404 Not Found – Ki őrzi meg az internetet? konferencia és kerekasztal-beszélgetés

2026/11/18
2026/10/27

November 18-án tizedik alkalommal kerül megrendezésre a webarchiválással és a digitális megőrzéssel foglalkozó konferencia és kerekasztal-beszélgetés.



[1]

404 Not Found – Ki őrzi meg az internetet?
A webarchiválás és a born digital tartalmak megőrzésének aktuális problémái
Konferencia és kerekasztal-beszélgetés

2026-ban tizedik alkalommal kerül megrendezésre a webarchiválással és a digitális megőrzéssel foglalkozó ***404 Not Found – Ki őrzi meg az internetet?* című konferencia és kerekasztal-beszélgetés.** A tizedik alkalom lehetőséget ad a hazai webarchiválás történetének áttekintésére és

az eddigi tapasztalatok összegzésére, emellett a digitálisan létrejött tartalmak aktuális szakmai kérdéseire is keressük majd a válaszokat. A korábbi évek programjaihoz képest változás, hogy idén a délutáni workshop helyett egy, az előadások témáihoz kapcsolódó kerekasztal-beszélgetéssel zárjuk a rendezvényt.

Időpont: **2026. november 18., szerda**

Helyszín: OSZK, V. emelet, **516-os terem**
(Budavári Palota F épület, 1014 Budapest, Szent György tér 4–5–6.) [2]

Fb-esemény [3]

A részvétel ingyenes, de regisztrációhoz van kötve.

REGISZRÁCIÓ [4]

Az érdeklődők **online** is bekapcsolódhatnak az élő közvetítésbe
az OSZK YouTube-csatornáján. [5]

PROGRAM

10.00 Köszöntők

Előadások

10.20 A Common Crawl előadói: *Who Gets Crawled? Coverage and Policies at Common Crawl*
(Mit mentünk? Gyűjtőkör és alapelvek a Common Crawlnál)

10.40 Drótos László (OSZK): *Az eltűnő web nyomában – Tíz év sikerei, buktatói és tanulságai a nemzeti könyvtár webarchívumában*

11.00 Kalcsó Gyula (OSZK): *A nemzeti könyvtár webarchiválási tevékenységének aktuális problémái*

11.20 Kávészünet

11.40 Bánki Zsolt (MNL): *Az eltűnő emlékezet – évtizedes veszteségek a kulturális örökségben*

12.00 Keller-Gera Adrienn (BTM): *A born digital tartalmak kezelésének jogi kihívásai*

12.20 Németh Márton (DE): *Webarchiválás a könyvtár- és információtudományi felsőoktatásban*

12.40 Ebédszünet

13.30 Sárközi-Lindner Zsófia (ELTE BTK): *Régi-új kérdések: webes források és archívumelmélet egymás tükrében*

13.50 Ferenci Tamás: *Lelkes amatőrként a digitális archiválás világában*

Kerekasztal-beszélgetés

14.10 *A webarchiválás és a born digital tartalmak megőrzésének aktuális problémái – Múlt, jelen és jövő*

Résztvevők: **Bánki Zsolt** (MNL), **Drótos László** (OSZK), **Horváth Adrienn** (VMK), **Kalcsó Gyula** (OSZK), **Moldován István** (PPKE BTK)

15.40 A konferencia zárása: kérdések, hozzászólások, vita

Az előadások címe és rövid leírása

A Common Crawl előadói: *Who Gets Crawled? Coverage and Policies at Common Crawl* (Mit mentünk? Gyűjtőkör és alapelvek a Common Crawlnál)
2008 óta a Common Crawl több mint 300 milliárd weboldalt gyűjtött össze, havonta körülbelül 2,5 milliárd oldallal bővítve a gyűjteményét. Kutatók, archiválók és fejlesztők világszerte használják ezt az ingyenes archívumot. Minden havi mentésnek van egy rögzített kerete, és mindennap sokkal több URL-t talál a keresőrobotunk, mint amennyit letölthet. Ezért a gyűjtendő tartalom kiválasztása a munka központi eleme. Hogyan hozza meg a Common Crawl ezt a döntést, és hogyan válik egyre nehezebbé ez a feladat? Az internet többnyelvű, de az azt megőrző mentések ritkán azok. A linkalapú rangsorolás a sok kapcsolattal rendelkező webhelyeket részesíti előnyben, így az angol és néhány nagy nyelv viszi el a költségvetés nagy részét, míg a kisebb nyelvi közösségek háttérbe szorulnak. Mi lenne szükséges ahhoz, hogy ez megváltozzon?
Ez az előadás a Common Crawl legújabb, egy kiegyensúlyozottabb archívum létrehozására irányuló törekvését mutatja be, beleértve a közösségi erőfeszítéseket az alulreprezentált nyelvek felkutatására, a zajos webes szövegek jobb nyelvi azonosítását, valamint azokat a kísérleteket, amelyek a mentések kereteit a nem angol nyelvű tartalmak felé tágítják. Ugyanakkor vajon bezárja-e kapuit a web? A generatív mesterséges intelligencia térnyerése óta egyre több webhely blokkolja a keresőrobotokat, gyakran anélkül, hogy különbséget tennének az archívumok, a keresőmotorok és a mesterséges intelligenciával foglalkozó cégek között. Ezeknek a preferenciáknak a kifejezésére vonatkozó szabályokat jelenleg átírják, többek között olyan szervezeteknél is, mint az Internet Engineering Task Force. Ki dönti el, hogy mit őriznek meg, és milyen feltételek mellett?

Drótos László (OSZK): *Az eltűnő web nyomában – Tíz év sikerei, buktatói és tanulságai a nemzeti könyvtár webarchívumában*

Az Országos Széchényi Könyvtárban 2017 elején kezdődtek el a magyar nemzeti webarchívum előkészítő munkálatai. Bár a környező országok többségéhez képest késésben voltunk ezen a területen, de a digitálisan születő tartalom megőrzésében már több évtizedes tapasztalattal rendelkező szakemberek kezdték el a webarchiválás infrastruktúrájának és munkafolyamatainak kialakítását. Ennek az előzetes tudásnak és a gyakorlati megvalósítás megkezdését meglepő gyorsasággal követő jogi szabályozásnak köszönhetően néhány év alatt sikerült ledolgozni a hátrányunkat, sőt nemzetközileg is elismert szolgáltatást kialakítani. Természetesen a sikerekhez nagyban hozzájárult az a szoftverfejlesztő munka is, amit a külföldi kollégák már korábban elvégeztek, és az a „know-how”, amit készségesen megosztottak velünk, miután az OSZK 2018-ban csatlakozott a International Internet Preservation Consortiumhoz. Szintén jelentős segítséget kaptunk az Internet Archive-tól (magyar címlista, teljes szövegű keresés a .hu országdomén alól lementett fájlokban) és legutóbb a Common Crawl projekttől (magyar nyelvet tartalmazó külföldi webhelyek adatai). Az elmúlt tíz év alatt természetesen több változás történt a magyar webarchívum informatikai, szervezeti és személyi hátterében, és persze sok minden változott és változik jelenleg is a digitálisan születő tartalmak előállításában és online terjesztésében, melyek folyamatos kihívást jelentenek számunkra. Az előadás az elmúlt tíz év eredményeit foglalja össze, de kitér a fenntarthatósági problémákra is.

Kalcsó Gyula (OSZK): *A nemzeti könyvtár webarchiválási tevékenységének aktuális problémái*

Az Országos Széchényi Könyvtár webarchívuma lassan tízéves tevékenysége során nemzetközi szinten is kiemelkedő eredményeket ért el. A fenntarthatósági problémákról az előző előadásban hallhattunk, ebben az előadó csak azokra az aktuális nehézségekre tér ki, amelyek a webarchiválásban a globális webes-informatikai világ szédületes tempójú változásai miatt adódtak, és komoly kihívást jelentenek. Elsőként azzal a kérdéssel foglalkozik, hogy mi az oka annak, hogy napjainkra két fő irányt vett a webarchiválás: a korábbi, HTTP-kéréseken alapuló crawlerek mellett megjelentek a böngészőalapúak. Bemutatja a két módszer közti különbséget, valamint azt is, hogy ezek milyen nehézségeket jelentenek a tömeges webarchiválásban.
Az előadás foglalkozik továbbá a mesterséges intelligencia rohamos elterjedésének a következményeivel a webarchiválásra nézve: hogyan detektálhatóak és kezelhetőek a robotkitiltások, milyen feladatokat támaszt mindez a webarchívumokkal szemben? Az előadás mindezeket a problémákat a magyar webarchívum működésének konkrét példáin keresztül mutatja be: milyen webhelyeket nem tudunk már jó minőségben archiválni HTTP-kérésen alapuló crawlerrel, ezekre hogyan használunk böngészőalapú crawlert, milyen nehézségek merülnek fel a webhelyekkel kapcsolatos metaadatok előállításában, hogyan kellett a robotkitiltások miatt átalakítani a képernyőkép-készítő scriptünket?

Bánki Zsolt (MNL): *Az eltűnő emlékezet – évtizedes veszteségek a kulturális örökségben*

A kulturális örökség részét képező és ezáltal a memóriaintézmények gyűjtőkörébe tartozó digitálisan született tárgyak, műalkotások, dokumentumok, kéziratok és egyéb objektumok megjelenése – ha némi önkényességgel az elektronikus levelezés elterjedésétől, vagyis az 1990-es évektől számítjuk – lassan negyven évvel ezelőttre tehető. A memóriaintézmények gyakorlatába, rutinjába azonban négy évtized alatt sem épült be evidensen ezen gyűjteményi tartalmak kezelésének módszertana, pótolhatatlan, egyre növekvő űrt teremtve kulturális örökségünkben. Az előadás fel kívánja hívni a figyelmet a gyűjtemények mulasztására, nagy vonalakban áttekintve az objektumok sokféleségét és az intézményi területek relevanciáját.

Keller-Gera Adrienn (BTM): *A born digital tartalmak kezelésének jogi kihívásai*

Előadásomat egy rövid jogi alapvetéssel kezdem, amelynek keretében áttekintem a born digital tartalmak megőrzése, archiválása és későbbi hozzáférhetővé tétele szempontjából releváns jogterületeket. Bemutatom a szerzői jog, a személyes adatok védelme, valamint kulturális örökség megőrzésének kapcsolódási pontjait. A személyes adatok védelme körében arra keresem a választ, hogy milyen feltételekkel archiválhatók olyan webes tartalmak, amelyek természetes személyekre vonatkozó adatokat tartalmaznak. Az előadás következő részében a mesterséges intelligencia térnyerésével összefüggő szerzői jogi kérdéseket tekintem át. Középpontba kerül a szerzőség és az eredetiség fogalmának értelmezése, az emberi alkotói hozzájárulás szerepe. Végezetül azt vizsgálom, hogy a technológiai fejlődés által felvetett új kérdésekre milyen válaszokat adhat a jogalkotás, a jogértelmezés és a közgyűjteményi gyakorlat.

Németh Márton (DE): *Webarchiválás a könyvtár- és információtudományi felsőoktatásban*

Előadásomban a webarchiválás beágyazásáról lesz szó a könyvtár- és információtudományi felsőoktatási képzési kínálatba. Bevezetésképpen kitérek előzményként az OSZK-ban megvalósuló akkreditált tanfolyamra, hogy ennek az oktatási anyagait miképpen lehet integrálni felsőoktatási keretek közé. Megemlítem még az OSZK munkatársaként tartott vendégelőadásaimra is az ELTE BTK Könyvtár- és Információtudományi Intézetében. Előadásom fő vázát a Debreceni Egyetem Bölcsészettudományi Karának könyvtár- és információtudományi mesterszakos képzésében 2025-ben elindult, webarchiválással foglalkozó kurzus felépítésének, témaköreinek ismertetése és az oktatói tapasztalatok megosztása adja. Kitérek az Informatikai Kar doktori képzésében futó, webarchiválással foglalkozó választható tárgy kereteinek ismertetésére is.

Sárközi-Lindner Zsófia (ELTE BTK): *Régi-új kérdések: webes források és archívumelmélet egymás tükrében*

A webarchívumok történeti forrásként való értelmezése több ponton is próbára teszi a hagyományos levéltári elveket. A webes források sajátosságai új helyzetet teremtenek, miközben a szelekció, a kontextus, a proveniencia, a reprezentáció és a közvetítő szerepek kérdései korántsem a webarchiválással jelennek meg először az archívumelmélet nemzetközi diskurzusában. A webes környezet ugyanakkor e korábbi problémákat új összefüggésbe helyezi és kiélezi, egyes pontokon pedig újragondolásukra késztet. Az előadás e folytonosságok és fordulatok mentén tekint a webarchívumokra mint a levéltári elvek egyfajta stressztesztjére, kitérve arra is, milyen szerepe lehet a történészi szempontnak a jövő történeti forrásbázisának formálódásában.

Ferenci Tamás: *Lelkes amatőrként a digitális archiválás világában*

Különleges helyzetben vagyok, hiszen teljesen laikusként kerültem bele a digitális archiválás világába: a 2026-os országgyűlési választások előtt pár nappal egy megérzéstől vezérelve elkezdtem politikusok Facebook-oldalait lementeni azzal a céllal, hogy egy nyilvános archívumot hozzak belőlük létre – hátha a választás után lesz, aki törli az oldalát. Ez a megérzés hamar beigazolódott, így a mentések is felértékelődtek, ezért 2026 nyarán nekiláttam a megfelelő weboldal létrehozásának, ami nem volt nyilvánvaló feladat (több mint 100 politikus 720 ezer posztját mentettem le, bennünk 650 ezer fényképpel és 1,5 TB videóval, melyeket szintén mentettem).
Előadásomban bemutatom az így létrejött és ma már a https://ferenci-tamas.github.io/fb-politika-archivum/ [6] címen bárki számára elérhető weboldalt, beszélek a hátteréről, a motivációmról, a munkamenetemről, a nehézségekről, a technikai megoldásokról, és külön kitérek arra, hogy milyen elemzési lehetőségeket nyit ki egy ilyen adatbázis az archiváláson túl. Remélem, hogy a profi megoldások mellett érdekes lehet egy ilyen hobbiprojekt (béka)perspektívája is, mint példa arra, hogy hogyan kapcsolódhatnak be akár magánszemélyek is az archiválási munkába.

2026/11/18 - 12:17

Forrás webcím: https://oszk.hu/rendezvenyek/404-not-found-ki-orzi-meg-az-internetet_261118

Hivatkozások:
[1] https://oszk.hu/sites/default/files/404-konferencia_261118_Fb-event.jpg
[2] https://www.google.hu/maps/place/Orsz%C3%A1gos+Sz%C3%A9ch%C3%A9nyi+K%C3%B6nyvt%C3%A1r/@47.4954231,19.0370263,17z/data=!3m1!4b1!4m6!3m5!1s0x4741dc3d2cbe10bd:0x95a7b3e3f4b9f737!8m2!3d47.4954231!4d19.0396012!16zL20vMDZycnls?hl=hu&entry=ttu&g_ep=EgoyMDI0MTEyNC4xIKXMDSoASAFQAw%3D%3D
[3] https://fb.me/e/ftSXGctpY
[4] https://oszk.hu/404-not-found-2026-regisztracio
[5] https://www.youtube.com/@nemzetikonyvtar/streams
[6] https://ferenci-tamas.github.io/fb-politika-archivum/
[7] https://oszk.hu/category/foszotar-es-pozicionalo/hirek
[8] https://oszk.hu/category/foszotar-es-pozicionalo/hirek-cimlapon
[9] https://oszk.hu/category/foszotar-es-pozicionalo/rendezvenyek