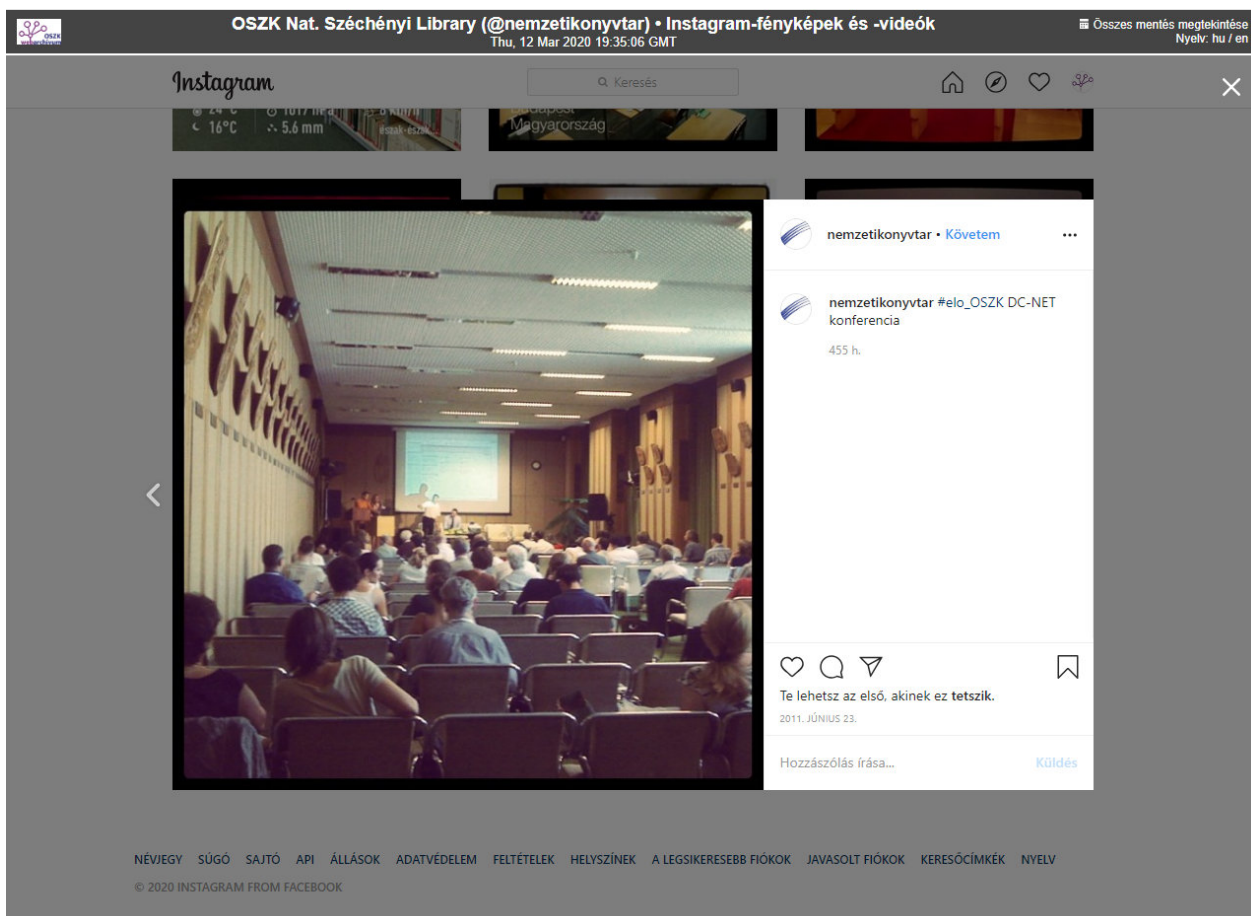


Az MNMKG OSZK Webarchívum 2024. októberi hírei

Archiválás és nyilvántartás

Sikerült megoldani a tömeges mentésekhez használt webharvest2 szerveren a szeptemberi újraindítás után felmerült, a megnyitható fájlok maximális számával kapcsolatos konfigurációs problémát, így be tudtuk pótolni a múlt hónapban el- vagy félbemaradt aratásokat. A nyilvános gyűjteményt archiváló és szolgáltató webadmin szerveren pedig az alapértelmezett konfigurációs fájlban 1 gigabájtról 2-re növeltük az egy adott webhelyről maximálisan letölthető tartalom mennyiségét, hogy minél teljesebb mentések készüljenek ezekről a honlapokról és blogokról.

Folytattuk az Instagram fiókok címlistájának bővítését és azok félautomatikus mentését, elsősorban az elektronikus periodikák, a közgyűjtemények és a tudományos intézmények, a könyvkiadók és a könyvkereskedők, valamint a kulturális intézetek és a művelődési házak oldalaival. A korábbi gyakorlattól eltérően most nemcsak letöltjük a legutóbbi mentés óta közzétett új posztokat, hanem a nyilvántartásunkban feljegyezzük azok darabszámát, valamint a legelső bejegyzésig visszalapozva annak dátumát és a már régóta nem frissülő fiókok esetében a legutolsó aktivitás napját is. Az eddig átnézett több mint 700 fiók közül a Nemzeti Sporté a legnagyobb (kb. 22,5 ezer poszt), a legrégebbi pedig az Országos Széchényi Könyvtáré, amely 2011. június 23-án tette közzé az első fotót, alig 9 hónappal azután, hogy megjelent az Instagram iOS applikáció. Októberben 280 gigabájtnyi képet és rövid videót, valamint szöveget és metaadatot tudunk archiválni az instagram.com szerverről az ArchiveWeb.page Chrome kiegészítővel, közel 200 darab tömörített WARC fájlban eltárolva.



A 2011 júniusi első bejegyzés a nemzeti könyvtár archivált Instagram oldalán

A Petőfi Irodalmi Múzeumból kérésünkre küldtek egy 29 tételes listát a PIM által gondozott honlapokról, melyet összevetettünk a saját nyilvántartásunkkal. Kiderült, hogy a címek közül 15-öt eddig még egyik tematikus gyűjteményünkbe sem vettünk fel, sőt volt 6 olyan is, ami a félévenként aratott – nagyrészt automatikusan gyűjtött URL-ekből álló – webtér-szintű címlistában sem volt benne. Utóbbinak részben az lehet az oka, hogy nem minden évben szedtük ki a WARC fájljokból a .hu végű doménekre és aldoménekre mutató linkeket. Ezért a decemberi webtér aratás előkészítése során azt tervezzük, hogy visszamenőleg a teljes állományból kigyűjtjük a .hu címeket.

Ismeretterjesztés és -szerzés

2024. október 1-jén Budapesten tartották meg a Common European Data Spaces program aktuális workshopját, melyen Kalcsó Gyula „A magyar nemzeti könyvtár webarchiválási tevékenysége” címmel tartott előadást. A rendezvényről szóló hír az OSZK Facebook oldalán olvasható.

Folyik a november 27-i „404 Not Found - Ki őrzi meg az internetet” című rendezvényünk szervezése. Az elmúlt hetekben nagyjából összeállt a program és a meghívandók listája, és elkészült a grafikai terv is.

Az „Internetes tartalmak archiválása” tanfolyam megújuló tananyagához elkészültek a Web Curator Tool keretrendszer aloldalainak képernyőfotói, valamint csináltunk egy táblázatot a WCT felhasználói felületén előforduló angol szakkifejezések magyarításához.

Felkérést kaptunk, hogy a Könyvtáros Tanárok Egyesületének november 11-én tartandó szakmai napján kb. egy órában mutassuk be az OSZK webarchiválási tevékenységét és közösen ötleteljünk arról, hogy hogyan tudnának segíteni az iskolai könyvtárosok, a tanárok, sőt akár a diákok is a magyar webtér minél teljesebb feltérképezésében és a webarchiváláshoz kapcsolódó egyéb részfeladatokban. Az intézményi és a személyes digitális archiválás fontosságáról és módszereiről 2020-ban készítettünk egy középiskolásoknak szánt [e-learning anyagot](#), ezért valószínűleg ennek hasznosításáról is szó lesz majd ezen a műhelybeszélgetésen.

Október 15-én „Heritrix Updates & Uses” címmel szervezett Zoom-alapú tájékoztatót az IIPC konzorcium, melyen az általunk is elsődleges archiváló eszközként használt Heritrix rendszer legfrissebb verziójának újdonságairól kaptunk tájékoztatást.

Az elmúlt hetekben lefutott aratások

Sport, testkultúra (3667 db seed URL)

Mezőgazdaság és élelmiszeripar (2764 db seed URL)

Vallások, hitrendszerek, egyházak (2986 db seed URL)

Könyv- és egyéb kiadók, kereskedők (1558 db seed URL)

Média, sajtó, műsorszórás (993 db seed URL)

Szolgáltatás, kereskedelem, szállítás, közlekedés (11638 db seed URL)

Kormányzat, önkormányzatok, politikai és civil szervezetek (7357 db seed URL)

Történelem, hely- és családtörténet (1502 db seed URL)

Életmód, szabadidő, hobbi (9211 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