



A webarchívumokra irányuló kutatási tevékenységek a WARCnet projekt tükrében – Az első két év

NÉMETH Márton

Az alábbi áttekintés a „404 Not Found – Ki őrzi meg az internetet?” című online konferencián elhangzott előadás bővített változata, kiegészítve a 2022. év első felének történéseivel.

Bevezetés

A következőkben bemutatom a webarchívumok tudományos célú felhasználására irányuló *WARCnet* projekt főbb célkitűzéseit, az eddig lezajlott munkaértekezletek tanulságait, továbbá szót ejtek az egyes munkacsoportok tevékenységeiről, céljairól, csakúgy, mint a Covid19-járvánnyal kapcsolatos nemzeti webarchiválási gyakorlatokat ismertető publikációkról.

A tárgyalt témakörökben a projekt honlapján (<https://warcnet.eu>) találhatóak részletesebb információk.

A WARCnet projekt célja és felépítése

Az elgondolás *Niels Brügger*nek, az Aarhusi Egyetem professzorának a fejében született meg, aki a Dán Kutatási Tanácshoz nyújtott be pályázatot egy olyan kutatási hálózatra megalkotásáról, amely segít értelmezni a nemzeti és nemzetközi felső szintű webdoménokból archivált anyag felépítését, rétegzettségét, történetét, tartalmi összefüggéseit (ilyenek például a *.hu* vagy a *.com* tartományok). Számos fontos társadalmi krízis, konfliktus, vitatéma tanulmányozásához elengedhetetlenek a webes források, így a hálózat a digitális kulturális örökség fontos szeletét jelentő globális társadalmi események európai szintű webes lenyomatát is igyekszik feltárni.

A sikeres pályázatot követően 2020-ban elindult projekt fő célja, hogy lehetővé tegye a webarchívumokban tárolt anyagokban rejlő összefüggések és a web történetére vonatkozó kutatások sok szempontú vizsgálatát (pl. kommunikáció- és médiatudományi, történettudományi, digitális bölcsészeti nézőpontokból). Egy webarchívum a nagy mennyiségű adatok tárházaként értelmezhető, melyre adatbányászati és adatvizualizációs tevékenységek építhetők. Tekinthetünk rá úgy is, mint modern digitális hiteles helyre, melynek forrásai jogviták esetén bizonyítékként szolgálhatnak. Sokszínű felhasználási lehetőségek nyílnak tehát, számos interdiszciplináris kapcsolódási ponttal. Az 1990-es évektől a számítógépes világháló fokozatosan a társadalmi kommunikáció, a tudásátadás, illetve tágabb értelemben a közélet megkerülhetetlen médiumává lett, természetéből adódóan a nemzetközi térben. Az itt zajló diskurzusok szinte egy pillanat alatt a múlt részévé válnak; ezek feltárása, elemzése nagyon fontos kutatási tevékenységként jelenhet meg. Az *Internet Archive*, illetve a nemzeti könyvtárak és levéltárak évtizedek óta gyűjtik a webes múlt fontos emlékeit. A webarchívumok kutatása azonban számos esetben speciális kompetenciákat igényel, és messze nem olyan triviális, mint a hagyományos levéltári vagy könyvtári kutatási tevékenység. A webarchívumok fontosságát általában alábecsülik kutatási szempontból, és miután a nemzeti szintű webes kutatási háttér mindenütt nagyon korlátozott mértékű Európában, ezért felvetődött egy olyan hálózat létrehozása, melynek segítségével egymásra találhatnak a szakemberek, és az erőforrások hatékony felhasználásával közös és magas színvonalú munkát folytathatnak. Így talán sikerül elérni azt a kritikus tömeget, melynek révén a webarchívumokra irányuló kutatások széles körű létjogosultságot nyernek.

A WARCnet projekt közös platformot nyújt a webarchívumokban tárolt anyag kutatóinak és a gyűjtemények működtetésében főszerepet játszó közgyűjteményi szakembereknek és informatikusoknak. A tevékenységeket hárman irányítják: Niels Brügger professzor, kommunikációkutató, *Valerie Schafer*, a Luxemburgi Egyetem professzora, aki jelenkori európai történelemmel foglalkozik, és *Jane Winters*, a Londoni Egyetem (*University of London*) digitális bölcsész professzora. A csapatban dolgoznak egyéni kutatók Dániából, Hollandiából, Izraelből, Luxemburgból, Németországból, Nagy-Britanniából, emellett tagjai a nemzeti könyvtári webarchívumok (holland, belga, dán, angol, luxemburgi, francia, valamint

az Országos Széchényi Könyvtár) illetve a nemzetközi internet-archiválásért felelős konzorcium (*International Internet Preservation Consortium*, IIPC) is. A finanszírozási háttérrel a Dán Kutatási Tanács biztosítja, a koordinátori szerepet a webtörténetírás atyjának is tekinthető Niels Brügger professzor látja el. Sajnos, a koronavírus-járvány következtében alig sikerült személyes részvételen alapuló találkozót tartani, melyekből eredetileg évente legalább kettőt terveztek a projekt időtartama alatt (döntően a Londoni, a Luxemburgi és a dán Aarhusi Egyetem szervezésében). Fontos cél lett volna az egymás intézményi, illetve kutatási tevékenységének jobb megismerését szolgáló, néhány napos tanulmányutak megszervezése is, amely a mostani helyzetben szintén háttérbe szorult. A projekt időtartamát azonban sikerült kitolni az eredetileg tervezett 2020–2021-es periódustól egészen 2023 tavaszáig. A munka természetesen a járványügyi korlátozások ideje alatt is zajlott, számos tevékenységet az online térben is hatékonyan lehetett végezni.

A projekt működésének három legfontosabb fóruma:

- az operatív irányítást végző szűk testület,
- a plenáris munkaértekezletek, illetve
- a szekciókban (ld. részletesen lentebb) folyó tevékenységek.

A *projektirányító bizottságot* a hálózat tagjai választották meg az első értekezleten 2020. május elején. A bizottság hattagú, a három, a hálózatot irányító professzor mellett egy fiatal kutató (*Sophie Gebeil*, Franciaország), egy informatikus (*Ulrich Karstorf Have*, Dánia), illetve egy webarchívumban dolgozó szakember (*Kees Teszelszky*, Hollandia) is bekerült a testületbe. A fő projektkoordinátor, Niels Brügger munkáját ezenfelül két diákasszisztens segíti az Aarhusi Egyetemenről. A rövid tanulmányutakra vonatkozó kérvények elbírálásával külön bizottság foglalkozik.

A projekt első szakaszában a tevékenységek az egyes munkacsoportok keretein belül, illetve a munkaértekezleteken zajlottak. Nagy hangsúly helyeződött a *WARCnet Papers* sorozatban megjelent publikációk előkészítésére is. A projekt második szakaszában az elért eredmények összegzése, illetve azok könyv formájában történő megjelentetése áll a középpontban. A *munkaértekezleteken* az általános projektmenedzsment-teendők elvégzése mellett a szekciók beszámolnak a folyamatban lévő tevékenységekről, az elért eredményekről. Ezek az alkalmak teret adnak átfogóbb, a webarchiválást esetleg csak érintő témáknak és olyan előadásoknak is, melyek új kutatási irányo-

kat alapozhatnak meg. Előzetes regisztrációt követően minden értekezleten ingyenesen részt vehetnek (tanácskozási joggal) nem projektag érdeklődők is. Közülük sokan – mint e sorok írója is, aki a második (online) értekezleten kapcsolódott be a munkába – a projekt tagjává is válhattak. A tagok számára a személyes értekezleteken való részvétel ingyenes, a szállásköltség is biztosítva van. A 2021 őszen Aarhusban tartott hibrid értekezletet követően 2022-ben, a járvány múltával egy nyári, londoni találkozóra került sor. A záróértekezletet – amelynek időpontját még a projekt időtartamának meghosszabbítása előtt jelölték ki – idén ősszel, Aarhusban rendezik majd meg. A későbbiekben az eredményeket összegző köteten fognak intenzíven dolgozni a projekt tagjai, online konzultációk keretében.

A munkacsoportok ismertetése

A WARCnet hat szekciója különféle szakmai témák köré szerveződik, munkájuk pedig szervesen kapcsolódik egymáshoz. Az *egyes számú szekció* egyrészt nemzeti és nemzetközi doméneket (webtereket) vizsgál (webhelyek száma, mérete, multimédia elemek használata, a hiperlinkekkel összefogott hálózat sajátosságai), másrészt a felhasználókra összpontosít (milyen jellemzőik vannak, és milyen webhelyeket látogatnak), harmadrészt pedig a globális közösségi hálózatok működését és társadalmi hatásait elemzi. Az első fő cél a fenti témakörökre vonatkozó szakirodalmi bázis összegyűjtése volt. Ezt követően jelenleg egy olyan útmutató megalkotásán dolgozunk, amely segíthet meghatározni a nemzeti webterek fogalmát, valamint a webhelyek leírásának módszertanát, a különböző webarchívumok összehasonlíthatóságának szempontjait és olyan kisebb korpuszok megalkotásának eszközeit, melyek modellezni tudják egy-egy nemzeti webtér működését.

A *második munkacsoport* a különféle kiemelt események webes megjelenésére és az ebből adódó kutatási, elemzési lehetőségekre fókuszál. Két fő kategória alakult ki: az egyik az előre nem látható történések csoportja (terrortámadások, természeti katasztrófák, menekültválság, Covid19-járvány), a másik az előre tervezhető kiemelt nemzetközi események kategóriája (pl. európai parlamenti választások, olimpiák). Az első lépés itt is a szakirodalmi bázis felmérése volt. Az első és a második munkacsoport tevékenysége szorosan összefügg, mivel az események webes megjelenése a nemzeti webterek elemzése során is értelmezhető. A vizsgálatok tárgyául szolgáló anya-

got döntően az IIPC keretében állítják elő, ahol szintén külön munkacsoport gondozza az eseményalapú mentéseket. A létrejövő nemzetközi gyűjteményekre támaszkodva különféle adatbányászati, adatelemzési, adatvizualizációs kutatásokat terveznek. A csoport munkájában az eddigi legfontosabb hozzáadott érték tematikus interjúk készítése volt a Covid19-járvány kapcsán végzett webarchiválási tevékenységekről az érintett közgyűjtemények képviselőivel. (Erre a későbbiekben külön fejezetben térek ki.)

A *harmadik szekció* a webarchiválással összefüggő módszertani teendőkre összpontosít. Utaltunk már arra, hogy a webarchívumban tárolt anyag jellegében is eltér más digitális könyvtári gyűjteményektől, ebből adódóan speciális eszközök, módszerek keltenek a feltárásához és tudományos vizsgálatához. A munkacsoport fő profilja tehát a rendelkezésre álló gyűjteményfejlesztési, gyűjteménykezelési és kutatási módszertani háttér és eszköztár felmérése, elemzése, tesztelése, illetve a különféle igényeknek megfelelő testreszabási lehetőségek meghatározása. A másik fő cél – a meglévő bázisra alapozva – igény szerinti új módszerek, eszközök, eljárások fejlesztése annak érdekében, hogy támogassák a webarchiválás tudományos célú hasznosítását, valamint az ezzel összefüggő gyűjteménykezelési és -fejlesztési tevékenységeket. Az ennek eredményeként születő nyílt forráskódú szoftvereszközöket a *GitHub*on, illetve egyéb hasonló platformokon kívánják közzétenni, az új eszközöket és a kapcsolódó módszertani elképzeléseket pedig workshopok és *hackathon*ok segítségével próbálják népszerűsíteni. A harmadik munkacsoport tehát egyfajta gyakorlati háttérrel nyújt az első két csoportban folyó tevékenységek hatékonyabb elvégzéséhez.

A *negyedik munkacsoport* a webarchiválás során létrejövő és a kutatás során használható adattömeg kezelésére koncentrál. Ez a kutatási adatok kezelésének egy speciális ága, amelyhez a közgyűjteményi szakembereknek és a kutatóknak együtt kell működniük az adattömegek előállítására és elemzésére terén. Ez a tevékenység még nemzeti keretek között is gyerekcipőben jár, nemzetközi szinten pedig eddig nem is nagyon foglalkozott vele semmilyen szereplő. A munkacsoport egyik kiemelt feladata egy olyan átfogó felmérés elkészítése, amely áttekinti az adatok kezelésének és megosztásának jogi és technikai követelményeit az egyes nemzeti webarchívumi környezetekben, választ keresve arra, hogy a kutatók nemzetközi szinten hogyan tudnának hozzáférni az egyes gyűjteményekben tárolt adatokhoz. A legálisan

beszerezhető adatkészletekből összeálló korpusz segítségével pedig egy kísérleti projektet szeretnének futtatni, amely az adatkészlet kutatásában rejlő lehetőségeket demonstrálná. Az egyes európai nemzeti könyvtárak számára kulcsfontosságú az az összehasonlító-elemző munka, melyet a projekt keretében végeznek, hiszen így egy egységes platformon átláthatóvá válnak egymás gyakorlatai, és európai szinten is javaslatokat lehet megfogalmazni a hatékonyabb adatkezelés jogi és technikai kereteinek biztosítására. A negyedik munkacsoportban zajló munka szervesen kapcsolódik az első kettőben folyó kutatási-elemzési tevékenységhez, emellett a harmadik munkacsoport kutatási-fejlesztési eredményei is hozzá tudnak járulni az adatok kezelésének előmozdításához.

A WARCnet projekt eredetileg a fentebb vázolt négy szekciót foglalta magában, időközben azonban a felmerült igényekre válaszul két újabb csoport is létrejött.

Az *ötödik szekció* voltaképpen a negyedik által lefedett kutatásiadat-kezelés egy speciális összetevőjével foglalkozik: a webarchívumokból származó kutatási adatok megosztásával. Jelenleg nem állnak rendelkezésre szabványos megoldások arra, miként lehetne a különféle adatkészleteket a kutatók számára megfelelő módon megosztani szabványos alkalmazásfejlesztési felületeken (API-kon) keresztül, közös adatkezelést valósítva meg máshonnan származó forrásokkal. Másképpen megfogalmazva: a webarchívumokból származó adatokat szemantikusweb-kompatibilis formában kellene publikálni, a nyílt adatkészleteket pedig a kapcsolt nyílt adatok (*linked open data*) univerzumának részévé kellene tenni. Egy közös adatkezelési nyelvre van tehát szükség, amely egyaránt szolgálja a közgyűjtemények és a kutatói közösség igényeit. Ennek érdekében első lépésként fel kell mérni a webarchívumokban jelenleg alkalmazott adatformátumokat és azok felhasználási lehetőségeit. Második lépésként el kell készíteni egy egységes leírást a webarchiváló intézmények és a kutatók által használható adatformátumokról, azok alkalmazási lehetőségeinek részletezésével együtt (*Web Archiving Codebook*). Nyitott kérdés, hogy felállítható-e egy mindenki által használható egységes adatformátum, vagy a különféle, már rendelkezésre álló formátumok interoperabilitására kellene helyezni a hangsúlyt. Az elkészült leírás alkalmazhatóságát a gyakorlatban is tesztelni kell, az első két munkacsoportban zajló kutatási tevékenység keretei között.

A *hatodik munkacsoport* arra összpontosít, hogy biztosítható legyen a kutatási hálózat hosszabb távon

történő fenntarthatósága a WARCnet projekt időtartamának vége után is. Ennek érdekében felméri a különböző európai uniós pályázati lehetőségeket, illetve az egyes nemzeti és nemzetközi kutatási hálózatokba történő bekapcsolódás formáit.

A projekt keretében zajló munka 2021 végétől 2022 közepéig

A 2021 novemberében lezajlott munkaértekezleten érdekes *keynote* előadást hallgathattunk meg a távol-keleti internet történetéről, fejlődési trendjeiről. Ennek kapcsán felvetődött a kérdés, hogy az összehasonlító vizsgálatok milyen formában, illetve milyen keretek között terjeszthetők ki globális méretekre. Önálló kerekasztal-beszélgetés foglalkozott azzal, hogy miként lehet a digitális megőrzésről gondoskodni azokban az országokban, ahol még nem létezik webarchívum, vagy ahol most tartanak az első lépéseknél ezen a téren.

A negyedik szekcióban zajló munka során sok szó esett a szerzői jogi, illetve adatvédelmi dilemmákról és jó gyakorlatokról is a webarchívumokhoz való hozzáféréssel összefüggésben.

Az első, a harmadik és az ötödik munkacsoportban fontos kérdésként vetődött fel, hogy miként lehetne becsatornázni a kutatói igényeket a webarchívumok szolgáltatási környezetébe, illetve hogyan lehet azokat egyáltalán megfogalmazni, továbbá milyen adatformátum és szoftveres eszköztár szükséges a kutatókhoz. Hogyan lehetne megteremteni egyfajta átjárható, közös alapokkal bíró szolgáltatási és adatformátum-környezetet a kutatók számára, amely egységes bázist biztosít az egyes nemzeti webarchívumokban zajló vizsgálódásokhoz? Az is szóba került, hogy a kutatási igények megfogalmazásával miképpen lehet támogatni az új webarchívumok indítását (elsősorban kutatóintézeti, felsőoktatási környezetben), ahol a gyűjtőkört és az archiválási módszereket is a specifikus kutatási igényekhez lehet igazítani. A WARCnet hálózat tagjai szerint általában nagyobb figyelmet kellene fordítani a webarchiválás különféle vonatkozásainak a felsőoktatás kereteibe történő beágyazására, oktatási és kutatási szempontból egyaránt.

A munkaértekezleten bemutatták az eseményalapú adatelemzési munka első eredményeit is a Covid19-gyűjteménnyel összefüggésben. Elkezdték leválogatni, hogy a felső szintű domének szerinti eloszlásban honnan és mennyi adatot sikerült begyűjteni, továbbá vizsgálják, hogy az egyes nemzeti webterekről archivált anyagoknak milyen sajátosságaik vannak,

milyen eltéréseket lehet tapasztalni az összetételt és a tartalmi jellemzőket tekintve, továbbá melyek voltak azok a főbb kulcsszavak, témakörök, melyek az archivált webtartalomban megjelentek, végezetül milyen összetételű az anyag általában, illetve a nemzeti webterek szintjén a webhelyek típusa szerint. A tartalomelemzésben rejlő lehetőségek érzékeltetésére külön *datathont* is szerveztek.

Niels Brügger és csapata a Dán Nemzeti Könyvtár webarchívumának 2006–2016 közötti anyagára alapozva olyan tartalomelemzésbe fogott, amelynek során felmérik, hogy melyek voltak a főbb témák a dán webterben, és ezek hogyan változtak az idők során. Itt több milliónyi kulcsszó leválogatásáról és elemzéséről van szó, amihez a megfelelő erejű informatikai infrastruktúrát csak a közelmúltban sikerült felállítani.

A harmadik szekció tagjai egy 28 kérdésből álló kérdőívet szerkesztettek, elsősorban kutatók számára, bár végül a válaszadók kétharmada közgyűteményi szakember volt. Arra voltak kíváncsiak, hogy milyen eszköztárral és szakmai kompetenciákkal bírnak a webarchiválás terén, és milyen igényeket tudnának megfogalmazni a területhez kötődően. A részletes eredmények későbbre várhatók, sajnos a 2022. júniusi londoni értekezleten sem tették még közzé azokat. Egy külön előadás foglalkozott a webarchiválás, illetve az arra irányuló kutatások etikai dimenziójával. Fontos kérdés, hogy már a gyűjtőkör meghatározásakor ki kell-e szűrni bizonyos típusú (pl. pornográf, gyűlöletkeltő) tartalmakat, vagy próbáljuk ezeket – akár célzott tematikus gyűjteményekben – közreadni a kutatók számára. Két fő szempont áll szemben egymással. Az egyik, hogy etikailag megkérdőjelezhető, illetve illegális tartalmakat begyűjthet-e legálisan egy webarchívum. Másrészt fontos kutatói alapállás, hogy előbb mindent be kell gyűjteni, hogy utána fel lehessen tenni a megfelelő tudományos kérdéseket. Ha eleve kizárjuk ezeket a típusú anyagokat, a jövőben sem lesznek már vizsgálhatók.

A munkaértekezlet színes mozzanata volt a Kees Teszelszky által szervezett *Twitter*-kihívás, amelyben az egyes résztvevők által gondozott vagy kutatót archivált tartalmakat lehetett közzétenni különböző témákban, majd ezek közül a zsűri győztest hirdetett, jelképes nyereményekért.

A projekt lényeges eredménye a *WARCnet Papers* sorozat, mely a kutatómunka során született eredmények gyors publikálására ad lehetőséget. A sorozat első darabjai 2020 tavaszán jelentek meg, és elsősorban a webarchiválással kapcsolatos kutatások pers-

pektíváinak felvillantásával foglalkoztak. 2021-ben Niels Brügger összefoglalót készített a projekt első évének munkájáról. A publikációk magvát a Covid19-járvánnyal kapcsolatos webarchiválási tevékenységet feltáró interjúk adják.

A 2022. június 13–15. között zajlott munkaértekezletre a Londoni Egyetem patinás főépületében, a *Senate House*-ban került sor. Az értekezleten a fő hangsúly a WARCnet keretében zajló különféle kutatási irányokat, eredményeket összegző kötet fejezetszerkezetének egyeztetésén, a főbb témakörök bemutatásán volt. Jómagam két fejezet összeállításában veszek részt. Az egyikben *Kalcsó Gyula* kollégámmal együtt az orosz–ukrán háború kapcsán létrehozott gyűjtemények szövegkorpuszain folytatott szövegelemzési-adatvizualizációs kutatási tevékenységekről számolunk majd be, ennek alapjait rövid előadás keretében az értekezleten is ismertettem. A másik téma a webarchiválás megjelenése különféle felsőoktatási képzési programok keretei között. Ezzel összefüggésben a hálózat tagjainak segítségével szeretném felmérni a meglévő jó gyakorlatokat, jövőbeli lehetőségeket.

Rövid áttekintés hangzott el arról, hogy az ötödik munkacsoport keretében összeállítás alatt áll egy szójegyzék, melynek legfőbb célja, hogy a webarchiválással kapcsolatos kifejezéseknek szabatos meghatározást adjon, illetve biztosítsa azt, hogy a kutatói közösség tagjai ugyanazt értsék a különböző fogalmak alatt.

A bevezető jellegű *keynote* előadások azzal foglalkoztak, hogy az irodalomtudomány területén az archivált webes tartalomnak milyen szeleteit és miképpen lehet kutatási célból használni. *Sandra Folie*, a Bécsi Egyetem oktatója az ún. *chick-lit* mint irodalmi műfaj, illetve az arról szóló webes diskurzus útját követte végig az *Internet Archive* által elérhető forrásokból a web kezdeteiig visszanyúlva.

Saskia Huc-Hepher, a Westminsteri Egyetem professzora a brit webarchívumot tanulmányozva kutat olyan értelmezések után, melyek speciális társadalmi csoportok irodalmi tevékenységét és az arról zajló diskurzust tárják fel az archívumból leszűrt mikrogyűjtemények révén. Ily módon rejtett összefüggésekre hívja fel a figyelmet például a londoni franciaajkú nők blogjaival kapcsolatban, melyek támogatják az egyes társadalmi csoportokról szóló tudásunkat, s teljesebb képet kapunk az ebben a társadalmi szegmensben zajló diskurzusokról is.

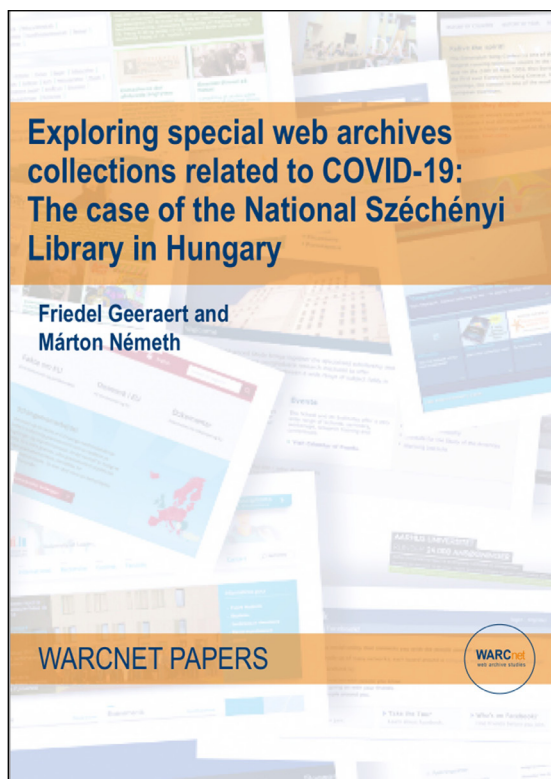
Az értekezlet különlegessége volt a *Tools Bazar* címmel megrendezett szoftverbemutató, melynek révén számos PhD-hallgató, egyetemi szoftverfejlesztő-kutató mutatott be fejlesztés alatt álló, webarchiválást segítő szoftvereket. (Pl. egy webarchívumok lekérdezését és az archívumban való kutatást segítő megoldás, egy Twitter-archiváló és -elemző szoftver, egy archivált weboldalakból strukturált metaadatokat kinyerni képes megoldás, illetve egy automatizált böngészést és archiválást segítő program.)

A Covid19-járvánnyal kapcsolatos webarchiválási tevékenység nemzeti gyakorlatainak összehasonlító feltárása interjúk révén

A projekt első szakaszának egyik leglátványosabb eredménye annak a publikációs sorozatnak az elkészítése volt a *WARCnet Papers* keretein belül, amely az egyes nemzeti webarchívumok – köztük a magyar – munkatársaival készített, azonos szerkezetű interjúk segítségével tárja fel a Covid19-járvánnyal kapcsolatos webarchiválási tevékenységek részleteit, illetve teszi egymással összehasonlíthatóvá a tematikus archiválási gyakorlatokat (1. ábra). A koncepció a második munkacsoportban született meg egy kérdéssor összeállításával, majd kijelölték az interjúkészítőket, és ezt követően 2020 elejétől elkezdték megkeresni az egyes nemzeti könyvtárakat. Mára a Francia Audiovizuális Archivum, illetve az amerikai Kongresszusi Könyvtár gyakorlatát és tapasztalatait is feltárták az interjúknak köszönhetően. A későbbiekben az anyag kutatási korpuszként is szolgálhat például ahhoz, hogy össze lehessen hasonlítani a szóhasználatot, vagy szógyakorisági vizsgálatokat lehessen végezni az interjúkon. Ez egy igen úttörő vállalkozás, hiszen bár már régebb óta zajlanak közös tematikus gyűjtések az IIPC keretein belül, ezek hátterét, az egyes nemzeti gyakorlatok számbavételét és összehasonlítását korábban még nem végezték el.

Zárszó

Jelen összefoglaló áttekintést nyújt a WARCnet projekt felépítéséről, illetve az első két év tevékenységeiről és eredményeiről. A munka több irányban, nagy lendülettel zajlik tovább. Reményeim szerint a projekt lezárultával mérleget lehet majd vonni arról, miképpen sikerült nemzetközi szinten intézményesíteni a webarchívumokra irányuló, illetve az archivált webes anyagot forrásként használó kutatásokat, és teljes összkép rajzolható majd fel a projekt eredményeiről.



1. ábra

A magyar webarchiválás gyakorlatáról készült interjúk kötetének címlapja

Felhasznált irodalom

BRÜGGER, Niels. The WARCnet network: The first year [elektronikus dok.]. Aarhus: WARCnet, 2021. 12 p. (WARCnet Papers, ISSN 2597-0615). ISBN 978-87-94108-03-4. Hozzáférhető: https://cc.au.dk/fileadmin/user_upload/WARCnet/Bru_gger_The_first_year.pdf [Megtekintve: 2022.07.25.]

GEERAERT, Friedel – NÉMETH Márton. Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary [elektronikus dok.]. Aarhus: WARCnet, 2020. 11 p. (WARCnet Papers, ISSN 2597-0615). ISBN 978-87-94108-03-4. Hozzáférhető: https://cc.au.dk/fileadmin/user_upload/WARCnet/Geeraert_et_al_COVID-19_Hungary.pdf [Megtekintve: 2022.07.25.]

NÉMETH Márton. Webarchívum mint a tudományos kutatások tárgya [online]. == Tudományos és Műszaki Tájékoztatás, 67. (2020) 12., p. 757–765. ISSN 1586-2984. Hozzáférhető: <https://tmt.omikk.bme.hu/tmt/article/view/12804> [Megtekintve: 2022.07.25.]

(Beérkezett: 2022. július 27.)