Az MNMKK OSZK Webarchívum 2025. januári hírei

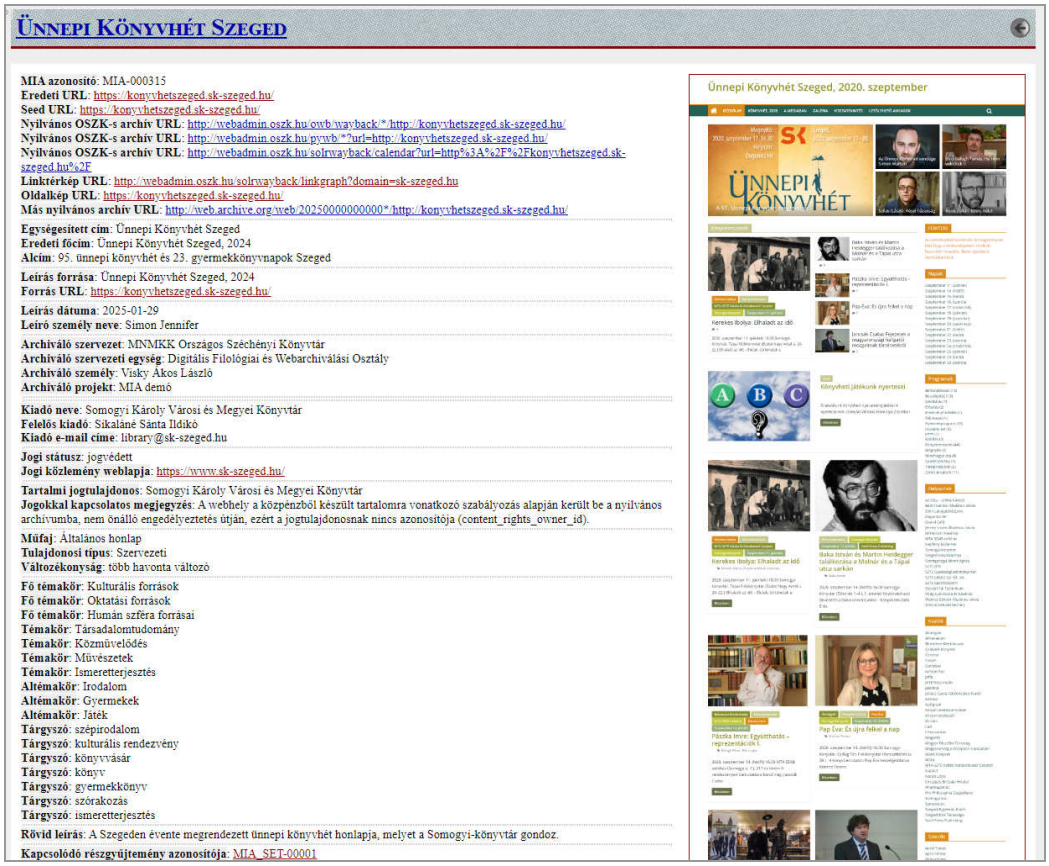
Archiválás

A webarchívum szervereinek költöztetése miatt megpróbáltuk januárban lefuttatni az első negyedévre betervezett összes tömeges aratást, hogy a szerverek másolása alatt ne keletkezzen rajtuk új tartalom. Ez nagyrészt sikerült is, mindössze két olyan job volt, amely február első napjaiban zárult le, a VALLAS nevű részgyűjtemény esetében pedig úgy döntöttünk, hogy azt a job-ot majd a költözés idejére beállított ideiglenes szerver teszteléséhez fogjuk használni. (A januári aratások listája a beszámoló végén található.)

Hasonló okból leállítottuk a nyilvános gyűjteményhez használt Web Curator Tool keretrendszerben a következő hetekre ütemezett ismétlő mentéseket. A publikus demó archívum januárban négy új tétellel bővült: két megyei könyvtári honlappal, valamint a Magyar Nemzeti Múzeum Közgyűjteményi Központ és a Civil Közoktatási Platform portáljaival. A zárt gyűjteményben is csináltunk egyedi mentéseket néhány fontos vagy megújulás/költözés előtt álló webhelyről a Heritrix és/vagy a HTTrack, illetve az AWP Express programokkal.

A „Könyvtári Kihívás” programmal kapcsolatos hírek és közösségi média tartalmak megőrzésére múlt év decemberében létrehozott részgyűjteményt ezentúl heti rendszerességgel bővítjük és mentjük az ArchiveWeb.page böngészőkiegészítővel. Január végén már több mint 100 weboldalt és 1550 szöveges bejegyzést, képet, illetve videót tartalmazott ez az esemény-alapú archívum.

Folytattuk a podcast csatornák szintén még tavaly elkezdett frissítését. 2025 első hónapjában 142 csatorna 6079 adását archiváltuk 409 GB összméretben a Podcasts nevű Chrome bővítmény segítségével.



ÜNNEPI KÖNYVHÉT SZEGED

MIA azonosító: MIA-000315
Eredeti URL: https://konyvhetszeged.sk-szeged.hu/
Seed URL: https://konyvhetszeged.sk-szeged.hu/
Nyilvános OSZK-s archív URL: http://webadmin.oszk.hu/owb/wayback/*/http://konyvhetszeged.sk-szeged.hu/
Nyilvános OSZK-s archív URL: http://webadmin.oszk.hu/pywb/*?url=http://konyvhetszeged.sk-szeged.hu/
Nyilvános OSZK-s archív URL: http://webadmin.oszk.hu/solrwayback/calendar?url=http%3A%2F%2Fkonyvhetszeged.sk-szeged.hu%2F
Linktérkép URL: http://webadmin.oszk.hu/solrwayback/linkgraph?domain=sk-szeged.hu
Oldalkép URL: https://konyvhetszeged.sk-szeged.hu/
Más nyilvános archív URL: http://web.archive.org/web/20250000000000*/http://konyvhetszeged.sk-szeged.hu/

Egységesített cím: Ünnepi Könyvhét Szeged
Eredeti főcím: Ünnepi Könyvhét Szeged, 2024
Alcím: 95. ünnepi könyvhét és 23. gyermekkönyvnapok Szeged

Leírás forrása: Ünnepi Könyvhét Szeged, 2024
Forrás URL: https://konyvhetszeged.sk-szeged.hu/

Leírás dátuma: 2025-01-29
Leíró személy neve: Simon Jennifer

Archiváló szervezet: MNMKK Országos Széchényi Könyvtár
Archiváló szervezeti egység: Digitális Filológiai és Webarchiválási Osztály
Archiváló személy: Visky Ákos László
Archiváló projekt: MIA demó

Kiadó neve: Somogyi Károly Városi és Megyei Könyvtár
Felelős kiadó: Sikaláné Sánta Ildikó
Kiadó e-mail címe: library@sk-szeged.hu

Jogi státusz: jogvédett
Jogi közlemény weblapja: https://www.sk-szeged.hu/

Tartalmi jogtulajdonos: Somogyi Károly Városi és Megyei Könyvtár
Jogokkal kapcsolatos megjegyzés: A webhely a központból készült tartalomra vonatkozó szabályozás alapján került be a nyilvános archívumba, nem önálló engedélyeztetés útján, ezért a jogtulajdonosnak nincs azonosítója (content_rights_owner_id).

Műfaj: Általános honlap
Tulajdonosi típus: Szervezeti
Változékonyság: több havonta változó

Fő témakör: Kulturális források
Fő témakör: Oktatási források
Fő témakör: Humán szféra forrásai
Témakör: Társadalomtudomány
Témakör: Közművelődés
Témakör: Művészetek
Témakör: Ismeretterjesztés
Altémakör: Irodalom
Altémakör: Gyermekek
Altémakör: Játék
Tárgyszó: szépirodalom
Tárgyszó: kulturális rendezvény
Tárgyszó: könyvvásár
Tárgyszó: könyv
Tárgyszó: gyermekkönyv
Tárgyszó: szórakozás
Tárgyszó: ismeretterjesztés

Rövid leírás: A Szegeden évente megrendezett ünnepi könyvhét honlapja, melyet a Somogyi-könyvtár gondoz.

Kapcsolódó részgyűjtemény azonosítója: MIA_SET-00001

Ünnepi Könyvhét Szeged, 2020. szeptember

Egy újonnan készült részletes metaadat-leírás a nyilvános webarchívumban

Nyilvántartás, metaadatolás, statisztika

A decemberi webtér-szintű aratáshoz a régi WARC fájlokban levő weboldalakról kigyűjtött, korábban még nem ismert *.hu* végű domén és aldomén címek ellenőrzését befejeztük január közepén. Az eredeti 17 ezres listából első körben több mint 3 ezer webhely tűnt valóban újnak és működőnek, majd ezeket alaposabban megnézve és témájuk szerint besorolva végül kb. 2300 honlappal és bloggal tudtuk bővíteni a negyedévente aratott tematikus részgyűjteményeinket, valamint az elektronikus időszaki kiadványok weboldalainak nyilvántartását.

A Digitális Képarchívum két könyvtárosának segítségével elkezdtük a pótolni a demó archívumban a hiányzó metaadatokat, valamint kiegészíteni a korábbi egyszerűsített leírásokat. Az elmúlt hetekben kb. 50 új vagy módosított XML fájl készült el, melyek HTML-re konvertálva megnézhetők, illetve a fontosabb adatmezők szerint kereshetők is a honlapunkon.

Az egyes tematikus és webtér aratások után rendszeresen készített statisztikák mellett összeállítottuk a 2024. évi összesített táblázatokat és grafikonokat is, melyek az „Alapinformációk és -adatok" aloldalon nézhetők meg a webarchívum honlapján. Tavaly 94 db tömeges aratást sikerült lefuttatni, melyek során közel 1,2 milliárd URL-ről töltött le tartalmat a robot. Ennek csaknem a fele volt új vagy módosult fájl, ami azután tömörített WARC konténerekben eltárolásra került 28,8 terabájt összméretben. (2017 óta ez volt a legnagyobb éves növekmény, de ennek az is az oka, hogy a 2023. évi második webtér aratást technikai okokból csak 2024 elején tudtuk elvégezni.) Készült kb. 1 millió oldalkép is az archivált webhelyek kezdőlapjáról.

Informatikai ügyek

Az oldalképek készítéséhez használt Puppeteer böngészővezérlővel január óta gondok vannak, egy-két kép után hibaüzenettel leáll, így az idei tömeges aratásokat követően nem jöttek létre ezek a PNG fájlok. Az informatikusok vizsgálják a probléma okát.

Ugyancsak az ő segítségükkel zajlik a KIFÜ felhőjéből a NISZ-hez való átköltözés előkészítése. Mivel ennek pontos kezdése és az archívum anyagának átmásolásához szükséges idő még nem ismert, ezért átmeneti megoldásként beüzemelünk egy helyi szervert, ahol a fontosabb aratások továbbra is elvégezhetők lesznek. Továbbá a webarchívum saját webhelyeit szolgáltató WordPress rendszert is átmásoltuk egy tesztgépre, és egy ideig párhuzamosan frissítjük a honlapunkat mindkét szerveren.

Összeállítottunk egy tervet arról, hogy az új infrastruktúrán hogyan kellene kialakítani a webarchívum újragondolt rendszerét. Ennek egyik lényeges eleme, hogy amit lehet, azt konténerben kell futtatni, elkerülve ezzel a függőségekből származó kompatibilitási problémákat.

Ismeretterjesztés

2025. január 22-én a Webarchiválási Csoport vezetője, Kalcsó Gyula „Mi a »born digital« gyakorlat, és milyen változást hoz a kulturális intézmények életében?" címmel tartott előadást az MMA Művészetelméleti és Módszertani Kutatóintézete által szervezett műhelykonferencián. Jelentkezett továbbá a május 13-15. között a győri Széchenyi István Egyetemen megrendezésre kerülő Networkshop konferenciára, melyre „A magyar webtér aratásával kapcsolatos kurátori feladatok" címmel adott be absztraktot. A scrapingtechnológiáról szóló tavalyi NWS előadásának szerkesztett változata pedig nemrég jelent meg a konferencia kiadványában, amely az MTA Könyvtárának repozitóriumából, a REAL-ból tölthető le.

A Könyvtári Intézet által eredetileg február-márciusra meghirdetett, megújított tematikájú tanfolyamunkat későbbi időpontra, májusra kellett halasztani. Az egyes modulok tananyagai közül a WCT keretrendszert részletesen bemutató prezentáción dolgoztunk januárban, amihez 67 db képernyőfotó és egy, az angol szakkifejezéseket magyarázó táblázat is készült.

Az elmúlt hetekben lefutott aratások

Mezőgazdaság és élelmiszeripar (2813 db seed URL)
Kutatóintézetek, tudományos szervezetek (1354 db seed URL)
Egyetemek, főiskolák (4535 db seed URL)
Idegenforgalom, vendéglátás (8154 db seed URL)
Egészségügy, szociális szféra (9279 db seed URL)
Sport, testkultúra (3840 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (2257 db seed URL)
Irodalom, irodalomtudomány és -történet (1613 db seed URL)
Kulturális intézmények, művelődési házak, rendezvényhelyszínek (1043 db seed URL)
Párkapcsolat, család (2134 db seed URL)
Természet- és műszaki tudományok, szakterületek (2636 db seed URL)
Közoktatás és egyéb képzések (7515 db seed URL)
Bölcsészet- és társadalomtudományok, szakterületek (6397 db seed URL)
Képző-, előadó-, zene- és filmművészet (9370 db seed URL)
Életmód, szabadidő, hobbi (9488 db seed URL)
Szolgáltatás, kereskedelem, szállítás, közlekedés (11817 db seed URL)
Podkasztok (4259 db seed URL)
Elektronikus periodikák (11024 db seed URL)
Történelem, hely- és családtörténet (1541 db seed URL)
Média, sajtó, műsorszórás (1009 db seed URL)
Könyv- és egyéb kiadók, kereskedők (1572 db seed URL)
Kormányzat, önkormányzatok, politikai és civil szervezetek (7465 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: webarchivum@oszk.hu