

Az OSZK Webarchívum 2021 novemberi hírei

Címlisták bővítése

Folytatódott a természet- és műszaki tudományos weboldalak címlistájának gyűjtése, novemberben főként a környezetvédelmi, a bányászati, a kohászati és az energetikai témakörökkel. Az elektronikus periodikák nyilvántartásába most csak 10 új tételt vettünk fel, viszont összehasonlítottuk a nemzetközi ISSN portálról (<https://portal.issn.org>) lekérdezhető magyar online kiadványok listáját a saját nyilvántartásunkkal. Az így talált, számunkra még ismeretlen webhelyek ellenőrzését és nyilvántartásba vételét a következő hetekben kezdjük el.

Technikai ügyek

Az archivált webtartalmak megjelenítéséhez használt PyWb szoftver új verziójában lehetőség van a felhasználói felület nyelvének megváltoztatására. Már korábban elkészítettük a magyar nyelvű fordítást, de technikai gondok miatt csak most sikerült beállítani a nyelvváltást és egyelőre csupán a nem nyilvános archívumnál, mert a publikus szerveren a rendszerkörnyezetet is frissíteni kell előbb hozzá. Ugyancsak részleges sikert értünk el a Web Curator Tool 3-as verziójának beüzemelésénél: az aratási folyamat most már elindul, de utána az indexelés nem fut le. A hiba okának kiderítésén az informatikus kolléga dolgozik. Többszöri próbálkozás után végül megtaláltuk a megoldást a „Chess - Sakklista” részben nem publikus archívumának megőrzésére. A lista gazdája egy csomagban küldte be a teljes állományt, amit sikerült úgy WARC formátumra konvertálnunk, hogy az eredeti URL címek alapján böngészhető a levelezőcsoport 2001 és 2021 közötti archívuma. Ugyancsak csomagként kaptuk meg az OSZK Facebook és Twitter fiókjának kiexportált tartalmát, amit egyelőre csak raktározni tudunk, mert ezekből nem lehet WARC fájlokat generálni.

pywb részgyűjtemény keresőoldala Nyelv: hu / en

A pywb részgyűjtemény(ek) URL szerinti keresése:

Eredmények megnyitása új ablakban

Illesztés típusa:

Dátum/Idő intervallum

Tól: Ig:

Szűrés

Szempont:

Illeszkedés:

Kifejezés:

- Tartalmazza
- Teljes egyezés
- Reguláris kifejezés szerinti egyezés
- Nem tartalmazza
- Nem egyezik
- Nem ezzel kezdődik

• Nincs szűrés feltétel

A Python-alapú Wayback megjelenítő magyarított felülete

Ismeretterjesztés és oktatás

A Petőfi Irodalmi Múzeum és az Országos Széchényi Könyvtár november 5-én „Born digital” címmel rendezett egynapos tanácskozást a digitálisan született tartalmak közgyűjteményi megőrzéséről. A rendezvényen mi is tartottunk egy előadást a webarchiválásról (https://webarchivum.oszk.hu/wp-content/uploads/2021/10/Drotos_Laszlo_Webtartalmak_megorzese.pptx). Megjelent az idei Networkshop konferencia válogatott előadásait tartalmazó kötet, benne Drótos László „A közösségi média archiválásának nehézségei” című írása. (A kiadvány az MTA könyvtárának repozitóriumából tölthető le PDF formátumban: <http://real.mtak.hu/132242/>) Ugyancsak novemberben jelent meg a Könyvtári Figyelő 3. számában Németh Márton és Drótos László cikke „A webarchiválás oktatásának nemzetközi keretei és hazai tapasztalatai” címmel, melyben az International Internet Preservation Consortium oktatási munkacsoportjának tevékenysége mellett a saját tanfolyamunk online formába történt átalakításának előnyeiről és hátrányairól is beszámolunk (<http://ojs.elte.hu/kf/article/view/2954>). November 9. és 12. között a Könyvtári Intézet szervezésében egy újabb távoktatási tanfolyamot tartottunk, melyre ezúttal 15-en jelentkeztek és közülük 13-an le is vizsgáztak.

„404 Not Found” workshop

Október és november jó része az idén ötödik alkalommal megrendezett „404 Not Found – Ki őrzi meg az internetet?” című műhelykonferenciánk előkészítésével telt. A járványhelyzet miatt ezúttal is a Teams alkalmazáson keresztül zajlott le a rendezvény. Ezzel a fajta a lebonyolítással kapcsolatban tavaly már több hasznos tapasztalatot szereztünk, viszont az előadók közül néhánynak akadtak technikai problémái az előzetes teszteken és egyiküknek még az előadása közben is, de ezt szerencsére sikerült gyorsan áthidalni. Új kihívást jelentett, hogy ezúttal két napos volt a rendezvény, az első napon jórészt külföldi előadókkal és angol nyelven. A szervezésben és lebonyolításban több mint 20 kolléga vett részt az OSZK különböző osztályairól, ami rengeteg egyeztetést igényelt. A közös munka eredményeképpen sikerült határidőre elvégezni minden részfeladatot: elkészültek a grafikai tervek, a meghívók és plakátok, a sajtóanyagok és a közösségi média posztok, az előadók és a meghívottak listája, a regisztrációs űrlap, a fordítások és a videó feliratozás, a szükséges jogi nyilatkozatok stb., végül pedig a rendezvény utáni gyorsjelentések és a prezentációkat tartalmazó weboldal is (<https://webarchivum.oszk.hu/404-not-found-workshop-2021-november-23-24/>). Az érdeklődés ez alkalommal is igen nagy volt, összesen 150-en regisztráltak és mindkét napon 108-an léptek be a Teams csatornára.

Az első nap Rózsa Dávid főigazgató köszöntötte a résztvevőket, majd a hagyományokhoz híven Dr. Latorcai Csaba közigazgatási államtitkár tartott megnyitó beszédet. Az ezt követő előadásokban a közép-európai országok nemzeti webarchívumai mutatkoztak be. Megismerhettük a cseh, a szlovák, az osztrák, a horvát és a szlovén kollégák tevékenységét, az általuk használt technológiákat, a munkafolyamataikat, a helyi jogi környezetet és a gyűjteményeik jellemzőit. A délelőtti programot az Országos Széchényi Könyvtár webarchívumának rövid ismertetése és a Lengyel Állami Levéltár bemutatkozása zárta, ahol most tervezik elkezdni ezt a munkát. Délután virtuális kerekasztal beszélgetésre került sor, melynek célja egy regionális együttműködés előkészítése volt „Central European Web Archives” (CEWA) néven. Ennek keretében kialakításra kerülhetne egy közös portál és keresőfelület, mód nyílna az informatikai fejlesztések és egymás jó gyakorlatainak megismerésére, közös publikálásra és előadások tartására, több országot is érintő témákban webcím listák gyűjtésére, sőt akár egyes archiválási tevékenységek megosztására is. Mivel a feladat méretéhez képest az emberi erőforrások mindenhol szűkösek és folyamatos informatikai fejlesztésekre van szükség, ezért lenne nagyon hasznos szorosabban fűzni a kapcsolatokat az ezzel foglalkozó intézmények között. A kooperáció szükségességének másik oka a térségben élő népek közös történelmi és kulturális öröksége, ami az interneten is leképeződik és megőrzésre érdemes.

A második napon az OSZK munkatársai beszéltek a digitálisan születő dokumentumok megőrzésének fontosságáról, a Webarchiválási Osztály elmúlt idejének eredményeiről, a hazai könyvtárakkal formálódó együttműködésről, a kutatási célú hasznosítással foglalkozó nemzetközi WARCNET projektről és a webarchívum metaadatolásához készülő RDA-alkalmazásprofilról. Szünet után az ELTE Digitális Bölcsés-

szet Tanszékének munkatársa a magyar hírportálok aratása során szerzett tapasztalataikat foglalta össze, majd pedig a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában folyó e-mail archiválási kísérletről hallhattunk egy előadást. Végül a Berzsenyi Dániel Könyvtár rendszergazdája a könyvtár honlapjának megújításáról és archívumbaráttá alakításáról számolt be.

Külföldi rendezvények

Aarhus

A webarchívumok kutatási célú felhasználásával foglalkozó WARCNET projekt (<https://warcnet.eu>) tanácsülésére november 3. és 5. között került sor a dániai Aarhusban, melyen Németh Márton képviselte az OSZK-t. A munkaértekezleten főként a következő félév teendőinek meghatározása került napirendre. A projekt hat munkacsoportból tevődik össze, ezek közül háromban veszünk részt: az 1. munkacsoportban, mely a webarchívumok összehasonlító kutatását, visszakereshetőségét szolgálja, az 5. munkacsoportban, ami a kutatók igényeinek megfelelő adatformátumok és adatsémák meghatározásával foglalkozik, a 6. munkacsoportban pedig minden projektag részt vesz. Itt a fenntartható kutatások kereteinek meghatározásán dolgozunk új pályázati lehetőségek felkutatásával 2023 tavasza utánra, mikor ez a projekt kifut majd. Elkezdődött a munkacsoportok teendőinek egyeztetése mellett az eredmények könyv alakban történő publikálásának tervezése is. Keynote előadások hangzottak el a távol-keleti webarchiválási projektekről, illetve kerekasztal beszélgetés keretében bemutatkozott több európai, illetve közel-keleti intézmény, melyek webarchiválást kívánnak végezni. Betekintést nyertünk abba, hogy a dán webarchívum miként biztosítja a kutatók számára a zárt gyűjteményben történő keresést, valamint célhoz kötött módon egyes adatkészletek rendelkezésre bocsátását. Archiválási módszertani kérdéseket is áttekintettünk a Dán Nemzeti Könyvtár, a Holland Királyi Könyvtár, illetve a Francia Nemzeti Könyvtár képviselőivel együtt. A különféle kutatói igények alakulásáról, a webarchiválással kapcsolatos oktatói, kutatói percepciók formálódásáról is részletesebb képet lehetett kapni a személyes részvétellel. Mélyebb ismereteket tudtunk szerezni a webarchiválással kapcsolatos egyetemi oktatói, kutatói igényekről és a kihívásokról ezek jogszerű kielégítése kapcsán (szerzői jog, személyes adatok védelme). A Dán Nemzeti Könyvtár aarhusi részlegében tett látogatás során megbeszéltük a SolrWayback szoftver legújabb fejlesztéseivel kapcsolatos kérdéseket. A Twitter archiválást végző komponens beüzemelésére itthon is lehetőség nyílhat. Áttekintést lehetett nyerni továbbá általánosabb kontextusban is a közösségi média archiválásának kihívásairól, módszereiről. Az egyes tagok által rendelkezésre bocsátott URL listákból létrejött IIPC COVID-19 gyűjtemény adatelemzési vizsgálatainak első eredményeibe is betekintést kaptunk. Ennek hosszabb távú haszna az lesz, hogy tematikus gyűjtési tevékenységünk sajátosságai nemzetközi összefüggésekben is értelmezhetővé fognak válni.

Varsó

A Lengyel Állami Levéltár Főigazgatója az OSZK főigazgatójánál kezdeményezte részvételünket a rendezvényen (<https://eveningmedia.pl/forum-on-the-archiving-of-the-polish-internet/>), melyre november 16-17-én került sor. Főigazgató úr kérésére szakértői szinten képviseltettük magunkat Németh Márton személyében. Nicola Herbert IT-adminisztrátor előzetesen meghatározta azokat a témaköröket, melyekben tájékoztatást szerettek volna kérni. Erre épült rá a 20 perces előadásunk, melyet 20 perc vita követett. A rendezvény első napja teljes mértékben a lengyel fél által hasznosítható nemzetközi esettanulmányok bemutatásáról szólt. Mondandónkat jól egészítette ki a portugál webarchívum vezetőjének, Daniel Gomesnek, illetve a spanyol webarchívum vezetőjének, Mar Perez Morillónak az előadása is. A rendezvény külön értékét adta a levéltári webarchiválási modellek bemutatása, a brit National Archive részéről Tom Storrar, a Holland Nemzeti Levéltár képviseletében pedig Antal Posthumus révén (ő magyar felmenőkkel bír, de már nem beszél magyarul). Silvia Sevilla az Európai Unió Kiadóhivatala képviseletében a komplex intézményrendszerre kiterjedő webarchiválásról adott áttekintést. A második napon a lengyel egyetemi szféra kutatóinak bevonásával fórumbeszélgetéseken járták körül a lengyel webarchívum megalkotásának szervezési, módszertani, jogi és információ-technológiai követelményeit.

Hazai együttműködések

Novemberben a „404-es workshop” és a webarchiválás tanfolyam mellett nem sok idő jutott a folyamatban lévő ügyek kibontására, újak kezdeményezésére. Október végén érkezett meg a Fővárosi Szabó Ervin Könyvtár válasza a megkeresésünkre, természetesen igenlő válasszal, úgy, hogy ők két téma gondozásában is érdekeltek lennének, a budapesti regionális és a szociológiai tematikus gyűjtemények vonatkozásában is. A hónap közepén a szolnoki Verseghy Ferenc Könyvtár és Közművelődési Intézmény jelentkezett a megbeszélések elkezdésére, velük eddig technikai okokból nem sikerült időpontot egyeztetni, és a mi lekötöttségünk miatt csak decemberben tudunk velük érdemben beszélni az együttműködés lehetőségeiről. Szintén a hónap fejleménye, hogy a székesfehérvári Vörösmarty Mihály Könyvtár felvetette, hogy a 2022-es Aranybulla Emlékév kapcsán össze lehetne állítani egy olyan tematikus gyűjteményt, mint a mi Rákóczi Archívumunk (<https://rakoczi2019.webarchivum.oszk.hu>). Bár a részletek megbeszélése még hátra van, szeretnénk minden szakmai támogatást megadni ehhez a projekthez, ami azért is példaértékű, mert önálló kezdeményezés és az általunk megosztott ismeretekre épül.

Az elmúlt hetekben lefutott tematikus és műfaji aratások

Irodalom, irodalomtudomány és -történet (1.363 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (1.937 db seed URL)
Közoktatás és egyéb képzések (7.155 db seed URL)
Képző-, előadó-, zene- és filmművészet (8.031 db seed URL)
Idegenforgalom, vendéglátás (5.832 db seed URL)

A tematikus aratások részletes statisztikai adatai a <https://webarchivum.oszk.hu/szelektiv-aratasok/> weblapon nézhetők meg. A projekt hírei a <https://webarchivum.oszk.hu/a-projektrol/hirek-esemenyek/> oldalon kísérhetők figyelemmel. Kapcsolati cím: mia@mek.oszk.hu