

Az OSZK Webarchívum 2023 májusi hírei

Archiválás és nyilvántartás

Mivel a tömeges aratásokat végző és a zárt archívumot szolgáltató két szerver közös tárterülete csaknem betelt, ezért egy kétszer nagyobb, 300 terabájtos tárhelyre kezdtük el átköltöztetni a gyűjteményt. Amíg a másolás zajlik, nem indítunk újabb mentéseket és május közepétől az ütemezett esemény-alapú aratásokat is leállította az informatikus.

Az április végi pápalátogatással kapcsolatos híreket a 2023-04-25 és 2023-05-16 közötti időszakban mentettük, május 2-án pedig közösségi médiaoldalakat is archiváltunk a Facebookról, az Instagramról, a Twiterről, Youtube-ról és a TikTokról (utóbbiaknál magukat a videókat nem töltjük le, csak a videók listáit).

Elkezdünk egy újabb eseménnyel kapcsolatos részgyűjteményt kialakítani „Európa Kulturális Fővárosa- 2023” címmel, melyben az idei nyertes Veszprém város és a Balaton Régió rendezvényeiről szóló híradásokat próbáljuk majd összegyűjteni, valamint archiváljuk a hivatalos honlapokat és a közösségi média oldalakat is.

Az „Életmód, szabadidő, hobbi” elnevezésű – szintén még csak gyűjtés alatt levő – válogatás címlistája már 6500 tételes. A hónap folyamán több mint 1500 címmel bővült, főleg életmód tanácsadás, coaching és ezoterikus témájú webhelyekkel. Az elektronikus periodikák nyilvántartásába 50, eddig még nem ismert kiadvány oldala került be, és a többi tematikus címlistát is bővítettük néhány száz címmel. A legfontosabb ezek közül a közgyűjtemények listája, melynél altémák szerint is besoroltuk az elmúlt hónapokban összegyűlt címeiket. Szintén elmaradt feladat volt az első alkalommal márciusban aratott „Egészségügy, szociális szféra” részgyűjtemény metaadat rekordjának az elkészítése, melyet most sikerült pótolni, így a többi tematikus válogatás leírásával együtt már ez is elérhető a honlapunkról.

A publikus demó oldalunkra 21 újonnan archivált webhely került ki, főként kormányzati honlapok, melyek szolgáltatásához nem kell külön engedélyt kérni. Az elsőre hibás vagy hiányos mentéseknél megpróbáltunk változtatni a Heritrix robot beállításain, a még ezután is problémás webhelyekről pedig készítettünk egy alternatív mentést is a HTTRack programmal.

A nyár közepén lesz a következő webtér szintű aratás, amely minden eddig felderített doménre és aldoménre kiterjed. Ennek kapcsán írtunk egy összefoglalást arról, hogy jelenleg hogyan zajlik ennek a hatalmas címlistának a nyilvántartása, szűrése és frissítése, és elkezdtük megtervezni a megújítását. A jövőben a bővítéshez jelentős segítség lesz majd az előző évben a .hu alá bejegyzett domének listája, melynek átadására ígéretet kaptunk az Internet Szolgáltatók Tanácsától.

IIPC rendezvény

Májusban tartotta meg taggyűlését és konferenciáját az International Internet Preservation Consortium, mely idén ünnepli a 20. születésnapját. (Az OSZK 2018-ban csatlakozott a szervezethez.) A rendezvény idén újra jelenléti formában valósult meg a hollandiai Hilversumban, de előtte egy héttel, május 3-án egy online napot is szerveztek, amin az előre feltöltött előadásvideókhoz kapcsolódó kérdéseket lehetett feltenni. Az IIPC irányítóbizottsága május 9-én a Királyi Könyvtárban tartott ülést, másnap pedig a taggyűlésre került sor a Holland Hang- és Képarchívum épületében. Az egyes munkacsoportok beszámolóit után a tagintézmények képviselői is szót kaphattak. A magyar webarchívum újdonságairól dr. Kalcsó Gyula, a Digitális Filológiai és Webarchiválási Osztály webarchiválási csoportvezetője tartott tíz perces prezentációt, akinek a rendezvényről, az érdekesebb előadásokról és workshopokról írt beszámolója az OSZK blogjában olvasható (<https://nemzetikonyvtar.blog.hu/2023/05/23/>).



Az IIPC idei taggyűlése

(A kép forrása az IIPC blogja: <https://netpreserveblog.files.wordpress.com>)

Egyéb ügyek

Szintén az IIPC-hez kapcsolódó hír, hogy kérésünkre frissült az OSZK webarchívumról szóló tájékoztató a szervezet honlapján (<https://netpreserve.org/about-us/members/orszagos-szechenyi-konyvtar/>).

A saját honlapunkon is aktualizáltunk pár dolgot: a projektet bemutató szórólapot, valamint a munkatársak és a közreműködők listáját, továbbá megújítjuk a szakirodalmi bibliográfiát is. Ezt korábban a Zoreroból exportáltuk ki többféle formátumban évente 1-2 alkalommal, de ezentúl naprakészen lesz elérhető a WordPressbe beágyazva.

Utolsó ellenőrzés céljából megkaptuk a webhelyek archiválási problémáiról szóló, Drótos László és Visky Ákos László által írt cikkünk betördelt változatát, ami remélhetőleg rövidesen megjelenik a Könyvtári Figyelőben.

Május 18-án az ELTE BTK Könyvtár- és Információtudományi Intézet BA képzésének „Tájékoztatás és információforrások” nevű kurzusán résztvevő hallgatók látogattak el az OSZK Digitális Bölcsészeti Központjába, ahol a digitális filológiáról és a webarchiválásról néztek meg egy-egy bemutatót.

Az elmúlt hetekben lefutott tematikus és webtér aratások

Irodalom, irodalomtudomány és -történet (1511 db seed URL)

Természet- és műszaki tudományok, szakterületek (2109 db seed URL)

Könyvtárak, levéltárak, múzeumok és galériák (1809 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: webarchivum@oszk.hu