

Az MNMKK OSZK Webarchívum 2025. márciusi és áprilisi hírei

Archiválás és nyilvántartás

Márciusban és áprilisban lefuttattuk a második negyedévre betervezett tömeges aratásokat az összes tematikus és a két rendszeresen mentett műfaji részgyűjteménynél, egyelőre még a régi infrastruktúrán. Jó hír, hogy a természet- és műszaki tudományok kivételével mindegyik rendben lefutott (és miután megismételtük, a TERMUSZ aratás is sikeresen befejeződött), sőt még a webhelyek kezdőlapjának oldalképei is elkészültek, pedig ezzel szintén gond volt korábban.

Csináltunk egy új mentést a nyilvános gyűjtemény valószínűsíthetően még frissülő 362 webhelyéről is a Web Curator Tool keretrendszerrel. Korábban ilyen mennyiségű aratást még nem indítottunk el a WCT-ben egyszerre, de annak ellenére, hogy a jelenlegi beállítás szerint párhuzamosan legfeljebb 10 job futhat, így is kb. 3 hét alatt befejeződött csaknem mindegyik. A szerver háttértárának betelése miatt az utolsó mentések nem záródtak le, de ettől függetlenül rendben lefutottak. A problémát egyelőre a túl nagyra nőtt naplófájl törlésével oldottuk meg.

Létrehoztunk egy eseményalapú részgyűjteményt a május 29-én induló Axiom Mission 4 küldetés apropóján, melynek magyar résztvevője is lesz Kapu Tibor személyében, aki a HUNOR Magyar Űrhajós Programban végez majd kutatásokat a nemzetközi űrállomáson. Az első alkalommal április végén, a Heritrix robottal learatott címlista 360 URL-t tartalmazott (nagyreszt címke alapján szűrt hírportálokat). Ezt az aratást egyelőre hetente futtatjuk, de május végén, amikor várhatóan megugrik majd a hírek száma, sűríteni fogjuk az ütemezést. Összegyűjtöttünk továbbá 28 egyedi weboldalt is ebben a témában, amiket az ArchiveWeb.page (AWP) programmal mentettünk le, valamint 26 darab közösségi média fiókot és hashtag alapú keresőlinket, melyeket szintén az AWP-vel archiválunk majd az űrmisszió ideje alatt.

Március 14-én elkezdtek különböző egyházak, gyülekezetek és egyéb vallási szervezetek Facebook fiókjait menteni, ami április 22-ig elég jó tempóban folyt, de ekkor a tulajdonos Meta cég kitiltotta az AWP-t a Facebookról és az Instagramról, így jelenleg ezekről a platformokról nem tudunk tartalmakat letölteni. Az öt hét alatt közel 400 Facebook fiókról készítettünk mentést, amelyek közül 261 esetben sikerült a legelső bejegyzésig visszagörgetni, vagy a teljes hírfolyamot archiválni. (A legkorábbi poszt dátuma 2009. november 9. volt.) A lementett anyagot az AWP saját WACZ formátumában is eltároltuk, de a visszakereshetőség érdekében kicsomagoltuk belőlük a tömörített WARC fájlokat, melyek összmérete közel 142 gigabájt lett.

Heti rendszerességgel folytattuk a „Könyvtári kihívás” programról szóló hírek és közösségi média posztok archiválását, de az AWP kitiltása ezt a munkát is akadályozza; a Facebook és az Instagram esetében jelenleg már csak az új bejegyzések számát tudjuk nyilvántartani, de nem lehet őket lementeni.

Március végével lezárult a podcast adások letöltésének aktuális fázisa. A hónap folyamán 232 csatornáról (köztük 42-ről első alkalommal) 12.473 darab hangfájlt archiváltunk, melyek összesített mérete 602 gigabájt volt.

Márciusban és áprilisban 63 új vagy bővített/javított, XML formátumú metaadat leírás került ki a honlapunkra és vált visszakereshetővé a nyilvános gyűjteményben levő webhelyekről.

Mindkét hónapban készült néhány egyedi mentés a HTTrack és az AWP programokkal olyan oldalakról, amikkel a Heritrix nem tudott megbirkózni, köztük egy-két, eddig még nem mentett OSZK-s webhelyről is. A nemzeti könyvtár saját online szolgáltatásainak mentéseit tartalmazó nyilvános weboldalra belinkel-tük az OSZK Indafotó fiókjából kiexportált képanyagot, mivel ez a népszerű képmegosztó szolgáltatás március végén bezárt.

File Edit View Window Help
REPLAY WEBPAGE / D:\ARCHIVEWEBPAGE_MENTES\HUNOR2025\hunor2025_weboldal_20250428.wacz / Axiom Mission 4
https://www.axiomspace.com/missions/ax4 2025. 04. 28. 18:05:47

AXIOM SPACE Axiom Station The Suit Missions Solutions About Shop

Axiom Mission 4

The Ax-4 mission will "realize the return" to human spaceflight for India, Poland, and Hungary, with each nation's first government-sponsored flight in more than 40 years. While Ax-4 marks these countries' second human spaceflight mission in history, it will be the first time all three nations will execute a mission on board the International Space Station. This historic mission underscores how Axiom Space is redefining the pathway to low-Earth orbit and elevating national space programs globally.

LAUNCH No earlier than May 2025

DURATION Up to 14 days docked to the ISS

AGENCIES Indian Space Research Organisation, European Space Agency, Hungarian Space Agency, Hungarian to Orbit Program

Meet the Astronauts



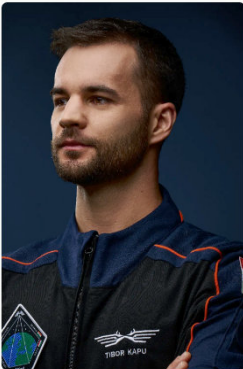
Peggy Whitson
Commander 



Shubhanshu Shukla
Pilot 



Sławosz Uznański-Wisniewski
Mission Specialist 



Tibor Kapu
Mission Specialist 

A május 29-ére tervezett űrmisszió hivatalos weboldalának mentése a ReplayWeb.page megjelenítőben

Egyéb ügyek

Elindult egy egyeztetési folyamat arról, hogy az MNMKK OSZK webarchívuma milyen módon és milyen feltételekkel kaphatná meg az Indafotó platformján eddig nyilvánosan szolgáltatott képeket és azok metaadatait.

A University of London doktoranduszától, Beatrice Cannellitől megkaptuk és áttanulmányoztuk az „Archiving Social Media: A Comparative Study of the Practices, Obstacles, and Opportunities Related to the Development of Social Media Archives” című PhD dolgozatát, melyben több nemzeti könyvtár webarchiválási gyakorlatát, azon belül pedig elsősorban a közösségi média tartalmak megőrzésének és szolgáltatásának módszereit hasonlítja össze.

Az ELTE BTK egyik informatikus könyvtáros hallgatója „A nyilvános internetes tartalmak gyűjtésének, megőrzésének és szolgáltatásának jogi kérdései” címmel készítette el alapszakos szakdolgozatát, aminek külső konzulense az OSZK webarchívumának kurátora volt. A dolgozatban a szerző az internet és a web, valamint a webarchiválás történetének rövid összefoglalása után részletesen bemutatja a webtartalmak archiválásának jogi hátterét, valamint a szerzői jogi védelem és a kötelezpéldányon alapuló hosszú távú megőrzés ellentmondásos helyzetét a digitálisan születő tartalmak esetében.

A honlapunkról korábban letölthető felhasználási szerződésminták már nem érhetők el, mert változni fog a nyilvános szolgáltatáshoz szükséges szerződések megkötésének munkafolyamata. A webhelyek tulajdonosainak a jövőben csak egy űrlapot kell majd kitölteniük, ahol megadják a szükséges adatokat, magát a szerződést pedig az OSZK munkatársa állítja össze és küldi el aláírásra.

A május végére meghirdetett „Internetes tartalmak archiválása” tanfolyamunk várhatóan nem indul el, mert kevés a jelentkező, de ettől még tovább dolgoztunk a Web Curator Tool keretrendszer egyes funkcióit bemutató prezentáción. A tanfolyamot ősszel ismét meghirdeti a Könyvtári Intézet, reményeink szerint már újra térítésmentesen.

Az IIPC konzorcium márciusban két online webináriumot is szervezett, melyeket mi is megnéztünk és sok új információt szereztünk. A március 17-i közvetítés címe „PANDORÆ and Hyphe: Open-Source Tools and Methodologies for Researching Web Archives” volt, és két, francia fejlesztésű, nyílt forráskódú eszközt mutattak be a készítőik és a használói, amelyekkel – elsősorban társadalomtudományi kutatásokhoz használható – webes korpuszok építhetők és elemezhetők. Három nappal később „Regex Club Live: Collaborative Exploration of Crawl Logs and Crawler Traps” címmel tartottak egy érdekes beszámolót az Egyesült Királyság Nemzeti Levéltárának munkatársai arról, hogy hogyan vizsgálják meg reguláris kifejezések segítségével a webaratások során keletkező naplófájlokat, így próbálva felderíteni a hiányokat és a robotcsapdákat, hogy a konfigurációs fájlok módosításával a jövőben teljesebb és kevésbé „szemetes” mentéseket tudjanak csinálni.

Az elmúlt hetekben lefutott aratások

Vallások, hitrendszerek, egyházak (3089 db seed URL)
Sport, testkultúra (3841 db seed URL)
Egyetemek, főiskolák (4539 db seed URL)
Kutatóintézetek, tudományos szervezetek (1355 db seed URL)
Mezőgazdaság és élelmiszeripar (2813 db seed URL)
Idegenforgalom, vendéglátás (8154 db seed URL)
Egészségügy, szociális szféra (9283 db seed URL)
Kulturális intézmények, művelődési házak, rendezvényhelyszínek (1046 db seed URL)
Irodalom, irodalomtudomány és -történet (1621 db seed URL)
Természet- és műszaki tudományok, szakterületek (2642 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (2276 db seed URL)
Párkapcsolat, család (2134 db seed URL)
Bölcsészet- és társadalomtudományok, szakterületek (6397 db seed URL)
Közoktatás és egyéb képzések (7515 db seed URL)
Képző-, előadó-, zene- és filmművészet (9385 db seed URL)
Történelem, hely- és családtörténet (1575 db seed URL)
Média, sajtó, műsorszórás (1012 db seed URL)
Könyv- és egyéb kiadók, kereskedők (1574 db seed URL)
Szolgáltatás, kereskedelem, szállítás, közlekedés (11839 db seed URL)
Életmód, szabadidő, hobbi (9506 db seed URL)
Podkasztkok (4269 db seed URL)
Kormányzat, önkormányzatok, politikai és civil szervezetek (7520 db seed URL)
Elektronikus periodikák (11065 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