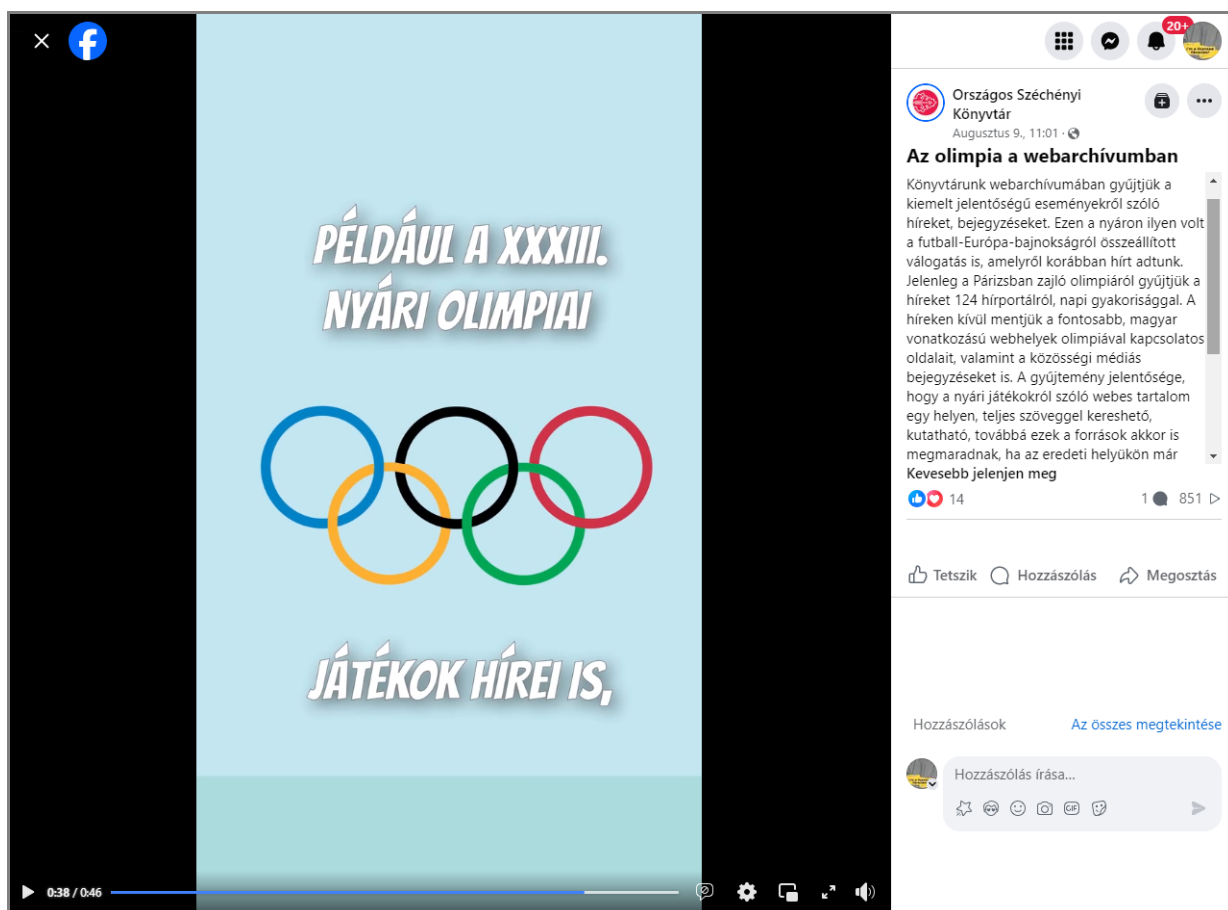


Az MNMKK OSZK Webarchívum 2024. augusztusi hírei

Archiválás és nyilvántartás

Felkerült a honlapunkra a nyári webtér aratás statisztikája, melyen az látszik, hogy bár most jelentősen kisebb volt a kiindulásnak megadott seed-URL lista mérete, mivel a 2022 óta használt közel 1,4 milliós tételes listát megtisztítottuk a már nem működő vagy nem releváns címektől, de összességében így is sikerült 130 millió URL-t letölteni a robottal, ami átlagosnak számít a korábbi webtér-szintű aratásokhoz viszonyítva. Augusztus elejére elkészültek az oldalképek is: 545 ezer darab PNG fájl, 1 terabájt összméretben.

Befejeződött SPORT részgyűjtemény címlistájának aktualizálása a hónap első felében. A júliusban átnézett 2 ezer tétel után augusztusban még további 1800 webhelyet ellenőriztünk egyenként megnyitva őket a böngészőben. Összességében 630 URL-t kellett megszüntetnek nyilvánítani, de ezek egy részénél sikerült megtalálni az új címet, ahová a honlap átköltözött – általában megújult formában. Így jelenleg kb. 3500 sport témájú honlapot és blogot tartunk nyilván, melyek következő aratása szeptemberben lesz.



A webarchívum „reklámfilmje” az OSZK Facebook oldalán

A párizsi olimpia kezdete óta napi rendszerességgel mentjük a Heritrix szoftverrel az eseménnyel kapcsolatos híreket és egyéb weboldalakat. Augusztus 7-én és 14-én pedig az ArchiveWeb.page programmal – emberi közreműködéssel – begyűjtöttünk több mint 10 gigabájnyi, robottal nem aratható közösségi média (Facebook, Instagram, Threads, X, Viber, YouTube) tartalmat is. Az augusztus 28-án

kezdődött paralimpiai játékok előtti héten a seed listát kibővítettük a paralimpiai hírekhez tartozó címkékkel és keresőkérdésekkel (93 db), az esemény befejezése után pedig szintén letöltjük a közösségi médiában megjelent posztokat is, amennyire ez technikailag lehetséges. A NYAROL2024 kódjelű gyűjteményhez beüzemeltünk egy teljes szövegű keresőt, az OSZK Facebook oldalán pedig megjelent róla egy rövid animáció: <https://www.facebook.com/reel/501833425863916>

A 45. nyílt és 30. női sakkolimpia Budapesten kerül megrendezésre 2024. szeptember 10. és 23. között. Az eseménnyel kapcsolatos hírekből, honlapokból, egyedi weboldalakból és webkettes tartalmakból egy új részgyűjteményt állítottunk össze és augusztus 28-án elindult ennek az archiválása is. A címlista már most több mint 100 tételes, de a verseny elindulása után megpróbáljuk még tovább bővíteni. A közösségi tartalmakat és az egyedi dokumentumokat pedig a rendezvény végén mentjük majd le, szintén az ArchiveWeb.page böngészőkiegészítővel.

Egyéb feladatok

Folyik a megújított tanfolyami tematikához tartozó tananyag összeállítása. Két modulhoz már elkészült a prezentáció, két továbbihoz pedig elkezdődött az anyaggyűjtés. A tanfolyam kiemelt témája a Web Curator Tool nevű archiváló keretrendszer, amellyel gazdag funkcionalitásának köszönhetően – feltelepítés után – különösebb informatikai tudás nélkül is létre lehet hozni egy komolyabb helyi webarchívumot. A gyakorláshoz szükséges tesztserveren végre sikerült elhárítani azt a problémát, amit a magyar nyelv környezeti változóként való beállítása okozott és ami miatt bizonyos ikonok nem jelentek meg a WCT felületén.

Felkerültek a YouTube-ra az áprilisi IIPC WAC (Web Archiving Conference) előadásainak felvételei. A rendezvényt idén Párizsban tartották meg, melyen az OSZK-t Kalcsó Gyula képviselte. Az érdekesebb videókat belinkeltük a MIA Wiki megfelelő szócikkeibe, a konferencián bemutatott újdonságokról pedig új szócikkeket írtunk, melyek egy része bekerül majd a tanfolyamhoz ajánlott források közé is.

Az elmúlt hetekben lefutott aratások

Képző-, előadó-, zene- és filmművészet (8976 db seed URL)
Közoktatás és egyéb képzések (7397 db seed URL)
Bölcsészet- és társadalomtudományok, szakterületek (6309 db seed URL)
Természet- és műszaki tudományok, szakterületek (2553 db seed URL)
Párkapcsolat, család (2102 db seed URL)
Irodalom, irodalomtudomány és -történet (1515 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (2204 db seed URL)
Kulturális intézmények, művelődési házak, rendezvényhelyszínek (1015 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 kísérhető figyelemmel. Kapcsolati cím: webarchivum@oszk.hu