

Webtörténetírás az Internet Archive-ból készített képernyővideókkal

Digitális történelem, webtörténet és webhelytörténet

A globális Internet Archive és a nemzeti webarchívumok a digitális történelem kutatásának fő forrásai, mivel összegyűjtik és megőrzik az eleve digitális formában születő kultúrát, s így olyan tartalmakat lehet megtalálni bennük, amelyek sehol máshol nem kutathatók. Az 1990-es évek második felétől már elképzelhetetlen teljes körűen megírni valaminek a történetét kizárólag csak papírújságra és -könyvekre alapozva, figyelmen kívül hagyva a téma internetes lenyomatait.

De ezek az archívumok magának az internetnek illetve a webnek a múltját is őrzik, hiszen ha megnézzük egy intézményi vagy tematikus honlap időbeli változásait, akkor nemcsak az adott szervezet vagy téma történetét követhetjük nyomon, hanem azt is, hogy hogyan fejlődött az internetes technológia, mikor milyen megoldások voltak népszerűek a weboldalakon. A különböző időpontokban, de ugyanott fotózott állóképekből összefűzött – az idő múlását látványosan érzékeltető – *time-lapse* videókhöz hasonlóan egyes weboldalokról is készíthetők ilyen „dokumentumfilmek” a webarchívumokban megőrzött mementóknak köszönhetően. A WayBack Machine (web.archive.org) – amely önmaga is az internettörténeti kutatás tárgya lehetne – meg tudja mutatni egy adott URL cím egyes mentéseit, melyekből azután végigkövethető, hogy hogyan módosult például a Google keresőjének, a *New York Times* online kiadásának, a *Fehér Ház* honlapjának, vagy éppen egy esküvőkkel foglalkozó webhelynek a tartalma és megjelenése az évek során.

A WayBack Machine

Az Internet Archive eredetileg ezzel a szlogennel hirdette a WayBack Machine szolgáltatását: „szűrőj a múltbéli weben” és elsősorban az eltűnt web-

oldalak helyén megjelenő 404-es hibaüzenet el-lenszeréül hozta létre. Az Alexa eszköztár (alexa.com/toolbar) telepítése után, ha a böngészőprogram egy 404-es hibába botlott, jó esetben villogni kezdett egy kis WayBack ikon, jelezve, hogy a keresett weboldalból van legalább egy mentés az archívumban, amit az ikonra kattintva meg lehet tekinteni. Sőt, a felhasználó is hozzájárulhatott az archívum bővítéséhez úgy, hogy az általa meglátogatott webhelyek URL címeit a szoftver elküldte az Alexa cégnek, amely azután – ha még nem volt róluk mentése – learatta és eljuttatta őket az Internet Archive-ba.

Az 1990-es évek közepén kifejlesztett WayBack Machine rendszer a minél simább szűrőlési élményt helyezte a központba, mert akkoriban még – komolyabb linkgyűjtemények és keresőgépek hiányában – ez volt a fő navigációs mód az élő weben is. Ennek érdekében, ha a felhasználó valamelyik ponton belép a WayBack Machine szolgáltatásba, szabadon kattintgathat a linkekre és a rendszer megpróbálja kikeresni és megjeleníteni az időben legközelebbi mentéseket a hivatkozott weboldalakból, illetve az azokat alkotó objektumokból. Böngészés közben tehát időben is ide-oda ugrálunk akár hónapokat vagy éveket is, így a WayBack Machine inkább egy, a korabeli élő webet rekonstruálni próbáló múzeum, semmint egy pedáns levéltár.

Történeti szempontú kutatáshoz két funkciót biztosított a WayBack Machine: Ha beírtunk egy URL címet, akkor a találati listában a mentési dátumok mellett egy csillag jelezte azt, amikor komolyabb változás történt az adott oldalon az előző mentéshez képest, sőt össze is lehetett hasonlítani az oldal két állapotát.¹ A másik hasznos funkció a

¹ A 2017 elején megjelent új verzióban ez a lehetőség megszűnt. (a referátum készítőjének megjegyzése)

WayBack Machine fejléc, amely az éppen nézett archív weboldal fölött egy idődiagram és lapozóniák formájában lehetővé teszi, hogy nagyobb időbeli ugrásokkal haladva vizsgáljuk meg az oldal változásait (pl. a kommentelő felületek megjelenését, majd eltűnését a hírportálokon).

Digitális forráskritika

A WayBack Machine által a mentésekből rekonstruált régi weboldalak valójában „digitálisan újjászűletett” tartalmak, amelyek különböző kritériumoknak kell megfeleljenek ahhoz, hogy hiteles forrásként felhasználhatók legyenek. Kérdés, hogy mennyire tekinthetők ezek másolatoknak, vagy legalább az eredetileg megjelent tartalom elfogadható helyettesítőinek? Hiszen nem (teljesen) az eredeti kód és adatok alapján, és nem az eredeti szolgáltatási technológiával állítja elő őket a WayBack Machine, nem (teljesen) egyidejű mentésekből származó digitális objektumokból. Mindazonáltal már volt rá példa, hogy bíróságon is elfogadtak az Internet Archive-ből kinyomtatott weboldalakat bizonyítékként.

A web mint hiteles információforrás más szempontból is kétséges, ha figyelembe vesszük a kalózok, hackerek, hamisítók, laikus szerzők, álhír- és összeesküvés-terjesztők tevékenységét. (Erre a problémára a legutóbbi amerikai választási kampány különösen riasztó példa volt.)

További kérdés az archivált weboldalakra való hivatkozhatóság. Az élő weben tapasztalható linkromlás jelensége ellen hatásos eszköz a webarchívum, mert stabil URL azonosítót tud biztosítani például egy tudományos publikációban való hivatkozáshoz, ami mögött nem tűnik el és nem változik meg a tartalom. Hogy mennyire tekinthető „valid”-nak egy archivált példány ilyen szempontból, az egyebek mellett attól függ, hogy az archív verzió időbélyege egyezik-e azzal a dátummal, amikor még az eredeti változat online elérhető volt. Az MLA hivatkozási stílus legújabb kiadása inkább kiegészítőnek, mintsem helyettesítőnek ajánlja az archív példány URL címét, amennyiben egy link már nem működik, mert az eredeti URL (kiegészítve az utolsó hozzáférés dátumával) is hasznos lehet például a webhely hitelességének megítéléséhez, vagy a forrás új helyének megkereséséhez az élő weben.

Az archívumokból rekonstruálható webhelyek sokszor nem felelnek meg a történelmi elvárásoknak, különösen az online újságok, melyeknél a más

szerverekről beágyazott tartalmak esetleg egészen más időpontokban készült mentésekből származnak, a dinamikus, interaktív elemek pedig nagyrészt hiányoznak. Az ilyen szedett-vedett módon rekonstruálható oldalak könnyen lehet, hogy sosem léteztek ebben a formában az élő weben, vagyis sokkal inkább újjáalkotott dokumentumoknak számítanak, mintsem hiteles másolatoknak.

Mindezeket tehát figyelembe kell venni a digitális történet kutatóinak, amikor a webet, illetve annak archivált változatát forrásként kívánják használni.

Webhistoriográfia

Az eddigiekben tárgyalt „életrajzi jellegű” weboldal- és webhely-rekonstrukció mellett még legalább három további céllal készülnek webarchívumok:

- A második típus az esemény-alapú archiválás, amikor például jelentős rendezvények, politikai változások, vagy éppen természeti katasztrófák alatt és után (és esetleg már előtte is) gyűjtenek össze mindenféle információforrást az internetről az adott eseménnyel kapcsolatban.
- A harmadik változat a nemzeti webarchívum. Ennél egy nemzeti könyvtár kijelöli azt a webteret, amelyet a kulturális örökség szempontjából fontosnak tart és rendszeresen lementi azt. Dániában ezt például így definiálták: minden, ami a .dk országdomén alatt van, vagy dánul íródott, vagy a témája dániai, vagy jelentőséggel bír a dánok és Dánia szempontjából. A nemzeti webtérhez tartozó online tartalmak válogatásával a könyvtárosok újra elkezdtek azt a web-kurátori tevékenységet, amelyet korábban a tematikus linkgyűjtemények szerkesztésével végeztek. (Ezek mára már szinte mind eltűntek, mert a web gyors növekedése és változása, valamint a keresőgépek népszerűsége miatt nem volt értelme karbantartani őket.)
- A legújabb, negyedik fajta archiválást „önéletrajznak” nevezhetnénk, mert ezek a tartalmak személyhez kötődnek, gyakran csak jelszóval védett közösségi médiaplatformokon vagy okostelefonos alkalmazásokon keresztül érhetőek el, s a hagyományos webarchiváló technikákkal nem menthetők le és nem jeleníthetők meg. A külalak és a funkcionalitás megőrzése helyett vagy csak a puszta tartalom mentése jöhet ilyenkor szóba (pl. egy Facebook-profilhoz tartozó szöveg- és képanyag exportálása), vagy mondjuk egy videofelvétel készítése a mobiltelefon képernyőjéről, miközben a tulajdonosa használja. Érdekes kísérlet ebben a műfajban a digitális művészettel foglalkozó *Rhizome* szer-

vezet webrecorder.io nevű és című szolgáltatása, amivel felvehető és visszajátszható egy webböngészési folyamat.

Képernyővideók

Az életrajzi jellegű webhistoriográfia egyik látványos formája képernyővideók (*screencast*-ok) készítése és kielemezése. *Eelke Hermens* 2011-ben a *nytimes.com* időbeli változásáról állított össze egy kisfilmet^[1], amin az látszik, hogy a *New York Times* kezdetektől fogva kísérletezett az online média nyújtotta új lehetőségekkel, melyek később vagy eltűntek a honlapról, vagy jelentősen átalakultak, vagy újak kerültek a helyükre.

Jelen cikk szerzője, *Richard Rogers* 2013-as könyvében^[2] egyebek mellett azt vizsgálta, hogy a 9/11-es terrorista támadás után hogyan változott meg a fontos közügyekkel foglalkozó *issues* oldal tartalma a *whitehouse.gov* honlapon: szinte csak olyan témák jelentek meg rajta, amelyekben a „biztonság” szó előfordult.

Livio, Mataly és *Schuh* 2012-es videójukban^[3] a *TheKnot.com* című, esküvői témájú webhely tizenhárom éves történetét dolgozták fel, képernyőfotókkal bemutatva az elüzettesítés folyamatát, amint az egyszerű tanácsadó oldalból egy komplex vállalkozás lett: esküvőtervező alkalmazással, termékelhelyezésekkel és webshoppal. (Illusztrálva ezzel egyben a fogyasztói társadalom hatását is az életünk személyes eseményeire.)

Jon Udell 2005-ben készített egy képernyővideót^[4] a „Heavy Metal Umlaut” című Wikipédia-szócikkről, aminek a témája a rockzenekarok nevében az ékezetes betűk használata. (Ezzel a tipográfiai megoldással próbálják németessé, régiessé, vagy egyszerűen csak idegenszerűvé tenni a nevüket, főként az angolszász bandák.) Udell a múlt rekonstruálásához nem a WayBack Machine-t, hanem a Wikipédia szócikk saját *history* lapját használta, amivel visszaállítható bármelyik korábbi állapot a sok száz változtatás közül (beleértve a vandálok által törölt – de az éles változatban néhány perccel később helyreállított – szövegverziókat is). A videó bemutatja, hogy pár év alatt hogyan lett az eredetileg egymondatos szócikkből egy minden részletre kiterjedő, jól strukturált és a kulturális érzékenység szempontjából is korrekt szöveg – az önkéntes szerkesztők kollektív tudásának és odafigyelésének köszönhetően.

Az amszterdami *Govcom.org* alapítvány által készített film^[5] a Google keresőoldalának alakulását követi nyomon 1998 és 2007 között, az Internet Archive mentéseit felhasználva. Megfigyelhetjük, ahogy az egysoros kereső alatt és felett sorra megjelennek, majd eltűnnek különböző gombok, linkek, menüpontok és fülek, attól függően, hogy az internetes cég milyen szolgáltatásaira akarta felhívni a figyelmet, vagy melyeket zárt be. Az egyik legszomorúbb sors az önkéntesek által épített Open Directory Project (*dmoz.org*) linkgyűjteményre alapozott Google Directory-nak jutott. Kezdetben még kiemelt helyet foglalt el a keresőoldalon (1. ábra), majd az egyik fülre került át (2. ábra), onnan később átvándorolt a *more* menüpont alá, még később pedig az *even more* al-almenübe (3. ábra)^[6], míg végül teljesen eltűnt az oldalról és már csak célzott kereséssel lehetett rábukkanni.² Az emberi – részben könyvtárosi – munkával gondosan válogatott és kategorizált linkgyűjtemény vándorlása a Google oldalán, majd végleges eltűnése annak a folyamatnak tükrözi, ahogyan a humán intelligenciát átvette a gépi. Amint a keresőrendszer mögött levő, folyamatosan finomított algoritmus kellően releváns találatokat tudott adni bármilyen kérdésre, egyre kevesebben használták már a tematikus webkatalógust.

Screencast készítés

Webarchívumból képernyővideót négy lépésben lehet készíteni: először kell egy linkgyűjtemény az archivált oldalakról, majd le kell őket fotózni az archívumban (vagy letöltés után), a képeket be kell tölteni egy videoszerkesztőbe és végül fel kell venni a narrációt. Az első lépéshez nagyon hasznos az Internet Archive WayBack Machine Link Ripper nevű szolgáltatása, ami kigyűjti egy adott oldal mentéseinek URL címeit (napi, tiznapos, vagy havi intervallumokban), majd az Output fülre kattintva letölthetjük ezt a listát TXT, CSV vagy HTML formátumban. Ezeket az URL címeket egy screenshot készítő programba (pl. a Grab Them All nevű Firefox bővítménybe) betöltve elkészíthetők a képernyőfotók. Ügyeljünk arra, hogy kellő várakozási időt adjunk a szoftvernek, amíg az az archívumból sorra lekéri és megjeleníti az egyes

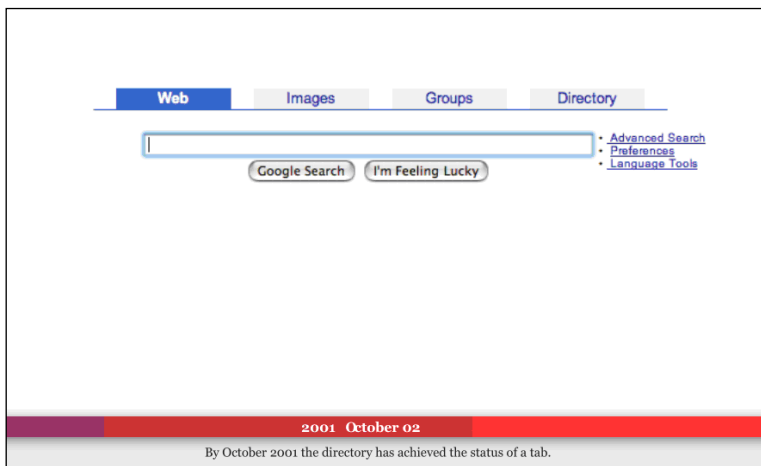
² 2011 nyarán véglegesen megszűnt a Google Directory, 2017 március 17-én pedig a *dmoz.org* is bezárt finanszírozás hiányában. (a referátum készítőjének megjegyzése)

mentéseket. A képek azután valamilyen videoszerkesztővel (pl. iMovie vagy Windows

Movie Maker) összefűzhetők és ezzel a hangalámondást is el tudjuk készíteni.



1. ábra



2. ábra



3. ábra

Hivatkozások

- [1] HERMENS, E.: The New York Times : a web historiography, Amsterdam, University of Amsterdam, 2011.
<https://vimeo.com/32319207>
- [2] ROGERS, R.: Digital methods, Cambridge, MIT Press. 2013.
- [3] LIVIO, M. – MATALY, J. – SCHUH, M.: TheKnot.com : a website historiography, Amsterdam, University of Amsterdam, 2012.
<https://www.youtube.com/watch?v=5cxVXJthETA>
- [4] UDELL, J.: Heavy metal Umlaut, 2005.
<http://jonudell.net/udell/2005-01-22-heavy-metal-umlaut-the-movie.html>
- [5] ROGERS, R. – Govcom.org: Google and the politics of tabs, Amsterdam, Govcom.org, 2008.
<https://movies.digitalmethods.net/google.html>
- [6] Az ábrák forrása: De GROOT, K: The Demise of the Directory : Web librarian work removed in Google, 2008.
http://www.govcom.org/publications/drafts/GCO_directoryfall.pdf

/ROGERS, Richard: Doing Web history with the Internet Archive: screencast documentaries. = Internet histories, 1. évf., 1-2. sz., 2017. p. 160–172.
<http://www.tandfonline.com/doi/full/10.1080/24701475.2017.1307542/>

(Drótos László)