

Az OSZK Webarchívum 2022 februári hírei

Archiválási munkák

Februártól kezdve a tematikus aratásoknál a korábbi háromról ötre növeltük azoknak az ugrásoknak a számát, amiket az archiváló szoftver a webhelyek kezdőlapjáról elindulva és a linkeket követve megethet, továbbá a több ezer URL-t tartalmazó részgyűjtemények esetében a futásidőt is nagyobbra vettük. A cél az, hogy az elmúlt évekhez képest idén nagyobb mélységben és nagyobb mennyiségben mentsünk fájlokat ezekről az oldalakról. Az eddigi tapasztalatok szerint a korábbiakhoz képest így körülbelül kétszer annyi tartalmat tölt le a robot darabszámban és méretben is, mielőtt eléri a fél terabájtos felső korlátot vagy lejár az aratásra szánt idő. Az elektronikus periodikák oldalai és az egyetemek honlapjai esetében már eddig is magasabb értékeket használtunk, tekintve, hogy ezek jellemzően nagy méretű webhelyek, és ezeket a paramétereket terjesztettük most ki a többi témakörre is, mert úgy ítéltük meg, hogy sok értékes tartalmat nem értünk el a három szint mélységű aratásokkal.

A címlisták karbantartása folyamatos volt ebben a hónapban is, főként a közgyűjtemények és a történelem témájúak bővültek több tucat címmel, az e-periodikák esetében pedig több mint 80 új seed URL-t vettünk fel. Az iskolai honlapoknál átnéztük azt a 93 címet, amelyek az őszi ellenőrzésnél még a megszüntetés alatt levő sulinet.hu doménon voltak, és ahol sikerült megtalálni az új webhelyet, ott javítottuk a címet, a régit pedig átsoroltuk a megszüntek közé. Rengeteg új tétellel bővítettük a KORONAVIRUS2020 jelzésű, heti rendszerességgel aratott részgyűjteményünk seed listáját is. A koronavírus okozta járvánnyal kapcsolatos híreket és egyéb online forrásokat 2020 februárjában kezdtük el archiválni. Az akkor összeválogatott 133 tételes címlista mára nagyrészt elavult és időközben sok újabb információforrás, valamint címke jelent meg, ezért már nagyon időszerű volt a frissítése. A jelenlegi listában 328 tételt (1012 URL címet) tartunk nyilván, típus szerint csoportosítva, melyből 48 közösségi média oldal (főként Facebook csoportok). Utóbbiakról a Windows alatt futó ArchiveWeb.Page szoftverrel egyedi mentéseket csináltunk a hónap közepén. Február 21-én elindítottuk egy újabb esemény-alapú gyűjtemény aratását: az orosz-ukrán konfliktussal kapcsolatos híreket mentjük hétfőnként 74 forrásból (115 keresőlinkről). A címlisták elérhetősége: <https://webarchivum.oszk.hu/esemeny-alapu-aratasok/>

Nyilvános gyűjtemény

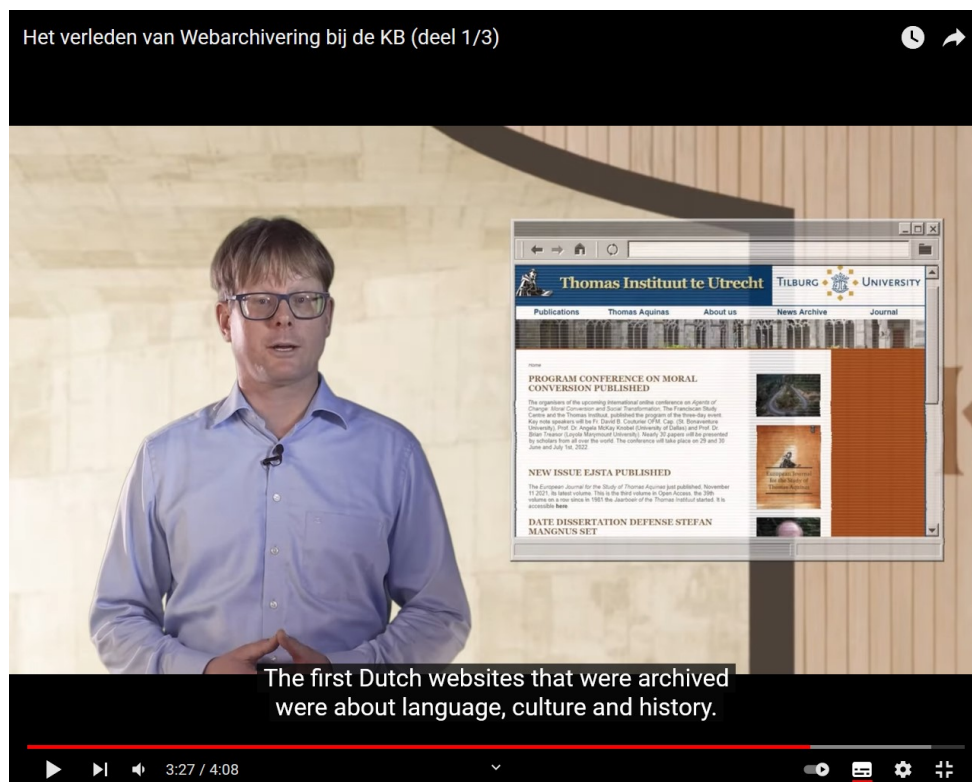
Folytatódott a lejáró felhasználási engedélyek megújítása. Emlékeztető leveleket írtunk azoknak a tartalomgazdáknak, akik még nem küldték el a szükséges iratot az OSZK-nak és ennek köszönhetően több szerződés is beérkezett a hónapban. Sajnos még mindig vannak olyanok, akik nem reagáltak a megkeresésünkre, így várhatóan 9 partnerünk 16 webhelyét kell kivennünk a nyilvános szolgáltatásból. Természetesen archivált anyagot nem törölünk, csak technikailag elérhetetlenné tesszük, ezért egy esetleges későbbi hozzájárulás után minden visszaállítható lesz. Tekintve, hogy 70 jogtulajdonos 94 webhelyére kellett megújítanunk az engedélyt és mindig számolni kell figyelmetlenségből, felelősségből fakadó veszteséggel, egészen jó arányt sikerült elérnünk. Nagy segítség volt és nagyon hasznosnak bizonyult, hogy a 626/2020. (XII. 22.) számú kormányrendeletben kapott felhatalmazás következtében, mintegy 30 „közpénzes” intézmény 111 honlapja esetében már nem kellett lefolytatnunk az engedélyeztetési procedúrát. A szerződések megújításának pedig volt egy olyan pozitív hozadéka, hogy új oldalakra is kaptunk felhasználási engedélyeket, összesen 6 db-ot. Ezekről elkészítettük az első mentéseket és beütemeztük folyamatos archiválásukat.

Ismeretterjesztés

A 2021. november 23-án és 24-én tartott „404 Not Found – Ki őrzi meg az internetet?” című online rendezvényünk előadásai – beleértve a Webarchiválási Osztály munkatársai által tartottakat is – rövidesen megjelennek cikk formátumba átdolgozva a Könyvtári Figyelő és a Könyv, Könyvtár, Könyvtáros folyóiratokban, Németh Márton pedig külön tanulmányban foglalta össze a nemzetközi napon elhangzott előadásokat és beszélgetéseket. Néhány napja már a konferencia videofelvételei is megtekinthetők a Videotoriumban a <https://videotorium.hu/hu/channels/5339/> csatornán.

Nemrég értesültünk róla, hogy felkerültek a YouTube-ra a lengyel Nemzeti Levéltár által 2021. november 16-án és 17-én szervezett és részben helyben, részben pedig online megtartott webarchiválási workshop videofelvételei, köztük Németh Márton „Brief introduction to the Hungarian Web Archive at the National Széchényi Library” című előadása (<https://www.archiwa.gov.pl/pl/157-forum-archiwizacji-zasobow-polskiego-internetu>).

A holland nemzeti könyvtár a január 31. és február 4. között lezajlott „webarchiválási hét” keretében egy három részből álló látványos sorozatot jelentetett meg a YouTube-on, melyben Kees Teszelszky, a digitális gyűjtemény magyar származású kurátora beszél a holland web kezdeteiről, az archiválás fontosságáról, a válogatási elvekről, a részgyűjteményekről és a jövőbeli tervekről. Hollandiában 15 évvel ezelőtt indult a szelektív webarchiválás, jelenleg már 21 ezer webhelyet mentenek, viszont webtér szintű aratásokra nincs lehetőségük az ottani jogi környezet miatt. A videók elérhetősége (a fogaskerék ikonra kattintva bekapcsolható az automatikus angol vagy magyar feliratozás.):
<https://www.youtube.com/watch?v=WGZltMnSQgE>
https://www.youtube.com/watch?v=ITK7eW_M14E
<https://www.youtube.com/watch?v=Seq44U-jXJE>



Részlet a holland ismeretterjesztő videóból

Segítségnyújtás

Február elején kereste meg a Webarchívumot az ELTE egyik végzős mesterszakos hallgatója, segítséget kérve szakdolgozata anyaggyűjtéséhez, melynek témája a koronavírus járvány egyes hullámainak reprezentációja az internetes sajtóban. Online csatornán keresztül beszéltük át a problémáját és hogy mi miben tudunk segíteni neki. Végül inkább jótanácsokkal tudtuk ellátni, többek között javasoltuk az ELTE Digitális Bölcsészet Tanszékével való kapcsolatfelvételt, mert az ott folyó tudományos munka közvetlenebb segítséget ígér, de emellett természetesen készséggel állunk rendelkezésére a továbbiakban is, ha szükséges.

Nemzetközi kapcsolatok

Február 16-án egy újabb „IIPC Updates” zajlott le egy Zoom csatornán, mintegy 30 résztvevővel. Egyrészt tájékoztatást kaptunk a nemzetközi szervezet megújult vezetéséről, az idén várható rendezvényekről, a téli olimpiával és paralimpiával kapcsolatos online hírforrások közösen összeállított archívumáról, másrészt rövid beszámolók hangzottak el egyes intézmények részéről a webarchiválást érintő terveikről és új eredményeikről. A Skót Nemzeti Könyvtár, az amerikai Kongresszusi Könyvtár és az Észak-Texasi Egyetem munkatársai mellett mi is szót kaptunk és ismertettük az Internet Archive és az OSZK között decemberben végén megkötött szerződés főbb elemeit. A résztvevők közül többen is gratuláltak ahhoz, hogy a magyar webarchiváló projekt néhány év alatt milyen messzire jutott.

Az elmúlt hetekben lefutott tematikus aratások

Közoktatás és egyéb képzések (6.506 db seed URL)
Könyvtárak, levéltárak, múzeumok és galériák (1.949 seed URL)
Irodalom, irodalomtudomány és -történet (1.386 db seed URL)
Képző-, előadó-, zene- és filmművészet (8.066 db seed URL)

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