

Az OSZK Webarchívum 2021 áprilisi hírei

A „404-es workshop” három előadása a Könyvtári Figyelőben

A Könyvtári Figyelő idei első számában megjelent a 2020 decemberében megrendezett „404 Not Found – Ki őrzi meg az internetet?” online workshop Latorcai Csaba (EMMI) államtitkár által tartott megnyitó beszéde, valamint a Webarchiválási Osztály munkatársainak két előadása „Az OSZK webarchívumának 2020-as újdonságai” és „Együttműködési lehetőségek a webarchiválás területén” címmel. A cikkek elérhetősége: <http://ojs.elte.hu/kf/issue/view/203>

Networkshop tutorial és előadások

Április 7. és 9. között került sor az online térben az idei Networkshop konferenciára. A nulladik napon az EPA munkatársával, Ipacs Eszterrel Drótos László és Németh Márton közösen tartott tutorialt az online webes tartalomként, illetve elektronikus kiadványként megjelenő folyóiratok archiválásáról és feldolgozásáról. A konferencián a Webarchiválási Osztály képviselőjében Drótos László a közösségi média szolgáltatások archiválásának kihívásairól, Németh Márton pedig a Covid-19 járványról szóló hírekből kialakított gyűjteménnyel kapcsolatos tapasztalatokról tartott előadást, bemutatva a pandémiával foglalkozó nemzetközi webarchiválási projekteket is. A prezentációk elérhetők a webarchívum honlapján: <https://webarchivum.oszk.hu/a-projektrol/eloadasok-prezentaciok-publikaciok/#2>

Online WARCNet konferencia

A webarchívumok anyaga tudományos hasznosításának, illetve a webarchívumokhoz kötődő oktatási tevékenységeknek különféle módjait kutató WARCNet projekt legutóbbi háromnapos online konferenciájára április 20-22. között került sor. Minden munkacsoport beszámolt a tevékenységéről, terveiről, megvitatásra kerültek az együttműködés főbb területei. Kerekasztal-beszélgetés járta körül a webarchiválással kapcsolatos oktatási kérdéseket, illetve azt, hogy az archivált tartalmak begyűjtésében, kezelésében az egyes résztvevő intézmények, egyetemi kutatók milyen kompetenciákra, illetve szoftveres eszköztárra támaszkodnak. A prezentációk elérhetők ezen a weboldalon: <https://cc.au.dk/en/warcnet/presentations/aarhus-meeting-2021/>

A Cseh Nemzeti Könyvtár bemutatkozó webinárium

Április 23-án az IFLA Nemzeti Könyvtári Szekció és az Illinoisi Egyetem könyvtárának szlavisztikai tájékoztató szolgálatának szervezésében online bemutatkozási lehetőséget kapott a Cseh Nemzeti Könyvtár. A járványhelyzet ideje alatt jelentősen ki tudták bővíteni digitális szolgáltatásaik palettáját, a korábnál sokkal nagyobb figyelem fordult az idén húsz éves évfordulóját ünneplő webarchívum nyilvános tematikus gyűjteményei felé is. Érdekes lehet majd az OSZK számára is a big-data alapú webarchívum adatfeldolgozási projekt eredménye, melyet nem régen kezdtek el. A teljes rögzített esemény megnézhető a Youtube-on: <https://youtu.be/c4xxpSiEk8A> A rendezvény prezentációi: <https://docs.google.com/presentation/d/1Er-IsOKdRbcwmkL49FKKsJ8at9IAhiZnCONCkIQbIKM/>

MIC minikonferencia

Holland szervezésben került sor – elsősorban a korábbi BOBCATSSS konferenciákon előadó szakemberek részvételével – a Mini Information Science Congress (MIC) nevű rendezvényre, ahol Németh Márton rövid, 15 perces előadást tartott a webarchiválásról. A konferencia keretében több magyar előadás is elhangzott az ELTE BTK szakmai képzési palettájáról, az egyetemi könyvtárvezetőkkel folytatott interjúk tapasztalatairól a könyvtári menedzsment vonatkozásában, illetve bemutatkozott a jövő évi BOBCATSSS konferencia szervezője, a Debreceni Egyetem Informatikai Kar is.

Kapcsolatfelvétel az ELTE Digitális Bölcsészeti Tanszékkal

Április 23-án megbeszélést folytattunk az ELTE BTK Digitális Bölcsészeti Tanszékének munkatársaival (Indig Balázs, Palkó Gábor, Sárközi-Lindner Zsófia). Tájékoztatást kaptunk a náluk folyó webarchiválási tevékenységről, illetve tájékoztatást adtunk a kutatókat érintő együttműködési lehetőségekről, nemzetközi projektekről. Az általuk összeállított cikkarchívum címlistáját megkaphatjuk, hogy össze tudjuk vetni a saját címlistánkkal, Ők pedig használni tudnák az OSZK-s MIA ID-eket az egyes kiadványok azonosításához. A szemantikus webes, illetve a nagy mennyiségű adatok kezelésre szolgáló kutatási projektjeik eredményeinek hasznosítása is releváns lehet a jövőben az OSZK számára. Az ELTE-n fejlesztett entitás-felismerő rendszert használhatnánk a webarchívumban nevek, fogalmak, események stb. beazonosításához. Ők is tesztelik majd a SolrWayback keresőt, ennek továbbfejlesztésében is nyílnak együttműködési lehetőségek. Azt is felajánlották, hogy az általuk végzett archiválási tevékenység során keletkezett WARC fájlokat hosszú távú megőrzés céljára az OSZK rendelkezésére bocsátják. A felek havi konzultációkat terveznek az együttműködés további részleteiről. A Történeti Intézet keretei között működő tanszék tevékenységéről itt érhetőek el információk: <https://elte-dh.hu/>

IIPC hírek

Aláírásra került az OSZK részéről is az International Internet Preservation Consortium (IIPC) újabb öt éves ciklusára, 2026 végéig érvényes konzorciumi szerződés. Az elnöki teendőket idén év elejétől Abbie Grotke, az amerikai Kongresszusi Könyvtár webarchívumának vezetője vette át a következő két évre. Az adminisztrációs feladatokat a British Library helyett immár egy amerikai cég látja el, a szakmai munka szervezéséért felelős Olga Holownia pedig az IIPC alkalmazásába került át a British Library-ből. Az idén június 15-16-án rendezendő éves konferencia a tavalyihoz hasonlóan az online térben lesz megtartva, a hosszú távú digitális megőrzés témájában kétvétenként rendezett RESAW konferenciával együtt. Június közepén így egy teljes hét a webarchiválásról szól majd: <https://netpreserve.org/ga2021/wac>

ArchiveBox új verzió

Április közepén megjelent az elsősorban személyes archiválásra hasznos ArchiveBox szoftver új, 0.6.2-es változata, melyben teljesítmény növelése mellett több hasznos új funkció is van. Az ArchiveBox különböző eszközöket (pl. wget, Chromium, youtube-dl) integrál és ezek segítségével ment le PDF, PNG, HTML, WARC, stb. formátumokban webes tartalmakat egy megadott URL cím vagy címlista (pl. könyvjelzők, böngészési előzmények, RSS linkek) alapján. Alapvetően parancsmódban használható, de van hozzá egy egyszerűbb grafikus felület is, ami böngészőből érhető el. Az általunk oktatási és tesztelési célokra összeállított Ubuntu rendszerű virtuális gépre is telepítettük már ezt az új változatot. A szoftver honlapja: <https://archivebox.io>

Az ArchiveWeb.page új verziója

Április 27-én a korábban online szolgáltatásként és önálló alkalmazásként is használható Webrecorder utódja, az ArchiveWeb.page felületén is megjelent az „autopilot szolgáltatás” (a 0.6.4-es verzióban), ezzel lehetővé válik közösségi médiatartalmak félautomatizált begyűjtése is. Sajnos az első tesztjeink szerint a Facebook esetében még nem működik rendesen ez a funkció. A projekt honlapja a <https://archiveweb.page> címen található, az autopilot fejlesztések összefoglalása pedig itt olvasható: <https://webrecorder.net/2021/04/21/autopilot-testable-automated-behaviors.html>

Twitter csatornák mentése

Az ArchiveWeb.page Chrome böngészőbe integrálható modulját használva elkezdtünk kísérletezni a Twitter csatornák archiválásával. Első lépésben 89 könyvtár, múzeum és galéria oldalát töltöttük le 600 MB összméretben. 75 esetben sikerült egészen az első bejegyzésig visszagörgetni, a legrégebbi elmentett tweet 2007 decemberi volt: <https://webarchivum.oszk.hu/webarchivum/bongesztes-twitter/>

ARCHIVE WEB.PAGE / TWITTER_01 / Pages

search://view=pages

Pages URLs Info

Search by Page URL, Title or Text

TWITTER_01

37 Pages Found

Download

Date	Page Title
2021. 04. 28. 17:43:00	OSZK / NSZL (@nemzetikonyvtar) / Twitter https://twitter.com/nemzetikonyvtar 35.6 MB
2021. 04. 28. 17:42:17	Nagykutasi Könyvtár (@NagykutasiKonyv) / Twitter https://twitter.com/NagykutasiKonyv 6.64 MB
2021. 04. 28. 17:41:29	Nagybjom könyvtár (@GyergyaiAlbert) / Twitter https://twitter.com/GyergyaiAlbert 6.87 MB
2021. 04. 28. 17:40:08	Nagy Miklós Városi Könyvtár (@nmvkszeghalom) / Twitter https://twitter.com/nmvkszeghalom 8.44 MB
2021. 04. 28. 17:37:52	MOME Könyvtár (@momekonyvtar) / Twitter https://twitter.com/momekonyvtar 6.72 MB

Sportesemények híreinek aratása

Egy újabb esemény-alapú aratást indítottunk heti rendszerességgel: a járvány miatt 2020-ról idénre halasztott Foci EB-vel kapcsolatos híreket gyűjtjük a főbb magyar nyelvű online forrásokból: <https://webarchivum.oszk.hu/bongesz-es-foci-eb-2021/> Továbbá újraindítottuk a tokiói Nyári Olimpiáról szóló cikkek archiválását, amelyet tavaly májusban leállítottunk, miután kiderült, hogy ez az esemény is halasztásra került: <https://webarchivum.oszk.hu/bongesz-es-nyari-olimpia-2020/>

TÖRTÉNELEM gyűjtemény

Április végén aratjuk első ízben az új, történelem tematikájú gyűjteményünket, ami magába foglalja a kapcsolódó területek webhelyeit is: a magyar és egyetemes történelem mellett ebbe a gyűjteménybe kerül a régészet és a művelődéstörténet, de a műemlékvédelem és a hagyományőrzés is. Külön csoportot alkotnak az interneten is népszerű témák, mint a had-, család- és helytörténet, de nem maradhat ki a politika-, a testület- és a szakmatörténet, ahogyan a történelem segédtudományai és az oktatás sem. A gyűjtés nem zárult le (jelenleg csak mintegy 500 webhelyet tartunk nyilván), folyamatosan szeretnénk bővíteni a címlistát ezután is. Ebben számítunk az érdeklődők segítségére is, mert a téma hatalmas és itt is érdemes kihasználni az ezzel foglalkozók vagy a közvetlen kapcsolatok előnyét. A már összegyűjtött webhelyek itt nézhetők meg: <https://webarchivum.oszk.hu/webarchivum/bongesz-es-tortenelem/> A gyűjteménybe illő webhelyet ezen az úrlapon lehet javasolni: <https://forms.office.com/r/Cz3GsHSQEd>

Az elmúlt hetekben lefutott egyéb aratások

Kormányzat, önkormányzatok, politikai és civil szervezetek (5946 db seed URL)

Média, sajtó, műsorszórás (949 db seed URL)

Könyv- és egyéb kiadók, kereskedők (1447 db seed URL)

Elektronikus periodikák (7047 db seed URL)

Az egyes 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