

Az MNMKK OSZK Webarchívum 2024. novemberi hírei

Archiválás és nyilvántartás

Egyelőre lezártuk az Instagram oldalak szeptemberben elkezdett tömeges archiválását. November elején még művelődési házak és egyéb kulturális helyek, pártok és politikusok, valamint néhány civil és kormányzati szervezet fiókjait mentettük le az ArchiveWeb.Page programmal (110 tétel, több mint 86 GB).

Elkezdtek viszont a podcast csatornák hanganyagának frissítését. Mivel nagyjából egy éve volt az utolsó mentés, ezért most az azóta közzétett adásokat töltjük le a Podcasts nevű Chrome kiegészítővel. Ebben a hónapban 75 csatornáról közel 3400 fájlt mentettünk le 210 GB összméretben. Ez a munka a következő hetekben is folytatódik majd.

A decemberi webtér aratás előkészítéseként a kezdetekig visszamenőleg kigyűjtjük a zárt archívumban található WARC fájlokból a .hu végű URL címeket, abban a reményben, hogy lesznek köztük olyan domének vagy aldomének, amelyekkel bővíteni tudjuk a seed-listát. A fájlok nagy száma és mérete miatt ez a lekérdezés több hétig fut, és utána lehet megvizsgálni majd, hogy az eddig számunkra még ismeretlen URL-ek közül melyeken van valamilyen (gyűjtőkörbe tartozó) tartalom.

Ismeretterjesztés és -szerzés

A belső használatú Redmine wikit három terjedelmes szócikkkel bővítettük („Podcast adások archiválása”, „Instagram fiókok archiválása AWP-vel”, „Tömeges aratások statisztikájának készítése”) és frissítettünk további hármat („A webtér seedlista karbantartása”, „Heritrix”, „Köztérképes projekt”).

November 11-én a Könyvtárostanárok Egyesületének (KTE) Őszi Szakmai Napján Visky Ákos László, az OSZK webarchívumának kurátora és Dömsödy Andrea, az ELTE Pedagógiai és Pszichológiai Kar könyvtárosa, a KTE alelnöke műhelybeszélgetést tartott a résztvevőkkel a webarchiválás fontosságáról és jelenlegi helyzetéről, valamint a lehetséges kapcsolódási pontokról ezen a területen a nemzeti könyvtár és az iskolai könyvtárak között. A beszélgetést felvezető prezentáció letölthető a honlapunkról.

A Petőfi Irodalmi Múzeum november 14-én rendezte meg „Born Digital 3.0 – Digitálisan keletkezett műtárgyak a múzeumi gyakorlatban” című műhelykonferenciáját, melyen Kalcsó Gyula és Horváth Dániel közösen tartottak előadást „E-mailek archiválása az Országos Széchényi Könyvtárban és a Petőfi Irodalmi Múzeumban” címmel.

November 25-én, az OSZK alapításának 222. évfordulóján rendezett ünnepségen Drótos László, a Széchényi Ferenc-emlékérem tavalyi kitüntetettje rövid előadást tartott „Megőrizni az internetet” címmel a Digitális Bölcsészeti Központban folyó webarchiválási munkáról. Ez a prezentáció is megtalálható a honlapunkon a <https://webarchivum.oszk.hu/eloadasok-prezentaciok-publikaciok/> oldalon.

Lezajlott a nyolcadik „404 Not Found – Ki őrzi meg az internetet?” konferencia és workshop, melyet az idén november 27-én rendeztünk meg. A résztvevők száma meghaladta a nyolcvanát, ebből kb. ötvenen online követték az eseményt. A rendezvény nyitó előadása az amerikai Internet Archive Community Webs projektjét mutatta be a szervezet vezetőjének prezentálásában. (Az előzetesen megkapott videóhoz magyar feliratot is készítettünk.) A helyi webes tartalmak archiválásának témájában a Fejér vármegyei Vörösmarty Mihály Könyvtár úttörő vállalkozásának, egy regionális webarchívum létrehozásának a feladatait, problémáit, megoldásait ismerhettük meg, külön kitérve annak könyvtári információs ökoszisztémába illesztésére, illetve egy helyi webarchívum technikai feltételeire. A konferencia másik kiemelt témája a digitálisan született (*born digital*) dokumentumok és egyéb tartalmak kezelése a közgyűjteményekben, amelynek kihívásait, jó gyakorlatait előadásokon és kerekasztal-beszélgetéssel járták körül a szakemberek. Az előadások prezentációi és absztraktjai, valamint a helyszínen készült fotók elérhetők a rendezvény weboldalán: <https://webarchivum.oszk.hu/404-not-found-konferencia-es-workshop-2024-november-27/>



A „404 Not Found – Ki őrzi meg az internetet?” konferencia és workshop megnyitója

A szervezőbizottság elfogadta Kalcsó Gyulának az International Internet Preservation Consortium jövő évi konferenciája poszter szekciójába „Web Scraping in the Hungarian Web Archive” címmel beadott tervezetét. A rendezvényre Oslóban kerül sor 2025. április 8. és 10. között.

Az IIPC két webinariumot is tartott ebben a hónapban. A november 6-i „Web Archive Collections as Data” címűn a norvég és a dán nemzeti webarchívumok munkatársai mutatták be, hogy milyen felhasználói felületet és szolgáltatásokat biztosítanak a kutatóknak. November 19-én pedig „Web Archiving in the Program Management Context: Web Archiving Roles” címmel volt egy online workshop, melyen holland, norvég, brit, kanadai és amerikai kollégák beszéltek a webarchiváló munkacsoportjaik összetételéről, a különböző munkakörökről és az új munkatársak kiválasztásának szempontjaitól.

Az elmúlt hetekben lefutott aratások

Irodalom, irodalomtudomány és -történet (1523 db seed URL)
Kulturális intézmények, művelődési házak, rendezvényhelyszínek (1015 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (2206 db seed URL)
Párkapcsolat, család (2102 db seed URL)
Természet- és műszaki tudományok, szakterületek (2559 db seed URL)
Bölcsészet- és társadalomtudományok, szakterületek (6313 db seed URL)
Közoktatás és egyéb képzések (7408 db seed URL)
Képző-, előadó-, zene- és filmművészet (9057 db seed URL)
Podcastok (4237 db seed URL)
Elektronikus periodikák (10907 db seed URL)

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