



Hol tart ma a webarchiválás?

Szakirodalmi áttekintés

HEGYKÖZI Ilona

Az utóbbi bő másfél évtizedben sok ország nemzeti könyvtárában vált a digitális kultúra megőrzése, a hálózati források archiválása a rendszeres feladatok egyikévé. Olyan webarchívumok jöttek létre – kísérleti jelleggel vagy állandósuló szolgáltatásként –, amelyek az internetes források kiválasztásával, gyűjtésével és megőrzésével foglalkoznak saját országuk webterét tekintve, esetenként a kötelezpéldány-szolgáltatással is szoros összefüggésben. A webarchiválás – mint látni fogjuk – az egyes országokban mind volumenében, mind módszereiben igen sokféleképpen zajlik.

Bevezetésként, a webarchiválással kapcsolatos fogalmak és munkafolyamatok előzetes gyors áttekintéséhez kapóra jön az ISO 2013-ban megjelent nemzetközi jelentése (*Technical Report 14873: Information and documentation – Statistics and quality issues for web archiving*), amelynek tárgya a webarchiválással kapcsolatos statisztika és minőségi értékelés. A tömör fogalmi tisztázó rész ezt tartalmazza:

„A webarchiválás sokféle internetes forrás kiválasztását és összegyűjtését jelenti, amelyeket aztán webarchívumokban tárolnak, tartósan megőriznek és hozzáférhetővé tesznek. A források begyűjtése túlnyomórészt automatikusan

és rendszeres időközönként *harvesting-szoftverek* (*crawlerek*) segítségével történik, amelyek megadott szabályok szerint és/vagy jegyzékek alapján vizsgálják át az internetet.

A begyűjtés (*webaratás*) átfogóan vagy válogató módszerrel történik. Az átfogó begyűjtés (*bulk crawl*) meghatározott formális szabályok szerint zajlik; például egy bizonyos nemzeti webtér (*domén*) tartalmát gyűjtik be (*.fr* vagy *.de*). Válogató begyűjtés esetén előre megadják az érintett webhelyek címeit: ezek vagy bizonyos témakörökkel (például az illető ország kultúrával és politikai életével), vagy bizonyos eseményekkel kapcsolatosan az összes internetes forrást begyűjtik (*event harvesting*). A válogató begyűjtést általában gyakoribb időközönként végzik, mint az átfogó begyűjtést, például fél évente készítenek egy-egy „pillanatfelvételt”.

A webarchiválás feladata legtöbbször szoros összefüggésben van a kulturális örökség hagyományos gyűjtésével és megőrzésével, például a köteleespéldányok szolgáltatásával. A gyűjtés indoka mindkét esetben azonos: fennáll a veszély, hogy a publikációk – mind a nyomtatottak, mind a webhelyek – elveszhetnek az utókor számára. A veszély az internet-tartalom esetében még nagyobb, mint a hagyományos publikációknál; közismert, hogy milyen gyorsan eltűnnek webhelyek. Sok országban létrejött már a jogi struktúra a webarchiváláshoz, a köteleespéldányokkal foglalkozó jogszabályokhoz hasonlóan. A létrejövő nemzeti webarchívumok nemcsak megőrzik a kulturális örökséget, hanem hosszú távon elérhetővé is teszik a kutatók és más érdeklődők számára.

A web archiválása az 1990-es évek végén kezdődött, elsősorban a nemzeti könyvtárakban, részben azonban a nemzeti levéltárakban és más, szakosodott gyűjtéssel megbízott intézményekben is. A kezdetektől világos volt, hogy ennek a hatalmas feladatnak a megoldásához összefogás szükséges. Ezért jött létre 2003-ban az IIPC (International Internet Preservation Consortium). A konzorcium a webarchiválás jogi és mindenekelőtt technikai vonatkozásaival foglalkozik. Nagy súllyal részt vett például a WARC-formátum (Web ARChive file format)

kifejlesztésében, amely a későbbiekben ISO 28500 néven a webtartalmak összekapcsolásának, tárolásának és feldolgozásának szabványa lett.

*

Szakirodalmi áttekintésünk elején visszautalunk Drótos László 2006-ban megjelent tanulmányára, amely igen alaposan, módszeresen és rendszerbe foglalva ismertette a webarchiválás akkori helyzetét. A következő kérdéseket járta körül: miért van szükség az internetes tartalom archiválására; mi a lényege a web archiválásának; ki végezze el ezt a feladatot; milyen külföldi kísérleti programokra került sor addig; milyen tanulságokat hoztak ezek a programok; milyen technikai vonatkozásokkal kell számolni.

Drótos László cikke második részében részletesen kidolgozott javaslattal állt elő egy *Magyar Internet Archivum (MIA)* létrehozására. Kiindulásképpen felvázolta egy 2–3 éves kísérleti projekt általános és szakosodott munkacsoportokban elvégzendő teendőit. Bár a könyvtáros szakma nagy elismeréssel fogadta előterjesztett tervét, tény, hogy – legalábbis tudomásunk szerint – a mai napig nem indult el a MIA vagy más, hasonló program megvalósítása. Nyolc év elteltével ideje megvizsgálni, hogy mi történt időközben világszerte, hol tart ma a webarchiválás más országokban.

Drótos László cikkében az Internet Archive, a skandináv országok közös kezdeményezése (Nordic Web Archive), az ausztrál, a brit nemzeti könyvtár, a Library of Congress és a cseh nemzeti könyvtár akkori gyakorlatát ismertette. Összeállításunkban ezekkel is foglalkozunk, de kitekintünk néhány más kezdeményezés tapasztalataira is. Forrásaink a webarchívumok honlapjai, valamint hazai és külföldi cikkek (utóbbiak főleg a Könyvtári Figyelő folyóirat-figyelő rovatának referátumai) voltak.

Amire sokszor hivatkoznak: Ausztrália

Ausztráliában a nemzeti könyvtár foglalkozik – már 1996-tól – a digitális információk hosszú távú megőrzésével. Az erre a célra létrehozott *PANDORA-archívumban* őrzik a nemzeti jelen-

tőségű, hosszú távú megőrzésre érdemes, válogatott webes forrásokat. 2013 márciusában az archívum mérete 8,19 terabájtot tett ki. A web tartalmának automatikus begyűjtésére, leírására és hozzáférhetővé tételére kifejlesztették a *PANDAS* szoftvert.

A PANDORA szelektív archívum, a válogatás fő szempontja a webhelyek jelentősége és kutatási értéke. Tíz partnerrel működnek együtt (ezek: az egyes államok könyvtáraival, társadalomtudományi kutatóintézetekkel, a nemzeti film- és hangarchívummal, a nemzeti galériával). Az egyes partnerek határozták meg saját gyűjtőkörüket. A nemzeti könyvtár értelemszerűen az országos jelentőségű dokumentumokra összpontosít, az állami könyvtárak az egyes államok és régiók szempontjából fontos dokumentumokra, míg a ScreenSound Australia a zenei és filmmel kapcsolatos webhelyekre specializálódik, és így tovább.

Begyűjtés és beszolgáltatás: Ausztria

Ausztriában az Osztrák Nemzeti Könyvtár (ÖNB) a műszaki egyetemmel közösen 2000-ben kezdett egy kísérleti program keretében a webarchiválással (az „osztrák internet” archiválásával) foglalkozni. Ekkor módosították az úgynevezett médiatörvényt: a beszolgáltatandó dokumentumok és egyben az ÖNB gyűjtőköre kibővült az offline kiadványokkal. A 2009. évi módosítás már felhatalmazta az ÖNB-t, hogy az online kiadványokat is gyűjtse, és létrehozza az osztrák webarchívumot (*Web@rchiv Österreich*). Átfogó, automatikus begyűjtésre évente max. négy alkalmat irányoztak elő (nem kötelezően, hanem megengedő jelleggel; jelenleg a Webarchiv honlapján kétévenkénti gyakoriságot említenek), ekkor az „időszaki elektronikus média” tartalmaira (webhelyek, e-folyóiratok, e-könyvek stb.) összpontosítanak, amennyiben azok az .at doménbe tartoznak vagy osztrák vonatkozásúak (a domének az utóbbi esetben változatosak lehetnek, pl. .com, .net, .org, .info, .cc, .eu stb.). A válogató begyűjtés körébe a gyorsan változó oldalak tartoznak (hatóságok, tudományos testületek webhelyei). Események-

hez kapcsolódóan is végeznek archiválást. Az ÖNB emellett egyedi médiatartalmakat is gyűjthet; erről a médiatulajdonosokat előre értesítenie kell. A szelektív begyűjtés szempontjából a következő ismérvekkel rendelkező dokumentumok jöhetnek szóba: amelyek nem jelentek meg nyomtatásban, tartalmuk zöme nem hangzó anyag vagy film, nem határozott ideig jelennek meg, valamint a tartalmak beadása, tárolása és megőrzése viszonylag egyszerűen kivihető.

A webhelyek begyűjtése mellett a webhelyek beszolgáltatására is van mód; ez szabványos interfészek közvetítésével zajlik, akár online, akár offline üzemmódban (DVD-n). Maga az archívum tudományos célokat szolgál. A Webarchiv nemzeti könyvtárban és az erre feljogosított könyvtárakban (ezek száma megközelíti a húszat) helybeni használatra áll rendelkezésre speciális terminálok. A nemzeti könyvtár honlapján alapos kérdezz-felelek összeállítás olvasható a webarchívumról, annak működéséről, a beszolgáltatás módjáról, a jogi és műszaki vonatkozásokról.

A csehek megcsinálták...

Csehországban 2000-től van napirenden a web archiválásának a kérdése, 2001-ben meg is kezdődtek a munkálatok. A *WebArchiv* projektben a nemzeti könyvtár partnerei a Morva Könyvtár és a brnói Masaryk Egyetem Informatikai Intézete. A projekt célja a cseh webes kulturális örökség megőrzése. Csehországban mind a nemzeti jelentőségű webforrások begyűjtését, mind a válogató, tematikus gyűjtést alkalmazzák. A kötelempéldányokról szóló törvény azonban még nem tér ki a hálózati dokumentumokra. A 2006. júniusi szerzői jogi törvény megfelel az EU 2001/29/EC direktívájának, azaz az archívum elérhető a nemzeti könyvtárban helybeni használatra. Egyes fontos internetes forrásokról megállapodást kötnek a kiadókkal annak érdekében, hogy a könyvtár jogosult legyen elérhetővé tenni azokat hálózatában; ezekhez külön szerveren biztosítják a hozzáférést.

Az internetes források begyűjtése automatikusan történik. A források beszerzését, indexelését

és megőrzését előre meghatározott paraméterek szerint végzik. A szoftverek nagy része nyílt forráskódú (például a Heritrix; a teljes szövegű indexelés a Nutch-csal történik, a hozzáféréshez a Nutchway rendszert és a WERA felületet használják), és az IIPC dolgozta ki. A projekt résztvevői saját szoftvereket is kifejlesztettek. A fájlok és az adatok tekintetében a szabványos archív formátumra törekшенek.

A cseh WebArchiv a következőket tartalmazza: az interneten szabadon elérhető digitális dokumentumokat, kutatási és művészeti tárgyú kiadványokat, híreket és aktuális információkat, időszaki kiadványokat, monográfiákat, konferenciái anyagokat, kutatási és más jelentéseket, tudományos kiadványokat, csak digitális formátumban létező szöveges, valamint egyes vizuális és hangzó dokumentumokat.

A csehországi projektnek az a célja, hogy mindent archiváljanak, amit valaha is közzétettek a cseh webtérben. Ez a cél természetesen technikailag nem teljesíthető, és a weben publikált dokumentumok mindegyike nem is alkalmas archiválásra. Ezért a begyűjtés és archiválás mindhárom szokásos módszerét alkalmazzák: az átfogó webaratást, a válogató begyűjtést és a tematikus gyűjtemények létrehozását (országos jelentőségű eseményekről). Mindhárom módszerhez részletes útmutató áll rendelkezésre.

A WebArchiv keresőfelületén első lépésben a WebArchivban és az Internet Archive-ban lehet keresni. Magát a gyűjteményt is három csoportba sorolva „tálalják”: átfogó témák (Subject Directory), átfogó webaratások és tematikus gyűjtemények (Európai Parlamenti választások, 2009; Cseh elnökség az EU-ban, 2009; Cseh-szlovákia 1968-as megszállásának negyvenedik évfordulója, 2008; elnökválasztás, 2008; a nemzeti könyvtár új épülete, 2007; olimpiai pályázat, 2007; árvizek, 2002 stb.).

A webaratások statisztikai adatai közül utoljára 2007-es adatok szerepelnek a WebArchiv honlapján. A .cz doménből 2007-ben 81 300 000 rekordot gyűjtöttek be, ezzel az állomány 3 600 000 megabájtnyira nőtt (az összes rekord száma: 206 827 686, az állomány 8 799 382 megabájt).

Webarchiválás együttműködésben: Dánia

Dániában a kötelezpéldány-szolgáltatásról szóló jogszabályt 1997-ben kiterjesztették a fizikai hordozón közreadott elektronikus publikációkra és a statikus hálózati dokumentumokra (módosítva 2005-ben). A Dán Királyi Könyvtár 1998-tól kezdte meg az internetes források szelektív archiválását. A regisztrált források közé a kiadók engedélyével bekerültek az elektronikus folyóiratok is. A jogszabály értelmében a kiadóknak be kell jelenteniük kiadványaikat, a Királyi Könyvtár így értesül azok címéről. Ezt követően a munkatársak begyűjtik, ellenőrzik, feldolgozzák és archiválják a dokumentumokat.

A Királyi Könyvtár, az Aarhusi Állami Könyvtár és az aarhusi egyetem internet-kutatási központja kísérleti jelleggel webarchívumot hozott létre (*Netarkivet.dk*) a különböző archiválási stratégiák tesztelése érdekében. A tapasztalatok alapján módosították a kötelezpéldány-rendeletet. A két együttműködő könyvtár közösen gondoskodik „a dán internet” tartalmának beszerzéséről és megőrzéséről. Közös archívumot építenek, amely biztonsági okokból mind Koppenhágában, mind Aarhusban megtalálható. Az archívumot csak kutatók használhatják, mégpedig azok, akik erre engedélyt kértek és kaptak a Dán Adatvédelmi Ügynökségtől.

A Netarkivet.dk statisztikai adatai szerint évi négy alkalommal készítenek pillanatfelvételeket. Legutóbb 2014. augusztus vége és október közepe között 32 854 767 537 265 bit adatot gyűjtöttek be, közel 610 millió dokumentumot.

A szelektív webaratás keretében 80–100 kiválasztott, gyakran frissülő webhelyet (pl. közszolgálati médiát, olyan gyakran látogatott weboldalt és portálokat, amelyek a civil társadalom, a hatóságok és a kereskedelmi szektor tevékenységét tükrözik, és wikiket, blogokat, egyedi és személyes oldalak tartalmát akár naponta többször is rögzítik).

A kiemelt eseményekhez (számuk évente 2–3) kapcsolódó webaratás körébe azok az események kerülnek be, amelyek felkeltették a közvélemény figyelmét. Ezek között vannak „ki-

számítható” (például a választások) és „kiszámíthatatlan” események egyaránt (például egy természeti katasztrófa vagy a Mohamed-karikatúrák nyomán kialakult válság).

Közös keresési lehetőség: Egyesült Királyság

Az Egyesült Királyságban a nemzeti webarchívumba (*UK Web Archive*) olyan webhelyek kerülnek, amelyek „a brit emberek életének, érdeklődésének és tevékenységeinek sokszínűségét mutatják be, és tükrözik a webes technológiához kapcsolódó innovációkat”. Gyűjtőkörébe beletartoznak az úgynevezett „szürke irodalom” forrásai is, vagyis azok a webhelyek, amelyek jelentéseket, állásfoglalásokat és más efemer jellegű, de igen fontos információt tartalmaznak. A webhelyekről rendszeresen pillanatfelvételek készülnek. Az 2004-től épülő archívum szabadon hozzáférhető a weben. Az archívumban rögzített webhelyek körének bővítésére bárki javaslatot tehet.

2013 áprilisától a British Library megkezdte a teljes brit webtér archiválását a nem nyomtatott dokumentumokra vonatkozó, 2013. évi jogszabály előírásainak megfelelően. A munkába a kötelezpéldányra jogosult öt másik könyvtár

is bevonták (a skót és a walesi nemzeti könyvtárat, a cambridge-i egyetemi könyvtárat, az oxfordi Bodley könyvtárat, a dublini Trinity College könyvtárat). További partnerek: a JISC, The Wellcome Library, The Women’s Library (London Metropolitan University) stb. Az archívum csak a kötelezpéldányra jogosult könyvtárakban érhető el, speciális engedéllyel pedig a webhelyek szűk köre külső helyszínről is.

A brit nemzeti webarchívumban vannak „különgyűjtemények” is. Ezekben az összeállításokban olyan webhelyek szerepelnek, amelyeket könyvtárosok, kutatók és más szakemberek az érintett intézményekkel együttműködve választottak ki. Ezek köthetnek egy-egy eseményhez (pl. olimpiai és paralimpiai játékok, 2012), témához (pl. a Credit Crunch Collection, amely 2008 júliusa és 2010 januárja között a globális pénzügyi válság eseményeit rögzítette bankok, vállalati, önkormányzati, fogyasztói és pénzügyi tanácsadói webhelyek és blogok alapján) vagy témakörhöz (pl. The British Countryside Collections).

A nemzeti webarchívum infrastruktúráját a British Library biztosítja. Igyekszik jobb keresési lehetőséggel és egyéb szolgáltatásokkal

The screenshot displays the Mementos search interface. At the top, there's a banner for the UK Web Archive with a grid of archived web pages from various dates (e.g., August 2005, November 2005, May 2006, June 2007, March 2009). Below this, the page title is 'Mementos' and the subtitle is 'Finding historical versions of web pages (mementos) captured by multiple archives across the world.' The main content area explains the Memento protocol and provides instructions on how to use the service. On the left, there's a sidebar with navigation links like 'Home', 'About', 'Search the archive', and 'Browse the archive'. On the right, there's a 'Quick links' section with questions like 'What is the UK Web Archive?' and 'Who is the UK Web Archive for?'. At the bottom, there are two example search results showing the history of 'www.webarchive.org.uk' and 'www.bl.uk'.

Weboldalak régi változatainak keresése több archívumban (Memento)

a „használói élményt” is javítani. A UK Web Archive-ot a British Library biztonságos digitális repozitóriumában is megőrzik.

A brit webarchívum *Mementos* néven hasznos szolgáltatást, közös keresési lehetőséget nyújt a világ webes archívumaiban, pontosabban azok metaadataiban. Ezzel egy-egy webhely történetét lehet áttekinteni, tekintet nélkül arra, hogy az illető webhelyet hol dokumentálták. A keresési eredmények megjelenítésekor kiderül, mely archívumokban és hány alkalommal rögzítették az adott webhely tartalmát egy bizonyos időszakban. Ennél bonyolultabb keresés is lehetséges különböző kiegészítő alkalmazások (pl. a Memento Fox) segítségével.

2014 novemberében az archivált webhelyek száma meghaladta a tizennégy ezret, az egy-egy címre vonatkozó pillanatképek száma a 67 ezret, az archívum mérete pedig majdnem elérte a 25 terabájtot.

Hatalmas ország – hatalmas archívum: USA

Az Egyesült Államokban a Kongresszusi Könyvtár webarchívuma (*The Library of Congress Web Archives, LCWA*) a webarchiválás fő intézmé-

nye. Szakreferensek és kurátorok által kiválasztott, archivált webhelyek gyűjteményeiből áll. A Kongresszusi Könyvtár arra törekszik, hogy a jövő kutatói számára értékelje, kiválassza, összegyűjtse, feldolgozza, megőrizze és hozzáférhetővé tegye a digitális dokumentumokat. A webarchívumok fejlesztését a Minerva fejlesztési program keretében kezdték meg.

Az indulás éveiben (2000-től) egyes témákkal vagy eseményekkel kapcsolatos webhelyeket gyűjtötték be (ilyenek voltak a választások, az iraki háború és a szeptember 11-i események). 2004-ben figyelmüket kiterjesztették a veszélyeztetett digitális tartalomra is.

A Kongresszusi Könyvtár meghatározott időszakonként és a legkülönbözőbb formátumokban archiválja a web kiválasztott tartalmát. A webarchívumot kiszervezték. Havi gyarapodása megközelíti az 5 terabájtot. Az amerikai kötelempéldány-rendelet nem tér ki a Kongresszusi könyvtár digitális megőrzési és webarchiválási feladataira, ezért minden érintett weboldal tulajdonosát igyekeznek megkeresni és értesíteni az archiválás tényéről, és ha szükséges, hozzájárulását kérné a szolgáltatáshoz.

The screenshot shows the Library of Congress Web Archives search results page. At the top, there is a navigation bar with 'LIBRARY OF CONGRESS', 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS'. Below this is a search bar with 'Archived Web Site' selected in the dropdown and a 'GO' button. The page is titled 'Archived Web Sites' and shows a search for 'Archived Web Site' with 940 results. The results are sorted by 'Relevance' and displayed in a grid format. The grid contains four thumbnails: 'The Nixon Center', 'Warblogging.com', 'European and Pacific Stars & Stripes', and 'Calon Legislatif Indonesia 2009'. The page also includes a 'Refine your search' section with options for 'Available Online' (940) and 'All Items' (940), and a 'FORMAT' section with options for 'Original Formats' (Archived Web Site) and 'Online Formats' (Image, Web Page, PDF, Dates).

A Kongresszusi Könyvtár webarchívumának katalógusa

A Kongresszusi Könyvtár webarchívumai között szerepelnek: (elnöki, kongresszusi és kormányzói) választási webarchívumok, kongresszusi képviselők és bizottságok weboldalai, a legfelső bíróság változásait dokumentáló források; oktatási, kutatási intézményekhez kötődő jogi tárgyú blogok. Emellett archiválnak olyan, konkrét eseményhez vagy témához nem kapcsolódó webhelyeket, amelyeket a könyvtár szakemberei választanak ki; ezek a digitális formátumban létrejött tartalmak az analóg állományt egészítik ki. Vannak olyan archívumok, amelyek egyedi webhelyeket tartalmaznak. Néhány téma a tematikus gyűjteményekből: a polgárháború 150. évfordulója, az 1812. évi háború 200. évfordulója, a darfuri (Szudán) válság, pápaválasztás. A távoli események dokumentálásába bevonják a Kongresszusi Könyvtár tengerentúli irodáit is (Kairó, Újdelhi, Rio de Janeiro, Iszlámábád, Jakarta). A könyvtár partnerként együttműködik a világ más intézményeivel különböző természeti katasztrófák és társadalmi történések megörökítésének érdekében (például a francia nemzeti könyvtárral az ukrajnai konfliktus és a tunéziai jázminos forradalom ügyében). A Kongresszusi Könyvtár hatással van a személyes archiválás elterjedésére is: egyre többen igyekeznek ugyanis elsajátítani, hogyan lehet digitális formában archiválni weblapjaikat, fotóikat stb. a családtörténet forrásaként a következő generációk számára.

Webes források a katalógusban: Németország

Németországban a Német Nemzeti Könyvtárról (Deutsche Nationalbibliothek, DNB) szóló törvény életbe lépésekor, 2006-ban a DNB megkapta azt a feladatot, hogy gyűjtse, feltárja, regisztrálja és archiválja a nem fizikai hordozón megjelent (hálózati) dokumentumokat; ezzel gyűjtőkörébe bekerültek a webhelyek is. A DNB egy szolgáltatót bízott meg, hogy kiválasztott weboldalakat automatikus eljárással gyűjtsön be. Ezzel az eljárással pillanatfelvételt készítenek. Egy adott URL-ről elindulva annak tartalmát, valamint az ehhez linkekkel kapcsolódó tartal-

makat tárolják. További külső partnerekkel is együttműködésre törekszenek.

A begyűjtés szelektív jellegű. Kiválasztott intézményekhez (például szövetségi hatóságok, érdekképviselői szövetségek, kulturális intézmények stb.), valamint bizonyos eseményekhez (például a 2013. évi Bundestag-választáshoz) kötődő weboldalakra összpontosítanak. A weboldalak körét bővítik. A webarchívumok kiválasztását a DNB munkatársai végzik tartalmi és formai kritériumok alapján. A weboldalakat működtetőiknek jelenleg nem kell bejelenteniük a nemzeti könyvtárnak, de lehetséges, hogy a későbbiekben lehetőséget adnak az önkéntes bejelentésre is. Ha szükséges és technikailag lehetséges, bizonyos eseményekhez kapcsolódva kiválasztott oldalakat a közösségi hálóról (Facebook, Twitter stb.) is begyűjtjenek. A közösségi háló tartalmának általános gyűjtését azonban nem tervezik.

A begyűjtést meghatározott időközönként, rendszeresen végzik el. A gyűjtés gyakoriságát az egyes weboldalaknál egyedileg állítják be, aszerint, hogy tartalmát mennyire tartják relevánsnak, illetve az milyen gyakran változik. Jelenleg a negyedévenkénti gyakoriság a legjellemzőbb.

A weboldalak jogi megfontolások alapján csak a DNB olvasótermeiben férhetők hozzá. A weboldalakat és az egyes pillanatfelvételeket mind a katalógusból, mind a webarchívumból belépve lehet hívni.

Az archivált weboldalakat a tájékozódást segítő kategóriákba csoportosítva sorolják fel. Egy-egy címre kattintva az illető cím összes archivált verziójáról áttekintést lehet kapni az archiválás dátumával és órájával egyetemben. A webarchívumban az archivált weboldalak tartalma teljes szöveggel is kereshető. A keresés szűkítéséhez két menüből lehet szempontokat választani. A keresést egy bizonyos időszakra is korlátozni lehet. A keresés eredményét megjelenítve az archivált weboldal egy keretben jelenik meg, ami azt jelzi, hogy nem egy aktív, hanem egy archivált weboldallal van dolgunk. A fejlécben megjelenik a begyűjtés dátuma, az eredeti URL és lehívható a fájl rekordja a DNB katalógusából.

The screenshot shows the DNB search results for the query "Willy" and "Brandt" and "100." and "Geburtstag". The search results table contains the following information:

Link zu diesem Datensatz	http://d-nb.info/1049442369
Titel/Bezeichnung	FAZ.net, Festakt zum 100. Geburtstag [Elektronische Ressource] [Archivierte Website]
Umfang/Format	Online-Ressource
Anmerkungen	Langzeitarchivierung gewährleistet
URL	http://www.faz.net/aktuell/politik/inland/festakt-zum-100-geburtstag-zwei-bundespraesidenten-ehren-willy-brandt-12708244.html
gehört zu	100. Geburtstag Willy Brandt
Zugehörige Zeitschnitte	1 Zeitschnitt 1. 17.03.2014 Zu: FAZ.net, Festakt zum 100. Geburtstag
Online-Zugriff	Webarchiv öffnen (nur im Lesesaal möglich)

Online források katalógusrekordja a DNB-ben: Száz éve született Willy Brandt

A téma iránt érdeklődők további hasznos információkat találhatnak az IIPC honlapján és a webarchiválásról szóló éves bibliográfiákban.

Felhasznált források

DRÓTOS László: Mi a MIA? Javaslat egy Magyar Internet Archivum létrehozására = Tudományos és Műszaki Tájékoztatás, 53. évf. 2006. 6. sz. 264–274. p.

Ausztria:
<http://www.onb.ac.at/about/webarchivierung.htm>
[2014. dec. 2.]

Ausztrália:
<http://pandora.nla.gov.au>; [2014. dec. 2.]

Csehország:
<http://en.webarchiv.cz/>
brosúra:
http://en.webarchiv.cz/files/dokumenty/letak/letak07_eng.pdf
[2014. dec. 2.]

Dánia:
<http://netarchive.dk/in-english>
[2014. dec. 2.]

Egyesült Államok:
<http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>
[2014. dec. 2.]

Egyesült Királyság:
<http://www.webarchive.org.uk/ukwa>
[2014. dec. 2.]

Németország:
http://www.dnb.de/DE/Netzpublikationen/Webarchiv/webarchiv_node.html
[2014. dec. 2.]

POLL, Roswitha: Internationaler Fachbericht für die Archivierung des Internets = B.I.T.Online, 17. 2014. 1. 23-26.
www.b-i-t-online.de; [2014. dec. 2.]

REYES AYALA, Brenda: Web Archiving Bibliography 2013. = UNT Digital Library. <http://digital.library.unt.edu/ark:/67531/metadc172362/>
[2014. dec. 2.]

Több információforrás a webarchiválásról az IIPC honlapján
<http://netpreserve.org/web-archiving/bibliography>
[2014. dec. 2.]