

Az OSZK Webarchívum 2023 áprilisi hírei

Archívumok és nyilvántartások

A hónap végi pápalátogatásról szóló hírekből és egyéb kapcsolódó weboldalakkól létrehoztunk egy több mint száz forrásból álló esemény-alapú részgyűjteményt. Az első aratásra április 25-én került sor és a látogatás ideje alatt, majd pedig utána 1-2 hétig ezt még többször megismételjük. Az eseményt követően pedig megpróbáljuk a közösségi médiából is összegyűjteni a releváns tartalmakat, amennyire ez technikailag lehetséges.

A könyvtáralapító Széchényi Ferenc életét és emlékezetét bemutató digitális gyűjteményünk a gróf születésnapjára, április 28-ára készült el. Az utolsó napokban még a részletes keresőn és egyéb kisebb hibák javításán dolgoztunk. A honlap rövidesen nyilvánosan is elérhető lesz.

Folytatódott az ELETMOD kódnevű új tematikus részgyűjtemény építése, amibe a magánélet különféle területeivel foglalkozó webhelyek adatai kerülnek bele. Áprilisban főként a gyűjtőszemély és modellezés, valamint a kertészkedés és a gasztronómia témájú oldalakból válogattunk közel 3 ezret. Ez a munka még májusban is folytatódik.

Az elektronikus periodikák nyilvántartása 70 címmel bővült a hónap folyamán, a többi tematikus címlista pedig több mint 200 tétellel.

Hazai rendezvények

Kalcsó Gyula kollégánk a veszprémi Networkshop konferencián a Katalist levelezőcsoport anyagának archiválásáról beszélt. A több mint 44 ezer levelet és 10 ezer mellékletet különböző szoftverekkel és scriptekkel sikerült szabványos formátumokra konvertálni és hosszú távú megőrzésre alkalmas beadási csomagot képezni belőlük. Egyebek mellett WARC fájlok is készültek, melyek – egyelőre kísérleti jelleggel – teljes szöveggel is kereshetők a webarchívumban használt a SolrWayback program segítségével.



Az Aranybulla-webarchívum kezdőlapja

„Távcső a történelemre” címmel rendezett konferenciát április 24-én a székesfehérvári Vörösmarty Mihály Könyvtár, ami egyben a 2022-es Aranybulla-émlékév alkalmából készült webarchívum (webarchivum.vmk.hu) nyilvános bemutatója is volt. A rendezvényen mi is tartottunk három előadást: Drótos László az OSZK webarchívumát ismertette, Visky Ákos László a Rákóczi Emlékévből létrehozott digitális gyűjteményről beszélt, Kalcsó Gyula pedig az archivált webes tartalom kutatási célú hasznosításáról.

Előadásaink prezentációi letölthetőek a honlapunkról (<https://webarchivum.oszk.hu/eloadasok-prezentaciok-publikaciok>).

Nemzetközi ügyek

„Le web : source et archive” címmel rendezett április elején három napos konferenciát a webarchiválás és a webarchívumokban való kutatás különböző részterületeiről a Université de Lille, melyet hála az élő közvetítésnek és az angol szinkronfordításnak mi is figyelemmel kísérhettünk.

Április 18-án az Artificial Intelligence for Libraries, Archives & Museums (AI4LAM) szervezett nyilvános online eseményt, melyen egy norvég projektet ismertettek a számi nyelv kutatásáról, valamint bemutatták a British Library rádióműsor archívumát. Számunkra a használt ASR (Automatic Speech Recognition) technológiák volt elsősorban érdekesek mindkét előadásban, amelyekkel kereshetővé és elemezhetővé tették a begyűjtött hangfelvételeket. Hasonlót tervezünk mi is a podcast archívumunkkal.

Ugyanezen a napon az International Internet Preservation Consortium webinárium-sorozatában a böngésző-alapú archiváló rendszer, a Browsertrix Cloud két fejlesztője mutatta be a legfrissebb verzió újdonságait, köztük azt a funkciót, amivel menet közben lehet figyelni a letöltésre váró URL címeket és közbe is lehet avatkozni, ha olyan linkeket próbálna követni a robot, amelyeket nem akarunk letölteni.

A nemzetközi konzorcium közgyűlése és konferenciája idén Hollandiában lesz május 10. és 12. között, előtte pedig május 3-án online előadásokat lehet megnézni. A rendezvényre mi is készülünk egy rövid összefoglalóval az elmúlt évi eredményeinkről és a jövőbeli terveinkről. E mellett rövidesen frissülni fog az OSZK webarchívumának ismertetője is az IIPC honlapján, mely még a 2017-2018-as pilot fázisról szólt.

A Society of American Archivists webarchiválási szekciójának blogjában (webarchivingrt.wordpress.com) megjelent Kalcsó Gyula írása, melyben az OSZK-ban folyó webarchiválást mutatja be.

Az elmúlt hetekben lefutott tematikus és webtér aratások

Történelem, hely- és családtörténet (920 db seed URL)

Média, sajtó, műsorszórás (911 db seed URL)

Könyv- és egyéb kiadók, kereskedők (1575 db seed URL)

Podkasztkok (4430 db seed URL)

Elektronikus periodikák (9830 db seed URL)

Kormányzat, önkormányzatok, politikai és civil szervezetek (6730 db seed URL)

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