

## Az OSZK Webarchívum 2022 októberi hírei

### Archiválás

A nyilvános gyűjteményben levő webhelyek archiválásához használt Web Curator Tool rendszert sikerült újra elérhetővé tenni az augusztusi leállítás után, így ismét a beállított – többnyire negyedéves – ütemezés szerint folyik a honlapok és blogok mentése.

Lezárult a magyar podkasztkok hanganyagának hosszú távú megőrzésére tett első kísérletünk, melyet májusban kezdtünk el. Az azóta felderített 1287 csatorna közül különféle segédprogramokkal 1243-ról sikerült letölteni adásokat: több mint 75 ezer hangfájl, 4,2 terabájt összméretben. A legrégebbiek 2006-ban kerültek fel az internetre és szerencsére még nem tűntek el teljesen. A gyűjtemény bővítését és aktualizálását jövőre is folytatni szeretnénk, de már kisebb volumenben. A hanganyagok archiválása mellett természetesen a podkasztkok és podkasztt gyártók weboldalait és RSS feed-jeit is mentjük a Heritrix szoftverrel, legutóbb október 21. és 24. között, amikor is közel 2,5 millió URL címről töltött le a robot 393 gigabájtnyi új tartalmat.

A hónap közepén elkezdtük egy újabb tematikus részgyűjtemény kialakítását bölcsész- és társadalomtudományi jellegű webhelyekből. A felsőoktatási és a kutatóintézeti gyűjtemények részeként már eddig is sok ilyen honlapot és blogot arattunk, így most olyan oldalakat válogatunk be, amelyek ezekből kimaradtak, illetve amelyek üzleti szolgáltatást nyújtanak valamelyik területen. (Például a JOG alá felvesszük az ügyvédi irodákat, a KÖZGAZDASÁG alá pedig a menedzsment tanácsadó vagy marketing cégeket is.) Már közel 2,5 ezer címet sikerült összegyűjteni, de ez a munka még folytatódni fog legalább az év végéig.

### Kutatás

Telefonos megkeresést kaptunk az ELTE Germanisztikai Intézetéből. A koronavírus járványhoz kapcsolódó, az Innsbrucki Egyetemmel közös kutatási projektjük keretében szeretnének hozzájutni elemzési célból a blikk.hu-ról archivált cikkekhez. A részleteket és a feltételeket egy online megbeszélés keretében egyeztettük a hónap végén, az adatbányászattal és -elemzéssel foglalkozó DBK-s kollégák bevonásával.

A csak helyben használható, a zárt webarchívumhoz is hozzáférést biztosító dedikált számítógépek átkerültek a felújított déli olvasóterembe, amit azonnal sikerült élesben is letesztelni. Egy olvasó a Magyar Nemzeti Bank honlapjáról szeretett volna visszakeresni egy feltöltött dokumentumot, amiben munkatársunk volt segítségére. Bár az MNB honlap jól archiválható, a keresett dokumentum a kezdőlaptól számítva olyan „mélységben” volt, ami már kívül esett az akkor használt archiválási paramétereken. Az eset rámutatott arra, hogy nemcsak adatbányászati célokra hasznosítható egy webarchívum, hanem valós igény a „hagyományos” kutatási kérések kiszolgálása is. Egyúttal össze-csengett azzal a tervünkkel is, hogy élve a kormányrendeletben kapott jogosultsággal, tömegesen szolgáltatassuk a közpénzből megvalósuló webtartalmak archivált változatát, könnyebbé téve a felhasználók számára ezek elérését. Ez a feladat az eddigi nyilvános szolgáltatás átalakításával is jár, ennek előkészítése, a technikai feltételek felmérése már elkezdődött.

### 404-es workshop

Élesre fordult a végül december 8-ra átkerült „404 Not Found workshop” szervezése: véglegesítettük az előzetes tervekhez képest újratervezett programot és elkezdődött a kapcsolódó feladatok (grafikai anyag készítése, meghívandók névsorának összeállítása, technikai feltételek megteremtése stb.)

kivitelezése is. A rendezvény végül a digitális megőrzés témáját, valamint az ehhez kapcsolódó hazai és nemzetközi archiválási és tudományos projektek ismertetését helyezi középpontba: szó lesz az OSZK-ban és a nemzetközi szervezeteknél zajló webarchiválási tevékenységről, valamint az előállt digitális tartalom kutatási lehetőségeiről, különös tekintettel a digitális bölcsészet szakterületeivel való kapcsolódási pontokra. A délutáni workshopon élő bemutató lesz a metaadatolásról, az egyedi webhelyek és weboldalak archiválásáról, és a webarchívumok keresőinek használatáról. A részletes program a hivatalos közzétételből tudható majd meg.

### **Együttműködések**

Mivel eredetileg össze terveztük a megyei könyvtárakkal formálódó együttműködések véglegesítését, megkerestük a formális válasszal még adós intézmények vezetőit a végleges álláspontjuk felmérésére. Sajnos az idei őszön más sürgős ügyekkel kell foglalkozniuk a könyvtáraknak (is), így nem is csodálkoztunk a kevés beérkező válaszon. Mivel a kollégák részéről is merült fel kérdés az együttműködés tervezett formájáról, amit előbb nekünk is tisztázni kell, így a szerződéseket idén már nem tudjuk tető alá hozni. Emiatt a „404-es” workshopon az együttműködéshez kötődő programot is újra kellett tervezni. A halasztástól függetlenül még volt egy megbeszélés a pécsi Csorba Győző Könyvtár munkatársaival, aminek folytatása a Pécsi Tudományegyetem Egyetemi Könyvtár és Tudásközpont bevonásával várható.

### **Bemutató**

Október 10-én a „Könyvtár, ami összeköt” program keretében az OSZK-ba látogató kárpát-medencei könyvtárosok számára egy rövid bemutatót tartottunk a webarchiválásról, melynek prezentációja letölthető a honlapunkról.

### **Rendezvények**

Előadással jelentkeztünk az ELTE BTK Digitális Bölcsészet Tanszék által szervezett nemzetközi konferenciára, amire novemberben kerül majd sor. Felkértek minket a webarchiválásról szóló kerekasztal-beszélgetésen történő részvételre is.

Sajnos a pozsonyi Egyetemi Könyvtárban szintén novemberre tervezett, a hosszú távú digitális megőrzéssel foglalkozó konferenciát, amelyre már az előadásunk szövegét is leadtuk, nem rendezik most meg, jövőre kerül majd rá sor.

Elfogadták a webarchívumokra irányuló tudományos kutatások kapcsán tartandó előadásunkat a 2023. januári BOBCATSSS konferenciára, mely hibrid formában kerül megrendezésre, így mi nagy valószínűséggel online fogjuk megtartani az előadást.

Októberben három külföldi konferencián is képviseltettük magunkat Németh Márton kollégánk személyében. A Cseh Tudományos Akadémia Irodalomtudományi Intézete „Mutual Learning Workshop to Improve Cultural Heritage Bibliographical Data” címmel rendezett háromnapos konferenciát és workshopot Prágában 2022. október 12-14. között. A rendezvényen könyvtáros szakemberek és tudományos kutatók részéről 18 előadás hangzott el a különféle digitalizált, illetve digitálisan születt forrásokhoz kapcsolódó metaadat- és bibliográfiai modellekről, és számos konkrét projekt került bemutatásra, hogy miként lehet bibliográfiai adatokat kutatási adatként felhasználni. A rendezvényt hibrid formában rendezték, az előadók és a nézőközönség soraiból is többen online csatlakoztak. Németh Márton a helyszínen a webarchívumok metaadatolásáról adott áttekintést, bemutatta az OSZK jelenlegi gyakorlatát és terveit is.



A WARCnet projektet lezáró rendezvényen készült fotó

A webarchívumokra, illetve webarchiválásra irányuló tudományos kutatásokkal foglalkozó WARCnet projekt zárókonferenciájára a dániai Aarhusban került sor 2022. október 17-18. között. Még a konferencia hivatalos kezdete előtt október 16-án este már munkamegbeszélést és vacsorát tartottak, melyen kiderült, hogy küszöbön áll a projekt eredményeit összegző kötet kiadásáról szóló szerződés aláírása. Ez lehetőséget nyújt majd az orosz-ukrán háború itthoni visszhangjával kapcsolatos webarchiválási tevékenység, illetve a gyűjtemény tudományos hasznosításával (természetes nyelvi elemző eszközök használata, adatvizualizáció) kapcsolatos tapasztalatok megosztására. Emellett tervben van egy nemzetközi körkép összeállítása is arról, hogy a webarchiválás miképpen jelenik meg a különféle tudományterületeket érintő felsőoktatási képzésekben.

A Covid-19 járvány miatt három év után a pozsonyi Comenius Egyetem Könyvtár és Információ-tudományi Tanszéke által október 20-21-én újra megrendezett Information Interactions konferencia idei fő témái a digitális információs környezetben jelentkező etikai kihívások voltak. Ez az esemény is hibrid formában került megrendezésre, számos cseh és lengyel résztvevő online tartotta meg előadását. A konferencián az első nap és a második nap délelőttjén tudományos előadások hangzottak el, melyek a digitális információ tárolásával, felhasználásával, szolgáltatásával, illetve a mesterséges intelligencia használatával kapcsolatos etikai dilemmákra fókuszáltak. A második nap délelőttjén az előadások folytatódtak, délután pedig doktori hallgatók ismertették kutatási programjaikat, amelyek azután fórum keretében kerültek megvitatásra. Kollégánk az első nap délutánján elhangzott előadásában a nemzeti könyvtárban zajló webarchiválási tevékenységről írt PhD dolgozatát ismertette, különös tekintettel az archívumok kutatásának erkölcsi dimenziójára.

Mindhárom fenti rendezvényről részletesebb összefoglaló is készült az OSZK blogja számára.

### **Az elmúlt hetekben lefutott tematikus aratások**

Történelem, hely- és családtörténet (771)

Podkasztkok (3742)

Média, sajtó, műsorszórás (878)

Könyv- és egyéb kiadók, kereskedők (1518)

Kormányzat, önkormányzatok, politikai és civil szervezetek (6292)

Idegenforgalom, vendéglátás (5928)

A tematikus 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: [mia@mek.oszk.hu](mailto:mia@mek.oszk.hu)