

404 Not Found – Ki őrzi meg az internetet? – A digitális megőrzés új irányai

2025/11/28



Az OSZK Digitális Bölcsészeti Központja 2025. november 12-én tartotta meg a webes tartalmak archiválásával és a digitális megőrzés aktuális kérdéseivel foglalkozó éves szakmai rendezvényét.



[1]

Az Országos Széchényi Könyvtár Digitális Bölcsészeti Központja 2025. november 12-én tartotta meg a webes tartalmak archiválásával és a digitális megőrzés aktuális kérdéseivel foglalkozó éves szakmai rendezvényét, a kilencedik 404 Not Found – Ki őrzi meg az internetet? című konferenciát és workshopot.

Az Országos Széchényi Könyvtárban megrendezett esemény központi témája idén a **webarchívum mint adat** volt, ez állt az egész napos program előadásainak és műhelyfoglalkozásának fókuszában. Már a megnyitók felhívták a figyelmet arra, hogy a webarchívumokban felhalmozott hatalmas

adattömeg kutatási hasznosítása stratégiai fontosságú mind a hazai, mind a nemzetközi tudományos életben.

A konferenciát a házigazda intézmény részéről **Gerencsér Judit**, az OSZK általános főigazgató-helyettese nyitotta meg, majd **Latorcai Csaba**, a Közigazgatási és Területfejlesztési Minisztérium közigazgatási és területfejlesztési miniszterhelyettese, parlamenti államtitkár köszöntötte a résztvevőket.



[2]



[4]



[5]

A szakmai program első részében a Skót Nemzeti Könyvtár két munkatársa, **Eilidh MacGlone és Joseph Marshall** mutatta be a *National Library of Scotland: Our web archive (1999-)* című előadásában az Edinburghban és Glasgow-ban működő kötelezpéldány-könyvtár két és fél évtizedes webarchiválási gyakorlatát. Részletesen ismertették, hogyan tette lehetővé számukra a 2013-ban megjelent Kötelezpéldány-könyvtárak (nem nyomtatott művek) szabályzata, hogy az engedélyezett gyűjtés mellett köteles példányként is archiváljanak weboldalakat, és miként alakították ki az azóta fejlődő, kutatói együttműködésekre építő munkafolyamataikat.

A program online bekapcsolódó előadója, **Kocsis Andrea** (University of Edinburgh) a *Hozzáérés engedélyezve. Kreatív utakon a webarchívumokhoz* című előadásában mutatta be az általános olvasók, az adathasználók és a „digitálisan kíváncsi” felhasználók bevonásának lehetőségeit. A University of Edinburgh, a Cambridge University Libraries és a National Library of Scotland közös projektjeire támaszkodva ismertette a gamifikált felületek, kiállítások, gondosan kurált metaadatkészletek és a Datasheets for Data dokumentáció szerepét, valamint két esettanulmányt – az *Archive of Tomorrow* és a *Digital Ghost* projekteket.

A Nemzeti Audiovizuális Archívum (NAVA) részéről **Bajnok László** a *NAVA archiválási tevékenysége – Audiovizuális dokumentumok katalogizálása a NAVA-ban* című előadásában az audiovizuális tartalmak gyűjtésének, digitalizálásának és katalogizálásának folyamatát ismertette. Részletesen bemutatta, hogy a NAVA a 2006. január 1-je óta gyűjtött rádiós és televíziós műsorszámokat miként teszi elérhetővé oktatási és kutatási hálózatokon.



[7]



[9]

A digitális információk megőrzésének alternatív eszközeiről szólt **Holl András** (MTA KIK) *Repozitóriumok – a digitális információk archiválásának eszközei* című előadása, amely a komplex webes tartalmak repozitóriumi elhelyezésének lehetőségeit tárgyalta, és egyúttal megemlékezett Dömölki Bálintról, a témához kapcsolódó NWS2025-ös előadás egyik szerzőjéről.

A nyelvtechnológiai kutatások szempontjait mutatta be **Sárossy Bence** (ELTE Nyelvtudományi Kutatóközpont) *Milyen magyar nyelvet tanul meg a mesterséges intelligencia? – Az internet tartalma mint tanítóanyag* című előadásával. Részletesen ismertette a PULI nyelvmodellcsalád 70 milliárd szavas tanítókorpuszának összeállítási folyamatát, az internetes adatgyűjtés kihívásait, valamint az adattisztítási és előfeldolgozási lépéseket, amelyek lehetővé tették a ChatPULI fejlesztését.

A nemzeti könyvtár webarchiválási tapasztalatait **Drótos László** (OSZK) *Adatok az adatokról – Az MNMKK OSZK webarchívuma 2025-ben* című előadása foglalta össze. Beszámolt a 2025-ben több terabájtal bővülő webtartalomról, az új és jelentősen fejlesztett részgyűjteményekről, az idei első webtér szintű aratás előkészítéséről és eredményéről, valamint a Heritrix robot és a Browsertrix crawler kombinált használatáról. Külön kitért arra is, hogy a nyilvános gyűjtemény több száz részletes metaadat-leírással gazdagodott, és elkészült annak az adatbázisnak a terve, amely az archivált tartalmak nyilvántartását és az automatizált munkafolyamatok támogatását egyaránt szolgálja.



[10]



[11]



[12]



[13]

A délelőtti programot **Kalcsó Gyula** (OSZK) *Hírportálok napi mentése Browsertrix crawlerrel* című előadása zárta, amely a válogatott magyar hírportálok napi archiválásának megújított eljárását mutatta be. Részletesen összehasonlította a Heritrix és a Browsertrix eszközöket, ismertette a dinamikus tartalmak archiválásának kihívásait, valamint azokat a szkripteket, amelyekkel a görgetést és lapozást igénylő hírfolyamok is sikeresen menthetők.

A délután 14 órakor kezdődő workshopon, amelyet szintén **Kalcsó Gyula** vezetett *Az archiválás új szintje – dinamikus generált, interaktív webhelyek mentése böngészőalapú crawlerrel (Browsertrix)* címmel, a résztvevők gyakorlati tapasztalatot szerezhettek a Browsertrix telepítésében, konfigurálásában és konkrét webhelyek próbamentéseiben. A nagy érdeklődéssel kísért műhelyfoglalkozás lehetőséget adott arra, hogy a szakemberek önállóan is képesek legyenek modern, interaktív webhelyek magas minőségű archiválására.

A konferencia kérdésekkel, hozzászólásokkal és szakmai vitával zárult. A nap során több új hazai és nemzetközi együttműködés körvonalazódott, többek között technológiai tudásmegosztás, új archiválási módszerek tesztelése, a magyar nagy nyelvi modell építésében való közreműködés, valamint az audiovizuális és webes gyűjtemények összekapcsolásának lehetőségei. Az esemény megerősítette, hogy a digitális kulturális örökség megőrzése olyan közös szakmai feladat, amelynek sikere az intézmények közti együttműködésen és a webarchívumok sokoldalú kutatási hasznosításán múlik.



[14]



[15]



[16]

Kapcsolódó tartalmak:

[A konferencia programja](#) [17]

[Az előadások listája, kivonatai és prezentációi](#) [18]

2025/11/25 - 14:13

Forrás webcím: http://193.6.201.226/hirek/404-not-found-ki-orzi-meg-az-internetet-digitalis-megorzes-uj-iranyai_251112

Hivatkozások:

- [1] http://193.6.201.226/sites/default/files/404_2025_FB_jav.jpg
- [2] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_03.jpg
- [3] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_01.jpg
- [4] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_07.jpg
- [5] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_02.jpg
- [6] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_16.jpg
- [7] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_12.jpg
- [8] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_25.jpg
- [9] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_29.jpg
- [10] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_24.jpg
- [11] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_32.jpg
- [12] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_38.jpg
- [13] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_42.jpg
- [14] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_76.jpg
- [15] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_54.jpg
- [16] http://193.6.201.226/sites/default/files/404-Not-Found_konferencia_251112_foto-VAL_68.jpg
- [17] <https://oszk.hu/rendezvenyek/404-not-found-ki-orzi-meg-az-internetet-konferencia-es-workshop-251112>

[18] <https://webarchivum.oszk.hu/szakembereknek/404-not-found-workshop/404-not-found-konferencia-es-workshop-2025-november-12/>

[19] <http://193.6.201.226/category/foszotar-es-pozicionalo/hirek>

[20] <http://193.6.201.226/category/foszotar-es-pozicionalo/hirek-cimlapon>