szolgáltatassuk az OSZK szerveréről, egyelőre 2018 végéig. Szerencsére sokan reagáltak pozitívan a kérésre, így 2018 februárjától már bárki számára megnézhető ez a több mint száz webhelyből álló kis gyűjtemény²³ az interneten, sőt azóta bővült is némileg, mert időközben mi is csináltunk egy űrlapot²⁴, amin keresztül már érkezett néhány javaslat, felajánlás. Az archivált tartalomban levő szövegek és képek közötti kereséshez beüzemeltünk két – egy saját és egy dán fejlesztésű – programot (*3. ábra*), valamint elkezdtük a webhelyek leírását, vagyis részletes bibliográfiai, adminisztrációs és technikai metaadatokkal való ellátását (*4. ábra*). Hogy

²⁰ <http://www.oszk.hu/okr-webarchivalas>

²¹ <http://mekosztaly.oszk.hu/mia/>

²² <http://mekosztaly.oszk.hu/mediawiki/index.php/IIPC>

²³ <http://mekosztaly.oszk.hu/mia/demo>

²⁴ <https://goo.gl/forms/Y1qIxcM7APPi443>

hogyan érdemes egy webarchívumot „katalogizálni”, arra vonatkozóan az amerikai könyvtári szövetség, az OCLC 2018 februárjában tett közzé egy ajánlást²⁵, ezt figyelembe véve alakítottuk ki a saját adatstruktúránkat.

Mindehhez olyan szoftvereket használunk, mint az Internet Archive által kifejlesztett Heritrix²⁶ „aratógép” és az OpenWayback²⁷ megjelenítő, az ezek kezelését és a munkafázisok nyilvántartását megkönnyítő új-zélandi Web Curator Tool²⁸, a dán NetarchiveSuite²⁹ és az amerikai WAIL³⁰, a már említett, szintén dán SolrWayback³¹ kereső és megjelenítő, valamint a Nimbus Screen Capture³² nevű böngészőkiegészítő, amivel képeket lehet készíteni weboldalakról. A metaadatokat pedig egyelőre az XML Notepad³³ segítségével rögzítjük, így még keresni nem lehet köztük, csak megnézni őket HTML vagy XML formában.

A „csinálva tanulás” mellett „tanulva tanítással” is foglalkozunk, abban bízva, hogy Magyarországon is lesznek olyanok, akik elkezdene érdeklődni ez iránt a nálunk még újnak számító könyvtári és levéltári tevékenység iránt, és akár saját gyűjteményeket hoznak létre, akár pedig bekapcsolódnak a nemzeti webarchívum munkálataiba. Ezért igyekszünk mielőbb megosztani másokkal is a megszerzett tapasztalatokat a projekt honlapján³⁴, ahol egy wikit³⁵ is írunk az internetes tartalmak megőrzésével kapcsolatos fogalmakról, projektekről, szoftverekről, szolgáltatásokról stb.; továbbá a MIA-L levelezőcsoportban, melynek jelenleg 28 tagja van; valamint cikkek³⁶ és előadások³⁷ formájában is. Mindezek mellett egy nagyrészt e-learning jellegű és egy hagyományos továbbképzési tanfolyamot is indítunk a Könyvtári Intézet szervezésében.

Helyi vonatkozású webtartalmak megőrzése külföldön

A webarchívumokról, vagyis az internetes tartalmak hosszú távú megőrzéséről szóló szakirodalomban – az Internet Archive³⁸ nevű amerikai non-profit szervezet 1996 óta épített hatalmas gyűjteménye mellett – leginkább a nemzeti és szakkönyvtárak projektjeit emlegetik. Hogy a közkönyvtárak és közlevéltárak gyűjtőkörébe és gyarapítási tevékenységébe ez a terület nem nagyon épült bele, annak valószínűleg a szükséges informatikai infrastruktúra és szakértelem hiánya lehet a magyarázata, bár ennek a mértékét sokan túlbecsülik, hiszen egy böngészőkiegészítő vagy egy ingyenes Windows program telepítésével egy átlagos PC-n is egészen jó helyi archívumot lehetne csinálni már régóta. És ha belegondolunk: az interneten digitálisan születő kultúra nagy része inkább a nagyközönséget szolgáló memóriaintézmények gyűjtőkörébe tartozik, mert a jelentősebb kulturális, tudományos és közéleti dokumentumok többsége még mindig – bár egyre csökkenő arányban – megjelenik valamilyen hagyományos hordozón is, így valószínűleg bekerül legalább a nemzeti könyvtárba és néhány szakterületi közgyűjteménybe. Viszont nem kerül be és ezért nem őrződik meg még középtávon sem az a rengeteg információforrás, amelyeket a társadalom különböző szegmensei: a hétköznapi

²⁵ <https://www.oclc.org/research/themes/research-collections/wam.html>

²⁶ <http://mekosztaly.oszk.hu/mediawiki/index.php/Heritrix>

²⁷ <http://mekosztaly.oszk.hu/mediawiki/index.php/Wayback>

²⁸ <http://mekosztaly.oszk.hu/mediawiki/index.php/WCT>

²⁹ <http://mekosztaly.oszk.hu/mediawiki/index.php/NetarchiveSuite>

³⁰ <http://mekosztaly.oszk.hu/mediawiki/index.php/WAIL>

³¹ <http://mekosztaly.oszk.hu/mediawiki/index.php/SolrWayback>

³² http://mekosztaly.oszk.hu/mediawiki/index.php/Nimbus_Screen_Capture

³³ <http://www.lovetsoftware.com/downloads/xmlnotepad/readme.htm>

³⁴ <http://mekosztaly.oszk.hu/mia>

³⁵ <http://mekosztaly.oszk.hu/miawiki>

³⁶ http://mekosztaly.oszk.hu/mediawiki/index.php/Magyar_nyelv%C5%B1_szakirodalom

³⁷ http://mekosztaly.oszk.hu/mia/404_workshop.html

³⁸ http://mekosztaly.oszk.hu/mediawiki/index.php/Internet_Archive

emberek, a helyi közösségek, a különféle szervezetek és intézmények tesznek fel az internetre, elsősorban jelen idejű felhasználásra, de amelyek rendkívül érdekesek és hasznosak lehetnének akár csak néhány év, évtized múlva is, ha nem tűntek volna el végleg a gyorsan távoluló digitális univerzumban is folyamatosan aktív, anyagelnyelő fekete lyukakban.

Hogy a széles közönségnek szolgáltató könyvtárak között is akadnak azért olyanok, amelyek korán felismerték a „ki őrzi meg a helyi webet?” kérdésben a saját felelősségüket, arra a legjobb példa az ausztrál PANDORA³⁹ archívum, melyet a web hajnalán, még 1996-ban indítottak. Bár ezt is egy nemzeti könyvtár kezdeményezte, de amint összeállt a rendszer és a munkafolyamat, 1998 augusztusában a State Library of Victoria máris csatlakozott a projekthez és példáját később a többi állami könyvtár és néhány egyéb intézmény, archívum is követte.

A kevés helyi kezdeményezés közül a külföldi publikációk legtöbbször Čačak webarchívumát⁴⁰ említik, mint követendő jó példát. Čačak egy 72 ezer lakosú város Közép-Szerbiában, melynek a könyvtárában kb. 10 évvel ezelőtt egy komolyabb digitalizálási és digitális gyűjteményfejlesztési programot indítottak. Ennek részeként gondoltak arra is, hogy a város és környékének kulturális ill. helytörténeti témájú webhelyeit archiválni kellene, ami azért is volt akkor különösen aktuális, mivel 2008-ban elkezdődött a szerb webszerverek átállítása a korábbi .yu-ról az .rs doménre és ezért sok webes tartalom elveszett, rengeteg link, hivatkozás megszűnt működni. 2009-ben a könyvtár digitalizációs központjában összeállítottak egy kb. 50 tételes listát a gyűjtőkörük szempontjából legfontosabb honlapokról, amit később 130-ra bővítettek. A mentéseket a HTTrack⁴¹ szoftverrel kezdték el és negyedévente megismételték. (Egyébként a PANDORA konzorcium tagintézményeiben is ugyanezt a Windows alatt is futó, felhasználóbarát és ingyenes programot használják, melynek magyar kezelőfelülete is van.) 2010-ben 255 ezer fájl, összesen 13 gigabájt volt az archívum mérete, egy 2011-es adat szerint pedig már 26 gigabájt. Hozzáférést nem biztosítottak, mert csak a szolgáltatók 10%-át tudták e-mailben elérni, bár azok szinte 100%-ban engedélyt adtak rá.

További példaként említhető Hollandia Groningen nevű tartományának és a hasonló nevű városnak a webarchívuma⁴², melyet – más digitális és hagyományos gyűjteményekkel együtt – a Regionaal Historisch Centrum Groninger Archieven gondoz, az üzemeltetője pedig az Archiefweb.eu. A helytörténet, helyismeret szempontjából érdekes webhelyeket és blogokat gyűjtik, beleértve az audiovizuális tartalmakat is, és a közösségi média mentése is szerepel a terveik közt. Egy év alatt kb. ezer site-ot válogattak össze és töltöttek le, de jelenleg nincsen nyilvános szolgáltatás belőle. A projekt egyik érdekes eleme, hogy bárki ajánlhat megőrzésre érdemes URL címeket egy űrlapon keresztül. Olyan javaslatokat várnak, amelyek megfelelnek a következő kritériumoknak:

- a webhely nyilvános, vagy rövidesen az lesz;
- beletartozik legalább az egyik kategóriába:
 - groningeni székhelyű szervezet, intézmény, cég tartja fenn;
 - groningeri lakos készíti vagy róla szól;
 - Groningen nyelvével, kultúrájával, kulturális örökségével, történelmével, vagy természeti környezetével foglalkozik.

³⁹ [http://mekosztaly.oszk.hu/mediawiki/index.php/PANDORA_\(ausztr%C3%A1l\)](http://mekosztaly.oszk.hu/mediawiki/index.php/PANDORA_(ausztr%C3%A1l))

⁴⁰ http://mekosztaly.oszk.hu/mediawiki/index.php/Web_Archive_of_Cacak

⁴¹ <http://mekosztaly.oszk.hu/mediawiki/index.php/HTTrack>

⁴² http://mekosztaly.oszk.hu/mediawiki/index.php/Web_Archive_Groningen

Annak a jogát természetesen fenntartják, hogy a javaslatok elbírálásánál olyan szempontokat is mérlegeljenek, mint az egyes témák arányai az archívumban, vagy hogy mennyire értékes és reprezentatív az ajánlott forrás, illetve hogy mennyi tárhelyet igényel a megőrzése.

Ugyan nem közkönyvtári indíttatású, de a helyi hírek megőrzését segíti elő a Local Memory Project⁴³, ami az amerikai Old Dominion University Web Science and Digital Libraries Research Group nevű munkacsoportjának 2017-ben nyilvánosságra hozott projektje. Az ehhez kifejlesztett rendszer segítségével a felhasználók – elsősorban – helyi vonatkozású híreket gyűjthetnek ki a Google találati listáiból, illetve a Twitterről, majd ezeknek a linkjeiből gyűjteményeket hozhatnak létre, sőt archiválhatják is a weboldalakat az Archive.is⁴⁴ szolgáltatással és megoszthatják őket másokkal. A hírek gyűjtése és mentése egy Chrome bővítményen keresztül történik, és a paraméterezési lehetőségeken belül a Download collection gombbal a saját gépükre is letölthetjük a linkeket (maguk a weboldalak nem menthetők el lokálisan). A projekt adatbázisában 2018 márciusában 183 ország 3151 városának online hírforrásai voltak nyilvántartva: több mint 12 ezer újság és egyéb időszaki kiadvány (ezek csaknem fele az USA-ból), valamint kb. ezer tévé-, ill. két és fél ezer rádiócsatorna.

2017 végén a washingtoni Institute of Museum and Library Services az Internet Archive támogatásával elindított egy Community Webs⁴⁵ nevű két éves projekt, melynek keretében 27 amerikai közkönyvtár több mint 35 terabájtnyi, a helyi közösségekkel kapcsolatos webes tartalmat (pl. helyi hírportálok, blogok, közösségi média anyaga) archivál az Archive-It⁴⁶ szolgáltatás segítségével. A projekt során a könyvtárosok képzést is kapnak a webarchiváláshoz szükséges ismeretekről. A tananyagot várhatóan 2019 elején nyilvánosan is elérhetővé teszik.

A helyi web megőrzésének lehetősége nálunk

Ahogy a fenti példákból látható, az internetes tartalmak gyűjtése nemcsak egy központi intézmény, mondjuk a nemzeti könyvtár feladata, hanem abban szerepet vállalhatnak a legkülönbözőbb típusú, méretű és gyűjtőkörű memóriaintézmények is. Annál is inkább szükséges ez, mert a folyamatosan evolválódó és elképesztő ütemben növekvő digitális univerzum megőrzése sokkal-sokkal nagyobb feladat, mint amivel ezek az intézmények valaha is találkoztak. Csak reprezentatív minták mentésére van lehetőség kisebb-nagyobb gyakorisággal, így még ha esetleg ugyanazt a webhelyet is archiválja két intézmény, az eltérő szoftver vagy annak beállításai, illetve a különböző időpontok miatt a két „mementó” nem lesz ugyanaz.

Ha egy intézmény úgy dönt, hogy belekezd a webarchiválásba, akkor egyrészt definiálnia kell egy gyűjtőkört, ami egy városi vagy megyei könyvtár esetében hasonló lehet a korábban bemutatott groningeni példához: a település ill. a régió történetével, társadalmával, kultúrájával, földrajzával stb. foglalkozó weblapok és webhelyek, például a helybéli intézmények honlapjai, a regionális hírportálok, a helyi közélet és kultúra szereplőinek nyilvános oldalai stb. Ezeknek azután össze kell gyűjteni az URL címeit, amibe a tájékoztató kollégák mellett be lehet vonni tanárokat, diákokat, a helyi online sajtó munkatársait, vagy akár a könyvtár internetes felületének használóit. Ha nyilvános szolgáltatást is szeretnének belőle, akkor a szerzői és személyiség jogi szabályok miatt erre engedélyt kell kérni a tartalom tulajdonosától, ami szólhat adott időszakra is, vagy pedig valamennyi türelmi idő lejártá után léphet érvénybe. Hogy milyen szoftvekkal érdemes az archiválást megkezdeni, az függ az archívum tervezett méretétől, a kiválasztott webhelyek műfajától, technikai megoldásaitól, és a rendel-

⁴³ http://mekosztaly.oszk.hu/mediawiki/index.php/Local_Memory_Project

⁴⁴ <http://mekosztaly.oszk.hu/mediawiki/index.php/Archive.is>

⁴⁵ http://mekosztaly.oszk.hu/mediawiki/index.php/Community_Webs

⁴⁶ <http://mekosztaly.oszk.hu/mediawiki/index.php/Archive-It>

kezésre álló informatikai szakértelemtől. A korábban említettek mellett szóba jöhet még a Firefox kiegészítőként telepíthető ScrapBook X⁴⁷ és annak folytatása: a Web ScrapBook⁴⁸, az egyedi weblapok szabványos WARC⁴⁹ fájlokba való mentésre szolgáló WARCcreate⁵⁰, az elsősorban webkettes oldalak megőrzésére optimális, egy böngészési folyamatot rögzítő Webrecorder⁵¹, vagy a Mink⁵² nevű Chrome modul, amivel nyilvános archívumokba lehet egy kattintással weboldalakat elmenteni. A megfelelő szoftverek kiválasztásához és használatához tudunk segítséget adni⁵³, és az OSZK-ban kialakítás alatt levő infrastruktúra is rendelkezésre fog állni minden közgyűjtemény számára, így különösebb informatikai ismeret nélkül is lehet majd részgyűjteményeket létrehozni a Magyar Internet Archívumban.

Hosszabb távon egy olyan együttműködési hálózatot kellene kialakítanunk, ahol munkamegosztásban zajlik a válogatás, az engedélykérés, a mentések minőségellenőrzése és a metaadatolás, maga az archiválás pedig vagy közvetlenül az OSZK-ban történik, vagy helyi archívumokban, melyek az ún. Memento protokollon⁵⁴ keresztül egy közös gyűjteményként is látszanak. Hasonló ez a feladat a tudományos és oktatási jellegű digitális dokumentumok archiválására szolgáló repozitóriumok létrehozásához, melyeknek már jól kiépült hálózata van a magyar könyvtárakban is, és amelyeknél az OAI PMH protokoll⁵⁵ biztosítja a közös kereshetőséget. De amíg a repozitóriumokba többnyire maga a tartalom előállítója küldi be a dokumentumot, a webhelyeken található értékes tartalmakért nekünk kell elmennünk és begyűjtenünk őket, mielőtt eltűnnek örökre.

Ajánlott irodalom

Dancs Szabolcs: Webarchiválási politikák

In: Könyv, könyvtár, könyvtáros, 2011. (20. évf.), 10. sz.

http://epa.oszk.hu/01300/01367/00248/pdf/EPA01367_3K_2011_10_14-20.pdf

Drótos László: Az internet archiválása mint könyvtári feladat

In: Tudományos és Műszaki Tájékoztatás, 2017. (64. évf.), 7-8. sz.

<http://tmt.omikk.bme.hu/tmt/article/view/1025>

Hegyközi Iлона: Hol tart ma a webarchiválás?

In: Könyvtári Figyelő, 2014. 4. sz.

<http://ki.oszk.hu/kf/2015/01/hol-tart-ma-a-webarchivalas/>

Németh Márton: Nemzetközi körkép a webarchiválás gyakorlatáról

In: Könyvtári Figyelő, 2017. (63. évf.), 4. sz.

http://mekosztaly.oszk.hu/mia/doc/KF_2017_4_webarchivalas_korkep.pdf

⁴⁷ <http://mekosztaly.oszk.hu/mediawiki/index.php/ScrapBook>

⁴⁸ http://mekosztaly.oszk.hu/mediawiki/index.php/Web_ScrapBook

⁴⁹ <http://mekosztaly.oszk.hu/mediawiki/index.php/WARC>

⁵⁰ <http://mekosztaly.oszk.hu/mediawiki/index.php/WARCcreate>

⁵¹ <http://mekosztaly.oszk.hu/mediawiki/index.php/Webrecorder>

⁵² <http://mekosztaly.oszk.hu/mediawiki/index.php/Mink>

⁵³ kapcsolati cím: mia@mek.oszk.hu

⁵⁴ http://mekosztaly.oszk.hu/mediawiki/index.php/Memento_Project

⁵⁵ https://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting

OSZK WEBARATÁS - DEMÓ ARCHÍVUM - Mozilla Firefox

http://mekosztaly.oszk.hu/mia/demo/

OSZK WEBARATÁS – DEMÓ ARCHÍVUM

KERESÉS A TELJES SZÖVEGBEN...

KÖNYVTÁR | LEVÉLTÁR | MÚZEUM | MŰVÉSZET | KUTATÁS | OKTATÁS | ÖNKORMÁNYZAT | TÖRTÉNELEM | KÖNYV
E-PERIODIKA | BLOG | SZEMÉLYES

Ez a kis gyűjtemény az OSZK-ban zajló **kísérleti webarató projekt** keretében készül azokból a mentésekből, amelyeknél az eredeti honlap tulajdonos engedélyt adott a Nemzeti Könyvtár számára az archivált példány(ok) nyilvános szolgáltatására – egyelőre 2018 végéig. (További felajánlásokat, javaslatokat örömmel veszünk **ezen az úrlapon**.) A célja az, hogy demonstráljuk vele a jelenlegi webarchiválási technológia lehetőségeit és korlátait. Bár a demóhoz kiválasztott webhelyek az automatikus módszerekkel viszonylag jól archiválhatók közül kerültek ki, még így is előfordulnak hibák és hiányok a lementett példányokban. Ezen problémák egy része a site-ok **robot-barát** és **archívum-barát** kialakításával orvosolható.

A piros nyilakra kattintva mentések a **Heritrix** szoftverrel készültek 2017 decemberével kezdődően, nem teljes mélységben, a videók és más nagy méretű állományok letöltésének kizárásával, és az eredeti webhelyen levő **robots.txt** fájlban levő tiltások tiszteletben tartásával. A megjelenítés az **Open Wayback** szoftverrel történik, melyben a mentés dátumára kell kattintani az archivált példány megtekintéséhez. A mentett változatra és az első mentés időpontjában készült képernyőfotóra mutató nyíl mellett van egy-egy link az amerikai **Internet Archive**-ban található mentésekre, valamint az eredeti "élő" honlapra is (amennyiben az még létezik). A nyilakra kattintva a weblapok új böngészőlapon nyílnak meg, így könnyen összehasonlíthatók a mentett verziók és az eredeti oldalak. A sárga gomb pedig az adott doménről kifelé mutató, illetve a rá kívülről hivatkozó linkek gráfját rajzolja ki - az archívumban levő mentések alapján (mivel a demó archívum még kicsi, ezért ez utóbbi, *ingoing* típusú linkek száma nagyon kevés). Az utolsó oszlopban levő barna nyilak az archivált webhelyeket leíró metaadat rekordokra visznek (a böngészőben a Ctrl+U megnyomásával látszik az eredeti XML forráskódjuk). Ezek a leírások fokozatosan kerülnek ki az oldalra, ahogy visszamenőleg elkészítjük őket.

A fejlesztés alatt levő **SolrMLA** keresők mellett kísérleti jelleggel a **Solr Wayback Search** nevű felület is kipróbálható, amellyel szintén a mentett webhelyek teljes szövegében lehet keresni és a találatok szűkíthetők doménnevekre, fájl típusokra és a mentés éve szerint. A találati listában a weblap vagy fájl (nagy betűvel kiemelt) címére kattintva jutunk el az archivált verzióra, az *Url*: sorban levő cím pedig az eredeti honlapra/fájllra mutat. A *Show full post* felirat alatt megnézhetők az adott találat részletes adatai. Egy találatra kattitva további információk is megjeleníthetők az adott oldalról vagy doménről, ha a bal felső sarokban levő *Toolbar* eszköztárat lenyitjuk.

KÖNYVTÁRI HONLAPOK

A webhely neve	URL címe	OSZK mentés	Képernyőfotó	Linktérkép	Internet Archive	Eredeti oldal	Metaadatok
Berzsenyi Dániel Városi Könyvtár, Marcali	www.marcalikonyvtar.hu	🔴	🔵	🟡	🟣	🟢	🟠
Csuka Zoltán Városi Könyvtár, Érd	www.csukalib.hu	🔴	🔵	🟡	🟣	🟢	🟠
Evangélikus Országos Könyvtár	konyvtar.lutheran.hu	🔴	🔵	🟡	🟣	🟢	🟠
Esztergomi Főszékesegyházi Könyvtár	www.bibliotheca.hu	🔴	🔵	🟡	🟣	🟢	🟠
EX LIBRIS Könyvtár	exlibriskonyvtar.hu	🔴	🔵	🟡	🟣	🟢	🟠
Fejér György Városi Könyvtár, Keszthely	www.fgyvk.hu	🔴	🔵	🟡	🟣	🟢	🟠
Helischer József Városi Könyvtár, Esztergom	www.vkesztergom.hu	🔴	🔵	🟡	🟣	🟢	🟠
Jókai Mór Városi Könyvtár, Pápa	www.jmrvk.papa.hu	🔴	🔵	🟡	🟣	🟢	🟠
Keresztély Gyula Városi Könyvtár, Bátaszék	www.kgyvk.hu	🔴	🔵	🟡	🟣	🟢	🟠
Dr. Kovács Pál Megyei Könyvtár és Közösségi Tér, Győr	www.gyorikonyvtar.hu	🔴	🔵	🟡	🟣	🟢	🟠
MaNDA DB	mandadb.hu	🔴	🔵	🟡	🟣	🟢	🟠
Martonosi Pál Városi Könyvtár, Kiskunhalas	www.vkhalas.hu	🔴	🔵	🟡	🟣	🟢	🟠
Nagy Gáspár Városi Könyvtár, Budakeszi	ngvk.hu	🔴	🔵	🟡	🟣	🟢	🟠
Széchenyi István Egyetem Egyetemi Könyvtár, Győr	lib.sze.hu	🔴	🔵	🟡	🟣	🟢	🟠

1. ábra. A demó webarchívum részlete, benne a két győri könyvtár honlapjának mentésével

OpenWayback - Mozilla Firefox

http://193.6.201.202/owb/wayback*/http://www.gyorikonyvtar.hu/

OSZK WEBARATÁS - DEMÓ ARC... x OpenWayback x +

OpenWayback

Enter Web Address: All Take Me Back Adv. Search

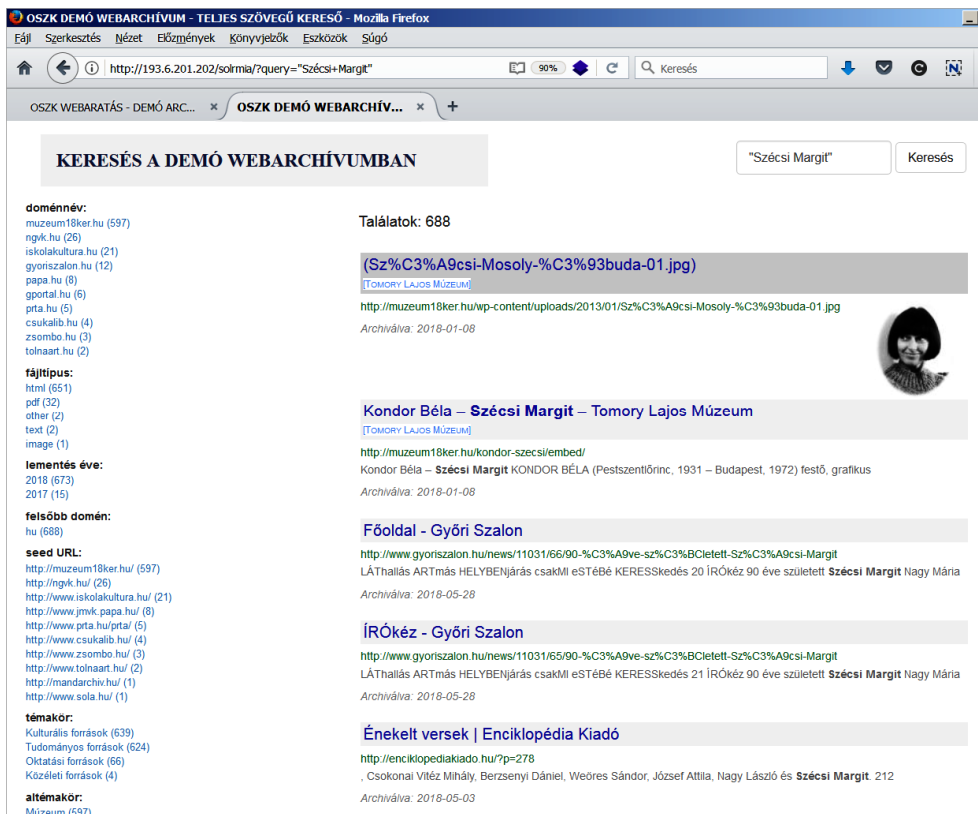
Searched for <http://www.gyorikonyvtar.hu/> Set Anchor Window: none 4 Results

Search Results for jan. 1, 2017 - dec. 31, 2018

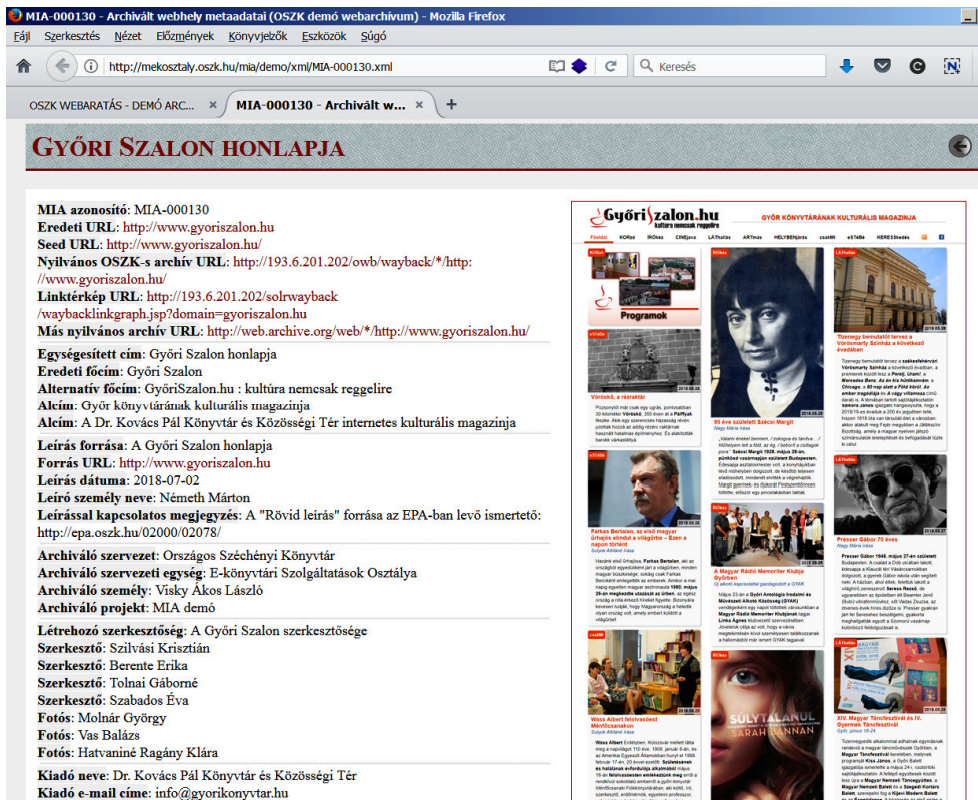
jan. 2017 - febr. 2017	márc. 2017 - ápr. 2017	máj. 2017 - jún. 2017	júl. 2017 - aug. 2017	szept. 2017 - okt. 2017	nov. 2017 - dec. 2017	jan. 2018 - febr. 2018	márc. 2018 - ápr. 2018	máj. 2018 - jún. 2018	júl. 2018 - aug. 2018	szept. 2018 - okt. 2018	nov. 2018 - dec. 2018
0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	2 pages	2 pages	0 pages	0 pages
								máj. 28. 2018 *	júl. 6. 2018 *		
								máj. 28. 2018 *	júl. 6. 2018 *		

Home | Help

2. ábra. A www.gyorikonyvtar.hu mementői az OpenWayback programban



3. ábra. Részlet a saját fejlesztésű keresőprogram egyik találati listájából



4. ábra. Az archivált Győri Szalon hírportál metaadatrekordjának részlete