A digitális örökség szolgálatában – jubileumi konferenciát rendezett az OSZK a webarchiválásról

2026/11/23



Tizedik alkalommal rendezte meg 2026. november 18-án könyvtárunk a *404 Not Found – Ki őrzi meg az internetet? A webarchiválás és a born digital tartalmak megőrzésének aktuális problémái* című konferenciát és kerekasztal-beszélgetést.



[1]

Tizedik alkalommal rendezte meg 2026. november 18-án könyvtárunk a *404 Not Found – Ki őrzi meg az internetet? A webarchiválás és a born digital tartalmak megőrzésének aktuális problémái* című konferenciát és kerekasztal-beszélgetést. A jubileumi rendezvény a magyar webarchiválás elmúlt évtizedének eredményeit és tanulságait, valamint a digitálisan születő tartalmak megőrzésének technológiai, jogi és szakmai kihívásait állította középpontba.

A konferencia végén Gerencsér Judit, az OSZK megbízott főigazgatója a nemzeti könyvtár nevében köszönetet mondott a nyugdíjba vonuló Drótos Lászlónak, a hazai digitális kultúra és webarchiválás meghatározó alakjának.

A konferencia megnyitóján **Gerencsér Judit** megbízott főigazgató köszöntötte a résztvevőket. Beszédében hangsúlyozta, hogy a *404 Not Found – Ki őrzi meg az internetet?* rendezvénysorozat mára önálló szakmai márkává vált, amely nívót, minőséget és elismertséget képvisel. Mint fogalmazott, a nemzeti könyvtár küldetése a magyar kulturális örökség összegyűjtése, megőrzése és hozzáférhetővé tétele, ez a feladat azonban napjainkra jóval túlmutat a hagyományos dokumentumokon. Kultúránk, tudásunk és közös emlékezetünk egyre nagyobb része eleve digitális formában születik, és sok esetben kizárólag az interneten vagy a közösségi médiában érhető el. A weboldalak, digitális dokumentumok, elektronikus levelezések, online művészeti alkotások és más digitálisan létrejött tartalmak megőrzése ezért a közgyűjtemények egyik meghatározó feladatává vált.



[2]



[3]

A megbízott főigazgató felidézte, hogy az OSZK Webarchívum létrehozásának előkészítő munkálatai 2017-ben kezdődtek meg. Az azóta eltelt időszakban az OSZK munkatársai jelentős szakmai munkával, nemzetközi együttműködésekre és tapasztalatokra támaszkodva alakítottak ki új szolgáltatásokat és valósítottak meg innovációkat. Kiemelte a nemzetközi pályázatok révén létrejött szakmai kapcsolatokat, valamint a Luxemburgi Nemzeti Könyvtár és a Skót Nemzeti Könyvtár képviselőinek korábbi látogatásait. Hangsúlyozta, hogy a webarchiválás ma már nem pusztán informatikai és technológiai kérdés: összefügg a nyelvi és kulturális sokszínűség megőrzésével, a mesterséges intelligencia térnyerésével, a szerzői joggal, az adatvédelemmel és a közgyűjtemények hosszú távú fenntarthatóságával is.

A konferencia előadói ezeknek az összetett kérdéseknek
a különböző szempontjait járták körül.


Common Crawl Use Cases
- Natural Language Processing
- Language modeling
- Information Retrieval (search indexes)
- Internet security research
- Web Science
- Semantic Web and Linked Data
- Sociopolitical research
- Economic research
- A lot more... [3]
- ... and many "derivative" datasets!
számos adatkészlet is született a Common Crawl adataiból,
ORSZÁGOS SZÉCHÉNYI KÖNYVTÁR
1802
BIBLIOTHECA NATIONALIS

[4]


Papers Citing Common Crawl
Egyébként ki használja ezeket az adatokat?
ORSZÁGOS SZÉCHÉNYI KÖNYVTÁR
1802
BIBLIOTHECA NATIONALIS HUNGARIAE

[5]

A szakmai programot a **Common Crawl előadóinak** angol nyelvű, online videóüzenete nyitotta meg *Who Gets Crawled? Coverage and Policies at Common Crawl (Mit mentünk? Gyűjtőkör és alapelvek a Common Crawlnál)* címmel. A nonprofit szervezet rendszeresen feltérképezi az internetet, és az összegyűjtött webes archívumokat ingyenesen hozzáférhetővé teszi. A 2008 óta több mint 300 milliárd weboldalt összegyűjtő projekt havonta hozzávetőleg 2,5 milliárd új oldallal bővíti archívumát, amelyet világszerte kutatók, archiválók és fejlesztők használnak. Az előadás a gyűjtési szempontok kialakításának nehézségeit mutatta be, különös tekintettel arra, hogy a keresőrobotok jóval több URL-t találnak, mint amennyit a rendelkezésre álló erőforrásokkal le lehet tölteni.

A linkalapú rangsorolás következtében az angol és néhány nagyobb nyelvű tartalom aránytalanul nagy részt kap az archiválási keretből, miközben a kisebb nyelvi közösségek háttérbe szorulnak. A Common Crawl új törekvései között ezért kiemelt szerepet kapott az alulreprezentált nyelvek felkutatása, a webes szövegek pontosabb nyelvi azonosítása és a nem angol nyelvű tartalmak gyűjtésének bővítése. Az előadás kitért arra is, hogy a generatív mesterséges intelligencia terjedésével egyre több webhely korlátozza a keresőrobotok hozzáférését, gyakran anélkül, hogy különbséget tenne az archívumok, a keresőmotorok és a mesterségesintelligencia-fejlesztő cégek között. Ez új kérdéseket vetett fel az archiválás feltételeiről és a megőrzendő tartalmak kiválasztásának felelősségéről.



[6]



[7]

Drótos László, az OSZK munkatársa *Az eltűnő web nyomában – Tíz év sikerei, buktatói és tanulságai a nemzeti könyvtár webarchívumában* című előadásában összegezte a magyar nemzeti webarchívum létrehozásának és működésének első évtizedét. Felidézte, hogy az előkészítő munka 2017 elején indult el, és bár Magyarország a környező országok többségéhez képest később kezdte meg a webarchiválást, a digitálisan születő tartalmak megőrzésében már több évtizedes tapasztalattal rendelkező szakemberekre építhetett. A meglévő tudás, a gyakorlati megvalósítást rövid időn belül követő jogi szabályozás, valamint a nemzetközi szakmai együttműködések révén néhány év alatt sikerült ledolgozni a hátrányt, és nemzetközileg is elismert szolgáltatást kialakítani.

Az eredményekhez jelentősen hozzájárult a külföldi szakemberek által korábban kifejlesztett szoftverek alkalmazása és az általuk megosztott szakmai tudás. Az OSZK 2018-ban csatlakozott az International Internet Preservation Consortiumhoz, emellett az Internet Archive-tól magyar címlistát és a .hu országdomén alól lementett fájlokban teljes szövegű keresést biztosító támogatást kapott. A Common Crawl projekt legutóbb a magyar nyelvű tartalmakat is tartalmazó külföldi webhelyek adataival segítette a munkát. Drótos László ismertette az idei évben elért eredményeket, majd kitért az elmúlt évtized informatikai, szervezeti és személyi változásaira, valamint a digitális tartalmak előállításában és online terjesztésében bekövetkezett átalakulásokra. Előadása a sikerek mellett a fenntarthatóság kérdéseire és a folyamatosan változó digitális környezetből adódó kihívásokra is felhívta a figyelmet.



[8]



[9]

A nemzeti könyvtár webarchiválási tevékenységének aktuális technológiai nehézségeit **Kalcsó Gyula,** az OSZK Digitális Megőrzési és Webarchiválási Osztályának vezetője mutatta be. *A nemzeti könyvtár webarchiválási tevékenységének aktuális problémái* című előadásában arra kereste a választ, miként alakította át a webes és informatikai környezet gyors fejlődése az archiválás módszereit. Részletesen ismertette a HTTP-kéréseken alapuló, hagyományos keresőrobotok és az újabb, böngészőalapú crawlerek közötti különbségeket, valamint azt, hogy az eltérő technológiák milyen nehézségeket okoznak a nagy tömegű webes tartalom mentésekor. Bemutatta, hogy bizonyos webhelyek a hagyományos módszerekkel már nem archiválhatók megfelelő minőségben, ezért esetükben böngészőalapú megoldások alkalmazására van szükség.

Az előadás a mesterséges intelligencia elterjedésének következményeivel és az egyre gyakoribb robotkitiltások kezelésével is foglalkozott. Kalcsó Gyula a magyar webarchívum konkrét példáin keresztül szemléltette, hogyan változott a webhelyek archiválásának gyakorlata, milyen nehézségek merültek fel a kapcsolódó metaadatok előállításában, és miként kellett a robotkitiltások miatt átalakítani az oldalképek készítésére szolgáló programot.



[10]



[11]

Bánki Zsolt, a Magyar Nemzeti Levéltár Digitális Szolgáltatásfejlesztési Osztály vezetője *Az eltűnő emlékezet – évtizedes veszteségek a kulturális örökségben* című előadásában a digitálisan született kulturális javak megőrzésének elmaradásából fakadó veszteségekre irányította a figyelmet. Rámutatott, hogy az elektronikus levelezés 1990-es évekbeli elterjedésétől számítva már közel négy évtizede jelennek meg a kulturális örökség részét képező digitális tárgyak, műalkotások, dokumentumok, kéziratok és egyéb objektumok. Ezek kezelésének módszertana azonban mindmáig nem épült be magától értetődő módon a memóriaintézmények gyakorlatába. A hiányosságok egyre nagyobb és pótolhatatlan űrt hagynak a kulturális örökségben. Az előadás a digitálisan létrejött objektumok sokféleségének áttekintésével és az érintett intézményi területek bemutatásával érzékeltette a probléma súlyát.



[12]



[13]

A digitális tartalmak megőrzésének jogi vonatkozásait **Keller-Gera Adrienn** jogász ismertette. *A born digital tartalmak kezelésének jogi kihívásai* című előadásában áttekintette az archiválás és a későbbi hozzáférhetővé tétel szempontjából meghatározó jogterületeket, különösen a szerzői jog, a személyes adatok védelme és a kulturális örökség megőrzésének kapcsolódási pontjait. Kitért arra, milyen feltételekkel archiválhatók a természetes személyekre vonatkozó adatokat tartalmazó webes tartalmak, valamint a mesterséges intelligencia térnyerésével összefüggésben felmerülő szerzői jogi kérdésekre is. Kiemelte a szerzőség és az eredetiség fogalmának értelmezését, valamint az emberi alkotói hozzájárulás jelentőségét. Az előadás végül azt vizsgálta, hogy a technológiai fejlődés új kihívásaira milyen válaszokat adhat a jogalkotás, a jogértelmezés és a közgyűjteményi gyakorlat.



[14]



[15]

Németh Márton a Debreceni Egyetem képviseletében a terület felsőoktatási képzésekbe történő beépítésének lehetőségeit ismertette *Webarchiválás a könyvtár- és információtudományi felsőoktatásban* című előadásában. Bemutatta, miként hasznosíthatók a webarchiválásról szóló, az OSZK-ban megvalósított akkreditált tanfolyam oktatási anyagai az egyetemi képzésben, és felidézte az ELTE Bölcsészettudományi Karának Könyvtár- és Információtudományi Intézetében tartott vendégelőadásait is. Előadásának középpontjában a Debreceni Egyetem Bölcsészettudományi Karán 2025-ben elindított, a könyvtár- és információtudományi mesterszakos hallgatóknak szóló webarchiválási kurzus felépítése, témakörei és oktatási tapasztalatai álltak. Emellett ismertette az egyetem Informatikai Karának doktori képzésében meghirdetett, webarchiválással foglalkozó választható tantárgy kereteit.


RÉGI-ÚJ KÉRDÉSEK:
WEBES FORRÁSOK ÉS
ARCHÍVUMELMÉLET EGYMÁS TÜKRÉBEN
Sárközi-Lindner Zsófia
ELTE BTK
ELTE HTK Irodalomtudományi
Kutatóközpont DigiPhil

[16]


értékelnie és szelektálnia
(Schellenberg 1956)

[17]

A webarchívumok történeti forrásként való értelmezését **Sárközi-Lindner Zsófia**, az ELTE Bölcsészettudományi Karának munkatársa járta körül *Régi-új kérdések: webes források és archívumelmélet egymás tükrében* című előadásában. Rámutatott, hogy a webes források sajátosságai több ponton is próbára teszik a hagyományos levéltári elveket. A szelekció, a kontextus, a proveniencia, a reprezentáció és a közvetítői szerepek kérdései ugyanakkor nem kizárólag a webarchiválással jelentek meg az archívumelméletben. A digitális környezet ezeket a korábban is ismert problémákat új összefüggésekbe helyezte, és egyes esetekben újragondolásukra késztetett. Az előadás a webarchívumokat a levéltári elvek egyfajta stressztesztjeként értelmezte, kitérve arra is, milyen szerepet játszhatnak a történészi szempontok a jövő történeti forrásbázisának kialakításában.



[18]



[19]

A konferencia előadói között **Ferenci Tamás** *Lelkes amatőrként a digitális archiválás világában* címmel egy magánkezdeményezésből létrejött archívum történetét mutatta be. Elmondta, hogy a 2026-os országgyűlési választások előtt néhány nappal, saját megérzésére alapozva kezdte el politikusok Facebook-oldalainak mentését, hogy az így összegyűjtött tartalmakból nyilvános archívumot hozzon létre arra az esetre, ha a választás után egyes oldalak eltűnnének. A feltételezés hamar beigazolódott, így a mentések jelentősége is felértékelődött. A projekt folytatásaként 2026 nyarán megkezdte egy önálló weboldal kialakítását. A munka során több mint 100 politikus 720 ezer bejegyzését mentette le, amelyek 650 ezer fényképet és 1,5 TB-nyi videót tartalmaztak. Az így létrejött archívum bárki számára elérhető. Az előadó ismertette a projekt hátterét, motivációit, munkafolyamatát, technikai megoldásait és nehézségeit, valamint az adatbázis archiváláson túlmutató elemzési lehetőségeit. Példája azt is szemléltette, hogy a digitális örökség megőrzéséhez magánszemélyek is érdemben hozzájárulhatnak.


ORSZÁGOS
SZÉCHÉNYI
KÖNYVTÁR
1802
BIBLIOTHECA
NATIONALIS
HUNGARIA

[20]


ORSZÁGOS
SZÉCHÉNYI

[21]

A konferencia előadásait *A webarchiválás és a born digital tartalmak megőrzésének aktuális problémái – Múlt, jelen és jövő* című kerekasztal-beszélgetés követte.

A résztvevők **Bánki Zsolt** (MNL), **Drótos László** (OSZK), **Horváth Adrienn** (Vörösmarty Mihály Könyvtár, igazgató), **Kalcsó Gyula** (OSZK), **Moldován István (**Pázmány Péter Katolikus Egyetem Bölcsészet- és Társadalomtudományi Kar, könyvtárvezető) és **Petrik Máté** (Társadalmi Kapcsolatokért és Kultúráért Felelős Minisztérium, ágazati digitalizációért és digitális szolgáltatásokért felelős helyettes államtitkár) voltak.

A beszélgetés az előadásokban felvetett kérdések közös megvitatására, valamint a webarchiválás és a digitálisan született tartalmak megőrzésével kapcsolatos múltbeli tapasztalatok, jelenlegi nehézségek és jövőbeli lehetőségek áttekintésére adott alkalmat. A rendezvény kérdésekkel, hozzászólásokkal és szakmai vitával zárult.



[22]



[23]

A jubileumi konferencia egyben egy meghatározó szakmai életút előtti tisztelgés alkalma is volt.

Gerencsér Judit megbízott főigazgató a rendezvény végén a nemzeti könyvtár nevében búcsúzott Drótos Lászlótól, aki 2026 végén fejezi be aktív munkáját az OSZK-ban. Beszédében a magyar internet egyik úttörőjeként és a hazai digitális kultúra meghatározó alakjaként méltatta a szakembert. Felidézte, hogy bár eredeti végzettsége szerint geológus mérnök, már a Miskolci Egyetemi Könyvtárban végzett munkája során a számítógépek, a hálózatok és az információhoz való új hozzáférési lehetőségek világa vált számára meghatározó szakmai területté. Saját megfogalmazását idézve: „nem a könyvek, hanem a számítógépek csábították a könyvtárba”.

Drótos László az internet magyarországi megjelenésének kezdetétől aktívan részt vett a hazai digitális világ alakításában. Megalapította a KATALIST szakmai levelezőlistát, 1994-ben társaival elindította a Magyar Elektronikus Könyvtárat, később pedig a Digitális Képarchívum és a webarchiválás létrejöttében is meghatározó szerepet vállalt. Ezek a kezdeményezések mára a nemzeti könyvtár digitális működésének alapvető pilléreivé váltak.



[24]



[25]

Gerencsér Judit külön kiemelte Drótos László korai felismerését, miszerint a digitális örökség megőrzése nem korlátozódhat a hagyományos dokumentumtípusokra: az interneten létrejövő tartalmakat és a folyamatosan változó digitális kultúra lenyomatait is meg kell őrizni. A szakember már 2006-ban megfogalmazott elképzelése az OSZK fejlesztési programjának köszönhetően később megvalósítható projektté, majd önálló szakmai területté fejlődött. A magyar webarchívum szakmai koncepciója, informatikai háttere és működésének megalapozása elválaszthatatlan Drótos László munkásságától.

A főigazgató méltatásában az alkotó kíváncsiságot, a szakmai igényességet és a közösség iránti felelősséget emelte ki. Hangsúlyozta, hogy Drótos László nemcsak létrehozott és fejlesztett, hanem tudását is széles körben megosztotta: publikált, előadásokat tartott, tananyagokat készített, és kollégák generációit segítette a digitális világban való eligazodásban. Munkáját türelem és természetes segítőkészség jellemezte, a hozzá fordulók mindig figyelmes, alapos és megértő válaszra számíthattak.

Gerencsér Judit a nemzeti könyvtár nevében megköszönte mindazt, amit Drótos László az elmúlt évtizedekben létrehozott, elindított és továbbadott, köszönetet mondva a szakmai kíváncsiságáért, tudásáért, türelméért, segítőkészségéért és a közös ügy iránti következetes elkötelezettségéért.

Drótos László szakmai teljesítményét számos elismeréssel jutalmazták. A Kalmár László-díj, a Magyar Elektronikus Könyvtár létrehozásáért kapott életműdíj és a Hungarnet-díj mellett 2024-ben az Országos Széchényi Könyvtár Széchényi Ferenc-emlékéremmel ismerte el munkásságát, ugyanabban az évben pedig a Szinnyei József-díjat is átvehette.



[26]



[27]

Kapcsolódó tartalom:
A konferencia programja [28]
2026/11/22 - 11:38

Forrás webcím: https://oszk.hu/hirek/digitalis-orokseg-szolgalataban-jubileumi-konferenciat-rendezett-az-oszk-webarchivalasrol_261118

Hivatkozások:
[1] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_04.jpg
[2] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_07.jpg
[3] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_02.jpg
[4] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_08.jpg
[5] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_06.jpg
[6] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_10.jpg
[7] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_13.jpg
[8] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_52.jpg
[9] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_51.jpg
[10] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_19.jpg
[11] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_18.jpg
[12] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_23.jpg
[13] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_21.jpg
[14] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_25.jpg
[15] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_26.jpg
[16] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_27.jpg
[17] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_28.jpg
[18] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_53.jpg
[19] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_31.jpg

[20] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_34.jpg
[21] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_40.jpg
[22] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_38.jpg
[23] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_55.jpg
[24] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_44.jpg
[25] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_49.jpg
[26] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_50.jpg
[27] https://oszk.hu/sites/default/files/404-Not-Found-konferencia_261118_foto-VAL_46.jpg
[28] https://oszk.hu/rendezvenyek/404-not-found-ki-orzi-meg-az-internetet_261118
[29] https://oszk.hu/category/foszotar-es-pozicionalo/hirek
[30] https://oszk.hu/category/foszotar-es-pozicionalo/hirek-cimlapon