

A webarchiválásról történeti megközelítésben

A tanulmánykötet esettanulmányok formájában az elsők között tesz kísérletet arra, hogy felvillantsa a webarchiválás történeti, illetve széles társadalomtudományi kontextusának számos fontos elemét. A szerkesztők előszava is kitér rá, hogy eddig inkább az volt a jellemző, hogy magáról a webarchiválási folyamatról, annak technikai részleteiről, a világháló archiválásához kapcsolódó kurátori tevékenységekről szóltak az összefoglalók. A szerkesztők az előszóban ez alkalommal is a legfrissebb szakirodalom segítségével vázolják fel a webarchiválás általánosabb kontextusát, eddigi történetének kronológiáját és főbb szereplőit. Jellemzik a főbb intézményeket, melyek e tevékenységeket végzik. Az *Internet Archive* úttörő szerepe mellett rámutatnak arra, hogy míg egyes országokban egyetlen vezető intézmény köré csoportosul e tevékenység (például Dániában), addig máshol intézményi koordináció tapasztalható világosan elkülönülő szerepkörökkel (pl. Franciaország, Nagy-Britannia). Rövid tájékoztatást kapunk arról, hogy milyen szoftverháttérrel történik az anyagok begyűjtése, s milyen módszerekkel lehet az eltárolt webes információk visszakeresését biztosítani (pl. az *Internet Archive* által fejlesztett *Wayback Machine* szoftverrel URL címekre kereshetünk, ezt egészíti ki a teljesszövegű index szolgáltatás, már amelyik gyűjteményben éppen elérhető).

Az általános összefüggésrendszer áttekintése után viszont hamar kiviláglik, hogy itt arról van elsősorban szó, hogy a digitális bölcsészeti, társadalomtudományi kutatások hatókörében milyen sokszínű módon lehet a webarchiválás által begyűjtött információkincset feldolgozni, értelmezni, újrahasznosítani. Voltaképpen a web a különféle társadalmi jelenségekre adott reflexiók feltárásának egy újfajta izgalmas felületét kínálja. A kötet különféle problémaköröket, kihívásokat, vagy éppen társadalmi csoportok vizsgálatát emeli ki az egyes tanulmányokon keresztül, s vázol fel az újfajta forrásanyagon alapuló következtetéseket. Kvantitatív, kvalitatív, illetve vegyes módszerekkel végzett kutatási projektekről is képet kapunk.

A világháló egyes szeleteinek összehasonlító vizsgálata

A kötet tanulmányait három nagy tematikai egységbe rendezték. Az első egységbe három publikáció tartozik, s itt elsősorban kvantitatív vizsgálatokról van szó, melyek a számítógépes világháló egy kiválasztott tartományát vizsgálják.

Meyer és munkatársai az első tanulmányban a brit webes szférára vonatkozóan tesznek megállapításokat, kiemelten oktatási domain (.ac.uk domain) vizsgálatában merülnek el. Rámutatnak arra, hogy a világháló fejlődésében az egyetemek mindig is úttörő, innovatív szerepet játszottak, hasonlóan a kormányzati szférához (.gov.uk domain), de eltérően a

kereskedelmi honlapokat üzemeltető szereplőktől (.co.uk domain). A tanulmány szerint hasonló következtetések adódnak a megfelelő amerikai domainek (.edu. org. com) esetében is. A szerzők vizsgálat alá vették az egyes webes tartományokon belüli kapcsolódási pontokat, s azok kapcsolódásait is. Az ilyenfajta vizsgálatokból az adott nemzet társadalmi életének változásaira vonatkozóan lehet következtetéseket levonni. Másfelől számos fajta összefüggés csak olyan összehasonlításokkal tárható fel, amikor nemcsak egy nagy nemzeti webtartomány alszeleteit vetjük össze egymással, hanem az egyes nemzeti webtartományok szerkezetét, rétegzettségét, illetve egyéb jellemzőit is összevetjük.

A második fejezetben, *Halke és munkatársai* tanulmányának is a brit nemzeti domain (.uk) a fő vizsgálati tárgya. A módszer azonban már vegyes (kvantitatív + kvalitatív), s a téma merőben eltérő. Azt vetik össze, hogy a *TripAdvisor* kereskedelmi weboldalon megjelenő webcímek összetételét miképpen lehet párhuzamba állítani az *Internet Archive* által begyűjtött brit utazási témájú anyaggal. A recenzens számára meglepő következtetésre jutnak: nagyon alacsony szintű a párhuzamosság a begyűjtött anyagok összetételében. A legnépszerűbb turistacélpontok webhelyeinek címei persze mindkét helyen megtalálhatók, a kevésbé ismert látványosságok felett azonban a *TripAdvisor* bizony elsiklik. Ebből természetes módon adódik a következtetés, hogy a kereskedelmi weboldal távról sem a reprezentatív bemutatás, tájékoztatás igényével gyűjti össze a webhelyek adatait. Amennyiben viszont az ilyen prominens weboldalak is ennyire megbízhatatlannak számítanak, az komoly kérdéseket vet fel a szerzők szerint a bölcsész, illetve társadalomtudományi kutatás számára, hogy miképpen lehet teljeskörű, illetve reprezentatív világhálós webhely címkészlethez jutni egy adott téma vizsgálata kapcsán.

Az első rész zárótanulmányát a szerkesztő *Niels Brügger és munkatársai* jegyzik. Vizsgálatuk fő tárgya a dán webarchívum (netarchive.dk), illetve az Internet Archive. Arra voltak kíváncsiak, hogy miképpen jelennek meg a domain nevek, indul a számuk radikális növekedésnek, illetve hogyan tűnnek el a világhálóról. Összehasonlítják a két webarchívumot a begyűjtött anyagok teljeskörűségének szempontjából, elemzik azokat a vizsgálati tényezőket, melyeket ennek kapcsán fel lehet használni. A vizsgálatok eredményeit összehasonlítva szerintük fontos következtetések vonhatók le egy adott archívum megbízhatóságának mértékéről. A másik érdekes vizsgálati területük, hogy a web mely szeletei és milyen arányban tűnnek el az exponenciális tartalomnövekedés közepette. A történészek érdeklődésére tarthat elsősorban számot annak regisztrálása, hogy mi veszett el a múltból, illetve a múlt mekkora szeletéről kell lemondanunk a világhálós anyag vizsgálatakor, mennyire torzíthatja a múltból alkotott képünket ez a hiány.

Média és Kormányzat

Ebben a részben vegyes kvalitatív, illetve kvantitatív szempontú kutatásokkal találkozhatunk. A negyedik fejezetben *Weber* az USA webterében megjelent helyi online sajtó elemzésével foglalkozik. Rámutat arra, hogy többször is drasztikus változások rendezték át a kínálatot országos szinten. A kínálat egyre szűkül, sokan elhullanak az idők során. Nem elégszik meg azonban a szerző azzal, hogy országos szinten elemez, hanem arra is kíváncsi, hogy az országos folyamatoknak milyen helyi szintű lecsapódásai vannak például New Jersey helyi online sajtójának vonatkozásában. Végül arra a következtetésre jut, hogy a változások országos trendjei hű lekötésre kerülnek helyi szinten is. A vizsgálat főleg a webes hivatkozás analízis módszerét használta fel, a másik módszer pedig az egyes online sajtótermékek jövedelmezőségének összehasonlító vizsgálata volt.

Az ötödik fejezetben *Cowls és Bright* kvantitatív vizsgálati módszerekkel jellemzi a BBC hírportálon megjelenő hírek egy szegmensének evolúcióját, méghozzá azokat a nemzetközi linkekét, amelyek kifelé vezetnek a portálról, vagy arra mutatnak rá. Divatos téma mostanában, hogy a hírpiac globalizálódásával összhangban a hírszolgáltatás mennyire tükrözi vissza a gazdagabb országok nézőpontját, illetve olyan országokét, melyeket erős katonai, gazdasági érdekek fűnek egy adott térséghez. A vizsgálatok arra utalnak, hogy a kevésbé békés országok kevesebb kifelé mutató linket kapnak a BBC News webhelyéről, még akkor is, ha a híryanagban bő tálalást kapnak, s az angol nyelvű országok felülreprezentáltak a kifelé mutató linkek terén. Miután a BBC News webhelye hírszolgáltatás terén globális referenciapontnak számít, a kutatók nagyon fontosnak tartják az ehhez hasonló vizsgálatokat. Bizonyításra vár, hogy a széles nemzetközi közönséghez eljutó BBC hírtartalmak valóban a globális egységről, egymásraultaltságról szóló feltevéseinket erősítik meg, vagy egyenlőtlen töredékes képet adnak, egyes régiókat felülreprezentálva, másokat pedig kevésbé láthatóvá téve.

Schafer tanulmánya arra irányul, hogy a francia kormány milyen módon kíván az eddigieknél is fokozottabban nyitni az állampolgárok felé a számítógépes világhálón, az interaktív kapcsolatteremtés növelésének szándékával. Az 1990-es években a francia politikusok hangzatos szövegekben dicsérték a digitális fejlődés vívmányait, míg a hálózati szolgáltatások a sajátos francia belső hálózat, a minitel köré összpontosultak. Nagyívű elképzelések születtek a közigazgatás hatékonyságának növeléséről, az állampolgárokhoz való közelítéséről, miközben valójában csupán néhány helyhatóság töredékes webes jelenlétéről beszélhettünk eleinte. Ezek megszületésükkor igen innovatív webhelyekké lettek, ám az idők során elavultak. Itt tehát egyfajta alulnézetből szemlélődő történetírásról van szó, mely a helyi szintű valóságot tárja fel az információs szupersztrádáról szóló központi politikai retorika mögött.

Politikai, társadalmi, kulturális kihívások webes lenyomatai

A harmadik nagy rész *Milligan* által írt tanulmánya talán a legkülönlegesebb. A *GeoCities* szolgáltatása gyűjtötte össze az egyik legkiterjedtebb és legösszetartóbb felhasználói közösséget a világháló hőskorában. Az emberek kialakíthatták saját webes otthonaikat a különféle felkínált szomszédság típusok hálózatában. A szerző elemzi, hogy miképpen mutatták be magukat az egyes felhasználók a közösség tagjai felé, illetve miképpen kommunikáltak egymás között. A *GeoCities* mögött álló földrajzi metafora mára már elvesztette relevanciáját, s gazdasági okok is hozzájárultak a szolgáltatás hanyatlásához. Civil közösség szerveződött azonban a szolgáltatáshoz kötődő tartalmak megőrzésére, a korábbi tagokként szereplő webarchivátorok révén, akik ennek kapcsán kapcsolatot találtak az offline világban a történeti műemlékek gondozását szervező civil szervezetek felé is. Megint rádöbbenhetünk arra e történet kapcsán, hogy a múltnak kis híján eltűnt szeletei mennyire fontosak lehetnek a történeti emlékezetünk szempontjából, a számítógépes világhálón is.

Ackland és Evans arról ad képet, hogy egy adott súlyos társadalmi vitakérdésnek milyen webes lenyomata tud lenni. Az abortuszkérdés ausztrál percepcióját tanulmányozták a szövegek, illetve a linkek analitikus vizsgálatával. Azok a kereskedelmi oldalak igen jelentős látogatottsággal bírnak, melyek abortuszt segítő gyógyszereket árulnak, mivel sokáig ennek törvényessége, társadalmi elfogadottsága meglehetősen kérdéses volt. Az analízis tárgyául szolgáló időszak elején az abortuszpárti és abortuszellenes vélemények nagy-

jából egyensúlyban voltak, aztán fokozatosan eltolódott az egyensúly az abortuszpárti álláspontok felé, ami a szerzők szerint a közhangulat változásáról is képet ad. A szerzők annak is megpróbálták utánajárni, hogy a *Google* megjelenése mennyire avatkozott bele az összképbe, a keresőszavak használatával miféle következtetéseket lehet levonni a témában. A big data alapú vizsgálatok azonban nagyon gyerekcipőben járnak, így a kutatás e szelétének tudományos relevanciája, illetve hasznossága bőven támaszt kétségeket a tudós társadalomban.

Webster a kilencedik fejezetben egy élénk társadalmi vitát kiváltó kérdés világhálós megjelenését állítja középpontba. A hagyományos iszlám jog, a *sharia* lehetséges nagy-britanniai alkalmazási módjait, az arra adott éles szélsőjobboldali reakciókat elemzi annak fényében, hogy miképpen nyilvánult meg ebben a kérdésben az anglikán egyház, illetve személyesen a Canterbury érsek. A webes hivatkozások analízise, illetve néhány, a témában érintett weboldal részletes vizsgálatának révén a kérdést tágabb kontextusba ágyazza a szerző, mely az anglikán egyháznak a brit társadalomban játszott szerepéhez kapcsolódik. Részletes bemutatásra kerülnek az ellentmondásos egyházi megnyilatkozások, illetve a szerző rámutat arra, hogy az érsek számos hivatalos állásfoglalása hosszú ideig nem volt elérhető online. A webes elemzés kiváló eszköze annak, hogy az ilyen éles vitakérdésekben feltűnő különféle egyházon belüli áramlatokhoz, frakciókhoz kötődő véleményeket fel lehessen tárni. Az eddiginél részletesebb képet kaphatunk az iszlám és más vallások viszonyát befolyásoló különféle szellemi tényezőkről.

A tizedik fejezetnek is az iszlám a főszereplője, ám egy másik kontinensen, s egészen más nézőpontból. *Dougherty* az iszlám punk kultúra észak-amerikai webes megjelenését, fejlődését vizsgálja. Ez a szubkultúra mára lényegében elhalt, így egy teljes összkép rajzolható fel annak megjelenéséről, virágzásáról, illetve bukásáról is. Az ilyen típusú vizsgálatok azért fontosak, mert rámutatnak, hogy a különféle időszakokban milyen kulturális normák hatják át az amerikai muszlim fiatalok viselkedését. Itt egy olyan zenei stílus megjelenéséről van szó, mely egyszerre reagált újszerűen az iszlám hagyományokra, illetve vitt be új színt a punk kultúra keretei közé. Tágabb kontextusban felvetődik az a kérdés is, hogy a nonkonformizmusnak mekkora tere lehetett annak az élénk társadalmi vitának a kapcsán, ami az iszlám Amerikában játszott szerepéről zajlott. Külön érdekes módszertani kérdés, hogy ebben az esetben az elsődleges forrásanyag jó részét nem az *Internet Archive*-ban lementett oldalak jelentik, hanem például a *Reddit* szolgáltatáshoz hasonló moderált beszélgetőcsoportok, blogok anyagai, a wikipédia bejegyzések, illetve a napi híryanag archivált szeletei.

Az utolsó tizenegyedik fejezet visszakapcsolódik Nagy-Britanniához, a brit nemzeti névtér nagymennyiségű adattömegének bölcsészettudományi vizsgálatához kötődő BUDDAH (*Big UK domain data for the Humanities*) projekt kapcsán. Itt voltaképpen számos olyan kísérleti tanulmány bemutatásáról van szó, melyek a brit társadalmi élet különféle szeleteit vizsgálják a webarchívumban elmentett adatanyagon keresztül. Ilyen téma például a brit hadseregről szóló weben keresztüli kép ábrázolása, a beat nemzedék költői életműveinek megjelenése, a szélsőjobboldali politikai csoportok webes tevékenysége, vagy éppen a fogyatékosokat támogató civil tevékenység világhálós lenyomata. A fejezet betekintést enged azokba a kihívásokba, amelyek a hatalmas adatanyagból való válogatásra, illetve annak bölcsészeti szempontú elemzésére vonatkozik. A legtöbb tanulmány szerzője ráadásul nem tudományos szakember. Ennek ellenére tevékenységük révén igen értékes betekintést nyerhetünk a társadalom különféle szegmenseinek működésébe, élet-

módjába, kulturális és történelmi fejlődési irányvonalak felrajzolására, dokumentálására is mód nyílik. A fejezet különösen azok számára ajánlott kezdőpontként, akik hasonló kutatási témákkal kívánnak foglalkozni, ahol a fő forrást magának a learatott webtérnek az anyaga jelenti.

Nagyon jól kapcsolódik a tizenegyedik fejezethez a zárszó, melynek szerzője *Jane Winters*, a fentebb említett BUDDAH projekt vezetője, az University of London digitális bölcsészettel foglalkozó professzora. Körüljárja azokat a tudományos diskurzusokat, melyek a nagymennyiségű adatok vizsgálatának, illetve az azokhoz kapcsolódó kvantitatív kutatási eljárásoknak a történettudományi hasznáról értekeznek. Rámutat arra, hogy várható volt, hogy éles diskurzusok zajlanak majd a témában, különösen a bölcsészetek területén, ahol szerinte kevés a megfogható, konkrét tény, s a tudósok egymásrautaltsága is kevésbé jellemző, mint más kutatási területeken. Jellemző azonban az új elméletek, vizsgálati módszerek berobbanása, majd fokozatos megszelídülése, hozzáidomulása a meglévő keretekhez.

Epilógus

Ez a tanulmánykötet igazi úttörő munka, olyan területre irányítja a figyelmünket, amely lebilincselően és az újdonság erejével hat, s számos tudományterülethez kötődő készségeket és képességeket kell összegezni a műveléséhez. Ki kell nyerni az adatokat, sok esetben megfelelő kvantitatív eszköztár segítségével hatalmas adathalmazokból, majd azokat lehet további elemzések tárgyává tenni. Számos példát látunk arra, hogy még csupán a kezdeti lépéseket megtéve milyen fontos új kutatási témák, összefüggérendszer tárnak fel a társadalom működéséhez kötődő webes forrásanyagok vizsgálatakor. Reméljük, hogy a jövőben a webarchiválással foglalkozó technikai szakemberek, illetve a digitális bölcsészet művelői hatékonyan együtt tudnak működni egymással, s a begyűjtött anyag társadalomtudományi, illetve bölcsészeti szempontú hasznosításának újabb és újabb útjai tárulhatnak fel.

Niels Brügger – Ralph Schroeder (eds.): *The Web as History: Using Web Archives to Understand the Past and the Present*. London, UCL Press, 2017.

Németh Márton

