



Egy véget nem érő tanulási folyamat – Összefoglaló az Országos Széchényi Könyvtár Webarchiválási Osztályának 2021. évi tevékenységéről

DRÓTOS László

Bevezetés

Harmadik alkalommal számolunk be a *Könyvtári Figyelő*ben az Országos Széchényi Könyvtár (OSZK) háromfős webarchiváló csoportjának munkájáról és az elmúlt évben szerzett új tapasztalatokról, hiszen az internetes tartalmak lementése, hosszú távon megtekinthető állapotban való tartása és a sok terabájnyi anyag kutathatóvá tétele egy véget nem érő tanulási folyamat. Az online világ változékonysága, a divattossá váló, majd eltűnő fájlformátumok, műfajok, szolgáltatások és platformok miatt a kidolgozott és bejárattott archiválási munkafolyamatokat időről időre újra kell tervezni, a korábban jól bevált letöltő és megjelenítő szoftverek helyett újakat kell kifejleszteni, megtanulni és beépíteni az üzemszerű működésbe. 2021 – az azt megelőző év nagy részéhez hasonlóan – nem volt egyszerű munkaszervezési szempontból.

A járvány miatt a kommunikáció nagyrészt elektronikus levelek, üzenetküldések és videóhívások formájában zajlott, a konferenciákat és workshopokat felváltották a webináriumok, a tanfolyamokat pedig a távoktatás. A Webarchiválási Osztály két belső levelezőlistáján tavaly közel 2500 e-mail ment át, és legalább ugyanennyi volt a közvetlen levélváltás külső partnerekkel vagy más OSZK-s kollégákkal. Szerencsére a feladat jellege miatt a távmunka nem csökkentette a hatékonyságot, és több területen előbbre is tudtunk lépni, de például a szakmai rendezvényeken szokásos személyes beszélgetések elmaradása vagy az online oktatás miatt a gyakorlati ismeretek átadásának ellehetetlenülése a tudásmegosztást eléggé megnehezítette. Ezért is kezdtük el áprilisban a honlapunkon és a MIA-L levelezőlistán közzétett havi „hírlevél” formájában összefoglalni egyrészt az OSZK webarchívumával kapcsolatos fejleménye-

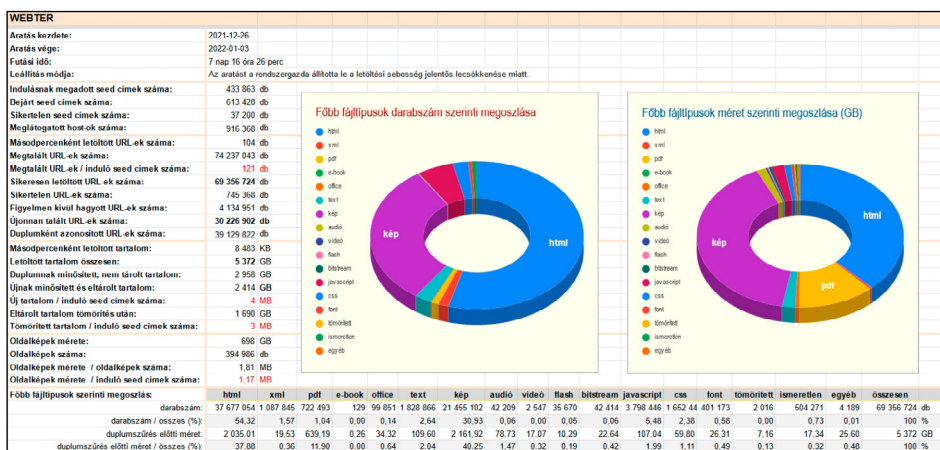
ket, másrészt azokat az új ismereteket, amelyeket a különféle archiváló technológiák tesztelése közben és a külföldi projektekről szóló beszámolók alapján szereztünk.¹ Az alábbiakban a 2021-es év eredményei és eseményei közül a legfontosabbakat emeltem ki, négy terület szerint csoportosítva.

Archívum

Az archívum publikus része² az év során több mint 80 webhellyel bővült, főként állami vagy önkormányzati támogatással működő intézmények honlapjaival, mert ezek nyilvános szolgáltatásához – a Kormány 626/2020. (XII. 22.) számú rendelete 6. § (3) pontja alapján – 2021 januárjától már nem kell egyedi szerződéseket kötnünk.³ Azoknál a webhelyeknél, amelyek nem tartoznak ebbe a kategóriába, és amelyeknél a korábban kötött szerződések 2021 végéig szóltak, elkezdtük a szerződések megújítását határozatlan időtartamra. Bár az ajánló úrlapon⁴ keresztül kevés javaslat érkezett, viszont több honlap, blog és

fórum gazdája közvetlenül keresett meg bennünket azzal, hogy archiváljuk a szolgáltatását. Öröndetes, hogy lassan kezd elterjedni, hogy a nemzeti könyvtár foglalkozik ezzel, és kezd tudatosulni az emberekben, hogy a sok munkával, évek alatt felépített webhelyeket érdemes a jövő nemzedékei számára is megőrizni, és felvenni velünk a kapcsolatot. Ha a szolgáltatás gazdájával együtt tudunk működni, akkor általában sokkal jobb minőségű mentés készíthető, szükség esetén bevonva mindkét oldalon a helyi informatikusokat is.

Az OSZK saját online tartalomszolgáltatásai közül 46 újabbat mentettünk le,⁵ és 85 tétellel gyarapodott a szintén nyilvános *Rákóczi archívum*⁶ is – ennek fele webes tartalom, a másik része pedig digitális könyv. A gyűjtemény nem publikus részét tartalmazó szerveren a 2020. évi 50-hez képest tavaly 63 nagy aratást futtattunk, és az előző évi 236,6 millió helyett 2021-ben 359,9 millió URL-címről töltöttünk le fájlokat, amiből 171,1 millió volt a korábban még nem mentett vagy időközben megváltozott cím (1. ábra).



1. ábra

A tematikus részgyűjtemények és az e-periodikák 2021-es aratásainak összesített adatai

Három új tematikus gyűjteményt alakítottunk ki (idegenforgalom és turizmus, történelem és helytörténet, természet- és műszaki tudományok), így összességében több mint 44 ezer, témák szerint válogatott web-címünk van (ez a szám 2020 végén 35,2 ezer volt). Az elektronikus periodikák weboldalait nyilvántartó címlistánk már több mint 6400 tételtől áll, ebből kb. 1200 db a tavalyi felvételű webcím. Az új kiadványokról az OSZK-ban működő Magyar ISSN Nemzeti Központ, valamint az Elektronikus Periodika Archívum és Adatbázis (EPA) rendszereiből érkező automatikus levelekből értesülünk. Az EPA

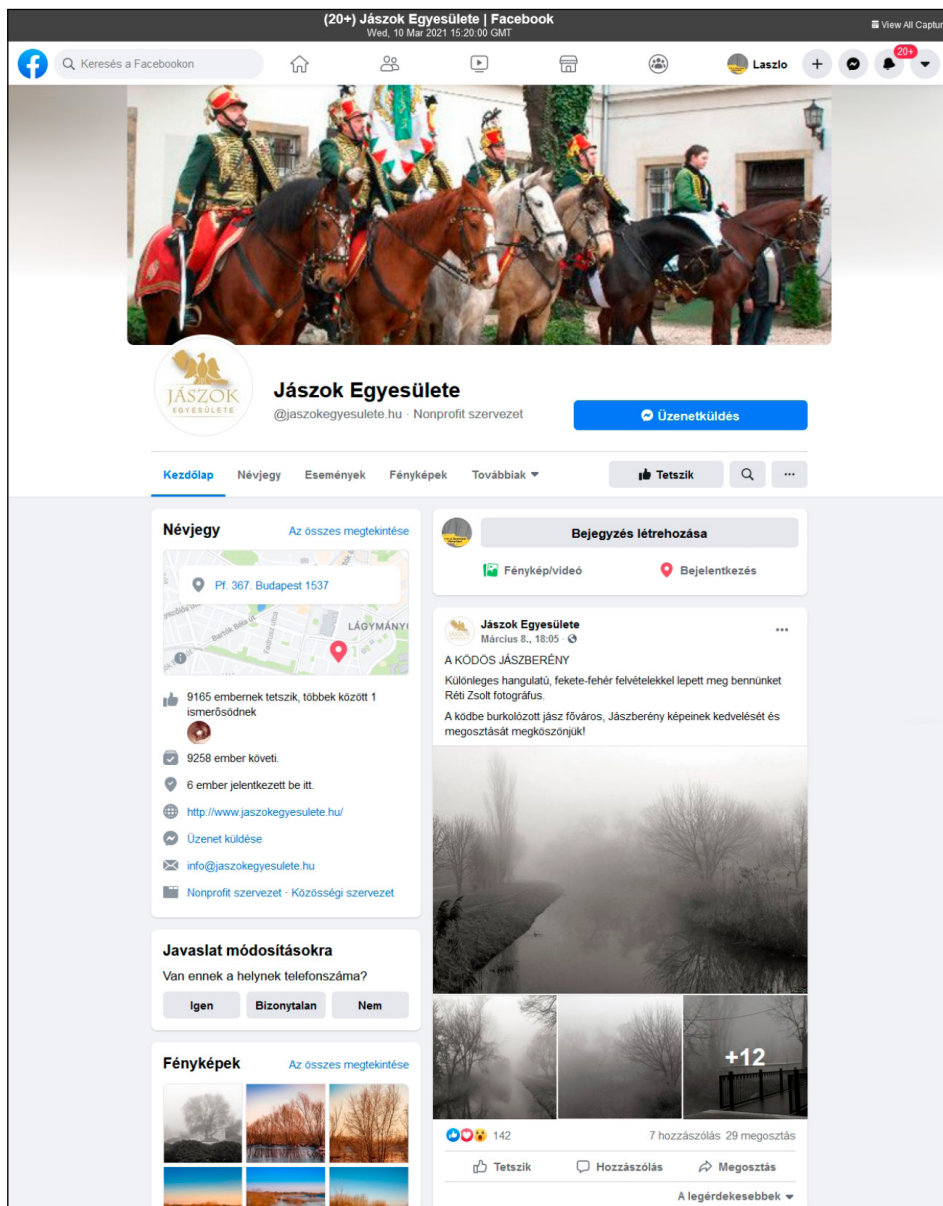
és a webarchívum nyilvántartásai azonosítók alapján kapcsolódnak össze, és a két gyűjtemény jól kiegészíti egymást: előbbiben kiadványszámok szintjén történik az archiválás és a feldolgozás, utóbbiban pedig ezek eredeti szolgáltatási felületét is igyekszünk megőrizni, illetve le tudunk menteni olyan online kiadványokat is, amelyek nem tagolódnak számokra. A koronavírus-járvány híreinek 2020 eleje óta futó gyűjtése mellett négy további eseményalapú archiválást indítottunk (*Nyári olimpia 2020 [2021]*, *Foci EB 2021*, *Nemzetközi Eucharisztikus Kongresszus 2020 [2021]*, *Országgyűlési választás 2022*). Napi

👁 EGY VÉGET NEM ÉRŐ TANULÁSI FOLYAMAT ... 👁

szinten mentettük továbbá 18 hírportál címlapját és az azokról elérhető híreket.

A közösségi média archiválására csak alkalmilag jutott idő, mert ez nagyon élőlátás-igényes folyamat. A Facebook esetében csupán az idővonalat mentjük,

és sokszor azt sem tudjuk egészen a fiók létrehozási dátumáig visszagörgetni. Tavaly főként hírportálok és egyéb periodikák, kormányzati és önkormányzati intézmények, politikai és civil szervezetek Facebook-oldalait archiváltuk egy-egy alkalommal (2. ábra).



2. ábra

Egy civil szervezett archivált Facebook-oldala a PyWb megjelenítőben

Az Instagramról júniusban töltöttünk le kb. 11 700 képet és rövid videót, húsz magyar hírportál fiók-jából. Tavasszal elkezdünk kísérletezni a Twitter-csatornák mentésével is, mert bár ez a platform

Magyarországon még kevésbé népszerű, de a Facebook helyett egyre többen választják ezt a felületet hír- és információmegosztásra. Első lépésben közgyűjtemények, kulturális intézmények, kiadók,


egyházak, civil szervezetek és médiacégek üzeneteit töltöttük le. Jelenleg 2485 Facebook- és 745 Instagram-oldalról, valamint 1003 Twitter-csatornáról van egy-két mentésünk.

Nemzeti webtér szintű aratást a kormányrendelet előírása szerint két alkalommal, júliusban és december végén végeztünk, közel 450 ezer domén- és aldoméncímről indítva a *Heritrix* szoftvert, amely 3,1, illetve 2,4 terabájtnyi webtartalmat archivált.⁷ A kiindulásként használt címlistát a 2020 decemberi 251 ezerhez képest úgy tudtuk lényegesen megnövelni, hogy az előző években lementett weboldalakból kigyűjtöttük a *.hu* végű URL-ekre mutató linkeket, majd ezeket duplumszűrés és többféle, részben emberi munkával elvégzett tisztítási művelet után hozzáadtuk a korábbi webtér-nyilvántartásunkhoz. A tavalyi év egyik fejlesztése volt az is, hogy az URL-címek mellé lekérdeztük ezek *title* metaadatát is (vagyis azt a nevet, amit a webhely gazdája adott az oldalnak), és mindkét adatot kereshetővé tettük.⁸ Mindent egybevéve a zárt archívum mérete több mint 12 terabájtjal nőtt 2021-ben. Az aratások során

oldalképek is készülnek a webhelyek kezdőlapjáról. Ilyenből tavaly közel 1 millió jött létre, 1,8 terabájt összméretben.

Technológia


A megnövekedett archiválási feladatok miatt 2021 augusztusában szükségessé vált a meglévő három mellé egy újabb virtuális szerver üzembe állítása a Kormányzati Informatikai Fejlesztési Ügynökség (KIFÜ) által biztosított felhőben, melyen kizárólag a *Heritrix* aratószoftver fut, hogy ne lassítsa az egyéb munkafolyamatokat. Több hónapos munkával októberre tudtuk csak leindexelni és kereshetővé tenni a *SolrWayback* rendszerrel a nem nyilvános archívumban levő csaknem egymilliárd fájl teljes szövegét. A keresés sebessége elég lassú, de reméljük, hogy ha sikerül winchesterekről SSD meghajtókra átköltöztetni az indexfájlokat, akkor lényegesen javul majd a válaszidő. Ennek a jelenleg még zárt archívumnak elkészítettük és teszteltettük a szolgáltatási felületét (3. ábra).



ORSZÁGOS
SZÉCHÉNYI
KÖNYVTÁR

OSZK Webarchívum

(nem nyilvános gyűjtemény)



OSZK
webarchívum

Az Országos Széchényi Könyvtárban 2017-ben indult az a webarchiválási projekt, melynek eredményeként létrejött egy – elsősorban a technológia demonstrálására szolgáló – nyilvános gyűjtemény, valamint az előbbinél egy jóval nagyobb zárt archívum. A projekt honlapja és a nyilvános szolgáltatás a <https://webarchivum.oszk.hu> címen található, a zárt állomány pedig a <http://webharvest.oszk.hu> oldalon kereshető és böngészhető a webhely neve vagy URL címe szerint az alábbi úrlappal. A találati listából megnyitható az aratások során készült oldalképek nézegetője, valamint az OSZK webarchívumában található mentett anyag kétféle megjelenítéssel: az OWB (Open Wayback) egy régebbi, áttekinthetőbb naplátnézeti szoftver, a PyWb (Python Wayback) pedig egy modern, kis méretű felületet használó eszköz. Előfordulhat, hogy egy archivált weboldal jobb minőségben jelenik meg az egyikben, mint a másikban. A két megjelenítést követően is elérhető a <http://webharvest.oszk.hu/owb/wayback/> és a <http://webharvest.oszk.hu/pywb/> címen.

Van még egy különgyűjtemény csak a Facebook oldalakról: <http://webharvest.oszk.hu/facebook/> és egy másik a naponta archivált hírtörtételekről, amely teljes szöveggel is kereshető a Solrwayback programmal: <https://hirdetelk.webharvest.oszk.hu/solrwayback/>. Ugyancsak külön Solrwayback kereső használható a koronavírus járvány témájú különgyűjteményhez: <https://koronavirus.webharvest.oszk.hu/solrwayback/> és van egy a teljes archívumhoz is, de ez jelenleg eléggé lassú (különösen akkor, ha a duplumok kiszűrésé céljából a GROUPED SEARCH opciót is bekapcsoljuk): <https://webharvest.oszk.hu/solrwayback/>.

A nyilvános és zárt archívumot alkotó különgyűjtemények, a különféle típusú aratások címlistái a <https://webarchivum.oszk.hu/webarchivum/bongeszse/> oldalon böngészhetők.

Részgyűjtemény kereső

Ezzel az úrlappal a zárt archívum tematikus és műfaji különgyűjteményeiben levő webhelyek nevében és/vagy URL címében lehet keresni. Ha mindkét mezőt kitöljük, akkor csak olyan találatok jelennek meg, amelyekre mindkét feltétel teljesül. Mivel többféle cím- és névvariació is előfordulhat, ezért az URL cím esetében a "http://" vagy "https://" előtagot, illetve a "www." aldoméncímet és az esetleges "index.html" vagy "index.php" stb. fájlnevet ne írjuk be, sőt általában a ".hu" országdómén is felesleges. (Tehát például a "https://www.mek.oszk.hu/index.php/ml" helyett elég csak erre rákeresni: "mek.oszk.") Ha pedig névre keressünk, akkor egy-két jellemző, egymás mellett szót kell megadni (pl. "elektronikus könyvtár"). Ez a kereső automatikus szótövezést használ, így a ragozott alakokat is megtalálja (pl. "Magyar Elektronikus Könyvtárért Egyesület").

A találati listában a vastag betűs sor a webhely neve, alatta pedig a központi és az esetleges alternatív URL címe látható (ha ez utóbbiból több is van, akkor csak az első). A következő sorban a szögletes zárójelben levő kód a különgyűjtemény rövid neve (a rövidítések feloldása [itt található](#)). A kereső zárójelben levő dátum pedig a különgyűjtemény utolsó aratásának időpontja. Ez nem feltétlenül jelenti azt, hogy magát a webhelyet is akkor archiváltuk legutóbb, mert ha ezen dátum után került bele a nyilvántartásunkba, akkor még nem készült róla mentés (hár webtér szintű aratásokba esetleg bekerülhetett). A találatok közt az aktuálisan aratott címlistában szereplő webhelyek mellett nem aratott, vagy már nem aratott tételek is megjelenhetnek. Előbbieket csak nyilvántartjuk, de valamilyen ok miatt nem mentjük (pl. extrém nagyok, vagy robbant nem bejárható, vagy más különgyűjtemény keretében archiváljuk). Utóbbiak pedig általában azért kapnak megszűnt státuszt, mert már eltűnt az illető webhely az internetről. Előfordulhat az is, hogy ugyanaz a webhely többször is megjelenik a listában, mert interdiszciplináris jellege miatt több helyen is nyilvántartjuk.

Webhely nevének részlete:

Webhely URL címének részlete:

Webtér szintű kereső

Ezzel az úrlappal a zárt archívumban található webtér szintű aratásba bekerült webhelyek saját, a legtöbb esetben automatikusan begyűjtött *title* adataiban és/vagy URL címében lehet keresni. A kereső működése megegyezik a Részgyűjtemény keresőjével, de mivel az adatok összegyűjtése nagyrészt automatikus módszerekkel történt, ezért ez az adatbázis nagyon "szemetes", sok benne a már nem működő URL, a duplum, a hiányos vagy semmitmondó név.

Webhely nevének részlete:

Webhely URL címének részlete:

3. ábra

A webarchívum nyilvánosan nem elérhető részének keresőfelülete

Az OSZK dolgozói egy védett terminálszerveren át már hozzá tudnak férni ehhez a gyűjteményrészhez, és várhatóan rövidesen az olvasók számára is elérhető lesz.

Az év során a már üzemszerűen használtak mellett tucatnyi további archiváló szoftvert teszteltünk *Windows* és virtuális gépen futó *Linux* alatt, amely szoftverek speciális részfeladatokra használhatók. Ilyen például az *ArchiveBox*,⁹ amely különböző eszközöket integrál egy közös kezelőfelület alá, és különböző formátumokba lehet vele weboldalakat elmenteni (*WARC* konténer, egyetlen *HTML* állomány, különálló fájlokat tartalmazó mappák, *PNG* kép, *PDF* dokumentum, *JSON* szövegfájl). Egy másik érdekes rendszer az önkéntesekből álló *Archive Team* által kifejlesztett *Warrior*,¹⁰ amely egy virtuális gépen fut, és a megszűnés előtt álló vagy más okból veszélyeztetett webes szolgáltatások közösségi archiválásába tudunk általa bekapcsolódni. A *Webrecorder* projekt keretében készülő *ArchiveWeb.page*¹¹ programot és böngészőkiegészítőt már mi is elkezdtük használni a webkettes tartalmak mentésére, mivel ez a *Chrome* segítségével előbb betölti és végiggörgeti az oldalt, végrehajtja a benne levő szkripteket, elindítja a beágyazott videókat, majd az összes fájlt elmenti egy *WARC* konténerbe, így az archív példány sokkal jobban hasonlít az eredetire, mint a *Heritrix*szel mentett. Hátránya, hogy a linkeket nem követi automatikusan, így valakinek végig kell kattintgatnia a letöltendő oldalakat. Hasonló elven működik a szintén ennek a projektnek a részeként fejlesztett *Browsertrix Crawler*,¹² de ebben, ahogy a neve is mutatja, már van egy robot komponens is, amely a konfigurációs fájlban megadott szabályok szerint végigmegy a lementett fájlokban talált linkeken, így akár teljes webhelyek is archiválhatók vele. Bár tömeges aratásokra nem jó, mert a böngészőn keresztül való mentés nagyon lassú folyamat, de még így is gyorsabb, mint az általunk a napi hírek archiválásához jelenleg futtatott *Brozzler*,¹³ úgyhogy tervezzük az erre való átállást az idei év folyamán.

Telepítettük továbbá az egyedi webhelyek mentésére használt *Web Curator Tool*¹⁴ keretrendszer májusban megjelent 3.0-s változatát, de technikai problémákba ütköztünk, és ez egyelőre nehezíti a nyilvános archívum bővítését is. Sikertült viszont beüzemelni és magyar nyelvre fordítani a *PyWb* megjelenítő eszköz legfrissebb verzióját, így most már ez működik a zárt és a nyilvános archívumban is.¹⁵

A tesztelt sokféle szoftver közül érdemes még megemlíteni a *scraper* programokat (pl. *DEiXTo*,¹⁶

TAGS,¹⁷ *Octoparse*,¹⁸ *Web Scraper*,¹⁹ *Screaming Frog*²⁰). Ezek nem teljes weblapok vagy webhelyek letöltésére valók, hanem a felhasználó által megadott elemek kigyűjtésére az oldalakból. Ezek az elemek lehetnek szövegrészek, képek és egyéb beágyazott médiafájlok, linkek, metaadatok stb. Elsősorban kutatási, adatbányászati célokra használják ezeket a szoftvereket, de egy webarchívum számára is hasznosak lehetnek, mert vannak köztük olyanok, amelyek a közösségi platformok programozható csatolóján (az *API*-n) keresztül tudnak letölteni fájlokat és adatokat. Így ha nem fontos az amúgy is nehezen archiválható és reprodukálható felhasználói felület megőrzése a *Facebook*, az *Instagram*, a *Twitter*, a *YouTube* és társaik esetében, akkor ezeknek a *scraper* programoknak a segítségével a lényegi tartalom kinyerhető és eltárolható.

Ismeretterjesztés

Amikor 2017-ben elkezdtünk foglalkozni az OSZK-ban a webarchiválással, húsz évnyi megtanulnivalónk volt azokhoz a külföldi könyvtáros és informatikus kollégákhoz képest, akik már a kilencvenes évek második fele óta dolgoznak ezen a területen. Annak érdekében, hogy minél gyorsabban csökkenteni tudjuk a lemaradást, a „csinálva tanulás” mellett döntöttünk, vagyis azonnal megpróbáltunk minden megszerzett ismeretet átültetni a gyakorlatba. A projekt indulása után nem sokkal ezt a *learning by doing* hozzáállást kiegészítette a *learning by teaching*, azaz a „tanítva tanulás” módszere is. A Könyvtári Intézet segítségével évente többször is meghirdetünk egy akkreditált továbbképzési tanfolyamot *Az internet archiválása mint közgyűjteményi feladat* címmel.²¹ A tanfolyami anyag naprakészen tartása, a résztvevők eltérő intézményi háttere és tudásszintje rákényszerít bennünket arra, hogy a napi munkához szükségesnél mélyebben és szélesebb körben is beleássuk magunkat az internetes tartalmak megőrzésének és kutathatóvá tételének egyes részterületeibe. 2021-ben tavasszal és ősszel két-két alkalommal tartottuk meg ezt a négy napos tanfolyamunkat, a pandémia miatt online formában. A távoktatás előnye, hogy így többen tudtak a képzésen részt venni, hátránya viszont, hogy a gyakorlati feladatok közös megoldására nem volt mód. Már jóval száz fölött van azoknak a könyvtárosoknak, informatikusoknak, múzeumi és egyetemi szakembereknek a száma, akik elvégezték ezt a tanfolyamot, és mivel az érdeklődés továbbra sem csökken, ebből arra következtethetünk, hogy egyre több hazai intézményben kezdenek el legalább a tervezés szintjén

foglalkozni vagy akár kísérletezni is a digitálisan születő webtartalmak archiválásával.

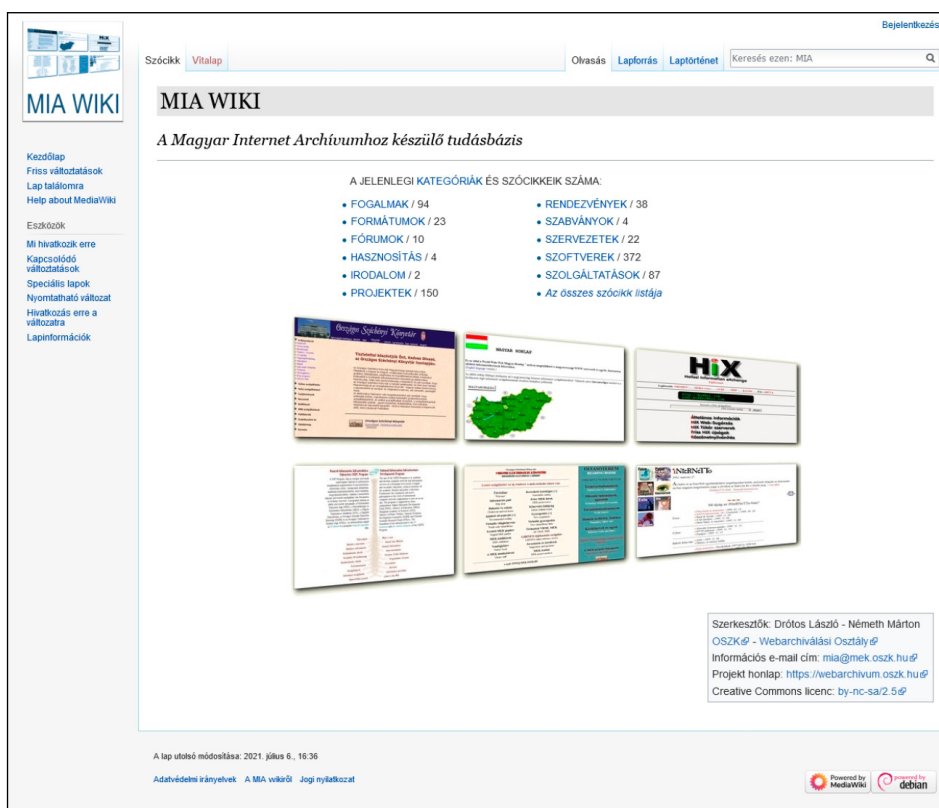
Az egyetemi oktatásba való beépülésnek is vannak már jelei. Az ELTE BTK Könyvtár- és Információtudományi Intézetének *Információforrások 2.* kurzusán *Németh Márton* webkönyvtáros kollégánk tavasszal egy bevezető jellegű vendéglőadást tartott a témáról, továbbá egy ELTE-s hallgató nálunk végezte el a nyári szakmai gyakorlatát. Megszületett az első magyar nyelvű disszertáció is erről a szakterületről, melyet szintén *Németh Márton* készített a Debreceni Egyetem Informatikai Tudományok Doktori Iskolájának hallgatójaként, és védett meg *summa cum laude* minősítéssel október 15-én. A dolgozat címe: *A webarchiválás elméletének és gyakorlatának alapelemei* (nyilvánosan elérhető az egyetem repozitóriumában²²).

Kollégáimmal együtt 2021-ben öt tanulmányt publikáltunk hazai szaklapokban, egyet pedig egy angol nyelvű szlovákiai folyóiratban, és további cikkeink

vannak megjelenés alatt. Írtunk továbbá egy fejezetet a webarchiválás módszertanáról a *Fehér Könyv* most készülő 2. kiadásába, mely korábban csak a közgyűteményi digitalizálással foglalkozott.

2021. november 23–24-én ötödik alkalommal szerveztük meg a „404 Not Found – Ki őrzi meg az internetet?” című konferenciát és workshopot, az előző évihez hasonlóan *Teams*-en, de ezúttal két napra bontva. Az első napon a környező országok webarchívumai mutatkoztak be, a másodikon pedig magyar előadások hangzottak el²³ – ezeket a lapban további írások foglalják össze. A saját rendezvényünkkel együtt 2021-ben 5 hazai és 10 külföldi (többségében online) szakmai eseményen vettünk részt, és 7 magyar és 4 angol nyelvű előadást tartottunk.²⁴

A nyár folyamán 217 új címmel bővítettük a webarchívum honlapján található szakirodalmi bibliográfiát²⁵ és 136 szócikkkel a wikit,²⁶ így már mindkettő több mint 700 tételből áll (4. ábra).



4. ábra

Az internetes tartalmak archiválásával foglalkozó wiki kezdőlapja

Szintén nyáron került fel a honlapunkra egy középis-kolásoknak szánt oktatási segédlet vázlatra *Mentsük le*

az internetet! – Internetes tartalmak megőrzése intézményi és személyes archiválással címmel.²⁷

Kapcsolatok

Májustól elkezdtük felvenni a kapcsolatot a megyei könyvtárakkal, hogy bevonjuk őket bizonyos munkafázisokba, segítve ezzel egyrészt a nemzeti szintű webarchívum építését, másrészt pedig a helyi digitális gyűjtemények webtartalmakkal való kibővítését. Az együttműködést az egyedi digitális dokumentumok (könyvek, időszaki kiadványok, képi dokumentumok) körére is szeretnénk kiterjeszteni, melyek a 2022-ben megújuló *Magyar Elektronikus Könyvtár* és a másik két kapcsolódó szolgáltatás, az *Elektronikus Periodika Archivum és Adatbázis*, illetve a *Digitális Képtár* állományát gazdagíthatják majd. A megkeresésre pozitívan reagáló intézmények közül augusztus óta 11-gyel tartottunk megbeszélést legalább egy alkalommal az együttműködés részleteiről, és 9 megyénél már le is válogattunk egy-egy címlistát a nyilvántartásunkból, amely a leendő regionális gyűjtemény kiindulópontja lehet. December elején a székesfehérvári Vörösmarty Mihály Könyvtár munkatársaival pedig már egy konkrét feladatról egyeztetünk. A könyvtár a 2022-es *Aranybulla Emlékév*hez kapcsolódva létre kíván hozni egy digitalizált és digitálisan született dokumentumokból, köztük archivált weboldalakból álló gyűjteményt, amihez az OSZK részéről módszertani és technikai segítséget tudunk biztosítani. Ugyancsak egy esetleges közös

Irodalmi hivatkozások

1. Az OSZK Webarchívum honlapja. Havi hírosszefoglalók aloldal [online]. Feltöltve: 2021. 05. 26. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/havi-hirosszefoglalok/> [Megtekintve: 2022.01.17.]
2. Az OSZK Webarchívum honlapja. Demó archivum aloldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/demo-kezdolap/> [Megtekintve: 2022.01.17.]
3. 626/2020. (XII. 22.) Korm. rendelet a webarchiválás részletes szabályairól. == Hatályos jogszabályok gyűjteménye [online]. Budapest: Wolters Kluwer, [s.a.]. Hozzáférhető: <https://net.jogtar.hu/jogszabaly?docid=a2000626.kor> [Megtekintve: 2022.01.17.]
4. OSZK Webarchívum, összeáll. Webhelyarchiválási javaslat [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://forms.office.com/r/Cz3GsHSQEd> [Megtekintve: 2022.01.17.]
5. Az OSZK Webarchívum honlapja. OSZK-s webhelyek archívuma aloldal [online]. Budapest: OSZK Webarchívum,

projektet előkészítő megbeszélésekre került sor vagy több alkalommal is az ELTE BTK Digitális Bölcsészeti Tanszékének munkatársaival a webarchívum kutatási célú hasznosításáról.

Nemzetközi téren aktívan részt vettünk az *International Internet Preservation Consortium* szervezet munkájában és webináriumain, valamint a *WARCnet* kutatási projektben. A *404 Not Found* konferencián kezdeményeztünk egy közép-európai együttműködést *Central European Web Archives* néven, illetve ennek részeként egy közös portált és keresőt. Ennek a projektnek a részletes kidolgozása és koordinálása 2022 egyik fontos feladata lesz.

December végén a felek aláírták a szerződést az OSZK és az Internet Archive között, melynek értelmében az amerikai archívum kialakít egy keresőfelületet a nála található, 1996 óta gyűjtött magyar webtartalomhoz, valamint átadja az általa aratott *.hu* domének listáját. Ennek az új keresőportálnak a beépítése a nemzeti könyvtár szolgáltatásai közé szintén kiemelt cél az idei évre.

E cikk írásakor még csak néhány nap telt el a 2022-es évből, de máris elkezdett szépen megtelni a Webarchiválási Osztály új feladatait nyilvántartó táblázat, úgyhogy minden esély megvan rá, hogy jövőre is lesz miről beszámolnunk a téma iránt érdeklődőknek.

6. Az OSZK Webarchívum honlapja. Rákóczi webarchívum aloldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <http://rakoczi2019.webarchivum.oszk.hu/> [Megtekintve: 2022.01.17.]
7. Az OSZK Webarchívum honlapja. Webtér szintű aratások aloldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/webarchivum/statisztikai-adatok/webter-szintu-aratasok/> [Megtekintve: 2022.01.17.]
8. Az OSZK Webarchívum honlapja. Keresés a webtér címlistában aloldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/keresesa-webter-cimlistaban/> [Megtekintve: 2022.01.17.]
9. ArchiveBox [számítógépes szoftver]. GitHub Inc., cop. 2022. Hozzáférhető: <https://archivebox.io> [Megtekintve: 2022.01.17.]
10. ArchiveTeam Warrior. == ArchiveTeam [online], 2009–. A lap utolsó módosítása: 2022. 01.14. Hozzáférhető: <https://wiki>.

- archiveteam.org/index.php/ArchiveTeam_Warrior [Megtekintve: 2022.01.17.]
11. Archiveweb.page [számítógépes szoftver]. GitHub Inc., cop. 2022. Hozzáférhető: <https://archiveweb.page> [Megtekintve: 2022.01.17.]
 12. Webrecorder/browsertrix-crawler [számítógépes szoftver]. GitHub Inc., cop. 2022. Hozzáférhető: <https://github.com/webrecorder/browsertrix-crawler> [Megtekintve: 2022.01.17.]
 13. Internetarchive / brozzler [számítógépes szoftver]. GitHub Inc., cop. 2022. Hozzáférhető: <https://github.com/internetarchive/brozzler> [Megtekintve: 2022.01.17.]
 14. WebCuratorTool / webcurator [számítógépes szoftver]. GitHub Inc., cop. 2022. Hozzáférhető: <https://github.com/WebCuratorTool/webcurator/> [Megtekintve: 2022.01.17.]
 15. Pywb részgyűjtemény keresőoldala [online]. == Az OSZK webadmin felülete. Hozzáférhető: <http://webadmin.oszk.hu/pywb/> [Megtekintve: 2022.01.17.]
 16. DEiXTo [számítógépes szoftver]. ΔEiXTo, cop. 2007–2022. Hozzáférhető: <https://deixto.com> [Megtekintve: 2022.01.17.]
 17. A TAGS – Twitter Archiving Google Sheet honlapja [online]. TAGS, cop. 2022. Hozzáférhető: <https://tags.hawksey.info> [Megtekintve: 2022.01.17.]
 18. Az Octoparse honlapja [online]. Octopus Data Inc., cop. 2021. Hozzáférhető: <https://www.octoparse.com> [Megtekintve: 2022.01.17.]
 19. A WebScraper honlapja [online]. Riga: Web Scraper, cop. 2022. Hozzáférhető: <https://webscraper.io> [Megtekintve: 2022.01.17.]
 20. A Screamingfrog honlapja [online]. Henley-on-Thames: Screaming Frog Ltd., cop. 2022. Hozzáférhető: <https://www.screamingfrog.co.uk/seo-spider/> [Megtekintve: 2022.01.17.]
 21. Az internet archiválása mint közgyűjteményi feladat [online]. == A Könyvtári Intézet honlapja. Képzés oldal [online]. Budapest: Könyvtári Intézet, cop. 2022. Hozzáférhető: <https://ki.oszk.hu/tanfolyamok/az-internet-archivalasa-mint-kozgyujtemenyi-feladat> [Megtekintve: 2022.01.17.]
 22. NÉMETH Márton. A webarchiválás elméletének és gyakorlatának alapelemei: A szervezett keretek között zajló webarchiválás kezdetei Magyarországon: Doktori (PhD) értekezés [online]. Debrecen: Debreceni Egyetem Természettudományi és Informatikai Doktori Tanács Informatikai Tudományok Doktori Iskola, 2021.== Debreceni Egyetem elektronikus Archivum. PhD dolgozatok: Informatikai Tudományok Doktori Iskola oldal. Hozzáférhető: <https://dea.lib.unideb.hu/dea/handle/2437/310638> [Megtekintve: 2022.01.17.]
 23. Az OSZK Webarchívum honlapja. "404 Not Found" workshop – 2021. november 23–24. oldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/szakembereknek/404-not-found-workshop/404-not-found-workshop-2021-november-23-24/> [Megtekintve: 2022.01.17.]
 24. Az OSZK Webarchívum honlapja. Előadások, prezentációk, publikációk oldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/a-projektrol/eloadasok-prezentaciok-publikaciok/#2> [Megtekintve: 2022.01.17.]
 25. Az OSZK Webarchívum honlapja. Szakirodalmi bibliográfia oldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/szakembereknek/szakirodalmi-bibliografia/> [Megtekintve: 2022.01.17.]
 26. Az OSZK Webarchívum honlapja. MIA wiki oldal [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/mediawiki/> [Megtekintve: 2022.01.17.]
 27. DRÓTOS László, szerk. Mentsük le az internetet! Internetes tartalmak megőrzése intézményi és személyes archiválással: E-learning tananyag vázlata [online]. Frissítés dátuma: 2021. 07. 08. == Az OSZK Webarchívum honlapja [online]. Budapest: OSZK Webarchívum, cop. 2022. Hozzáférhető: <https://webarchivum.oszk.hu/szakembereknek/tanfolyam-es-e-learning/mentsuk-le-az-internetet/>

Felhasznált irodalom

- „404 Not Found – Ki őrzi meg az internetet?” workshop [online]. Budapest: Országos Széchényi Könyvtár, 2021. Hozzáférhető: <https://webarchivum.oszk.hu/404-not-found-workshop-2021-november-23-24/> [Megtekintve: 2022.01.17.]
- DRÓTOS László. Az OSZK webarchívumának újdonságai. == Könyvtári Figyelő, Ú.f. 30. = 66. (2020) 1., p. 67–73. ISSN 0023-3773 Hozzáférhető: <http://ojs.elte.hu/kf/article/view/1065> [Megtekintve: 2022.01.17.]
- DRÓTOS László. Az OSZK webarchívumának 2020-as újdonságai. == Könyvtári Figyelő, Ú.f. 31. = 67. (2021) 1., p. 31–38. ISSN 0023-3773 Hozzáférhető: <http://ojs.elte.hu/kf/article/view/2296> [Megtekintve: 2022.01.17.]
- DRÓTOS László. Az idő fogságában: Ki őrzi meg a közösségi médiát? == Tudományos és Műszaki Tájékoztatás, 68. (2021) 7., p. 428–439. ISSN 1586-2984 Hozzáférhető: <https://tmt.omikk.bme.hu/tmt/article/view/13062> [Megtekintve: 2022.01.17.]
- NÉMETH Márton – DRÓTOS László. A webarchiválás oktatásának nemzetközi keretei és hazai tapasztalatai. == Könyvtári Figyelő, Ú.f. 31. = 67. (2021) 3., p. 405–412. ISSN 0023-3773 Hozzáférhető: <http://ojs.elte.hu/kf/article/view/2954> [Megtekintve: 2022.01.17.]
- VISKY Ákos László. Együtműködési lehetőségek a webarchiválás területén. Könyvtári Figyelő, Ú.f. 31. = 67. (2021) 1., p. 39–45. ISSN 0023-3773 Hozzáférhető: <http://ojs.elte.hu/kf/article/view/2297> [Megtekintve: 2022.01.17.]

(Beérkezett: 2022. január 17.)