Gyorsmérleg

– az OSZK Webarchívumának és néhány könyvtárnak a KDS-K projekt keretében történt együttműködéséről

készítette: Visky Ákos László

utolsó módosítás: 2020-07-08

Most, hogy a végéhez értünk a Közgyűjteményi Digitalizálási Stratégia könyvtári ágának keretében (KDS-K) megvalósult együttműködésnek – mely során az Országos Széchényi Könyvtár (OSZK) Webarchívuma és a projektben nyertesként résztvevő megyei hatókörű városi könyvtárak elsősorban a nemzeti webtér regionális vonatkozású webhelyeinek feltárásában működtek együtt –, illő rövid összefoglalót adni a közös munka eredményéről.

Talán a legfontosabb eredmény, hogy reményeink szerint nem ért véget a közös munka… Nem ért véget, mert néhány résztvevővel (elsősorban a kecskeméti Bács-Kiskun Megyei Katona József Könyvtárral és a tatabányai József Attila Megyei és Városi Könyvtárral) már konkrét megbeszéléseket folytattunk a lehetséges további együttműködésről, és amivel kapcsolatban már más könyvtárak is kifejezték érdeklődésüket. Az együttműködés szorosabbra fűzése érdekében elkezdtünk előkészíteni egy megállapodás tervezetet is a jövőbeli közös munka lehetséges kereteiről, amelyről minden érintettel egyeztetni fogunk. Örömmel láttuk, hogy sokan felismerték az élő és az archivált webtartalom fontosságát, helyét a saját, helyi gyűjteményeikben.

Nem ért véget a közös munka, mert a java még hátra van… Egyrészt a webtartalom változékonysága megköveteli a most létrehozott állomány folyamatos gondozását, hogy a közös eredmény ne enyésszen el idővel: hiába gyarapodott most a Webarchívum sok új és értékes webhellyel, hiába kerülnek be ezek a tételek a helyi gyűjteménybe, ha a ma adatai inflálódnak az idő múlásával. Másrészt célszerű bővíteni az együttműködők körét újabb partnerek bevonásával (nemcsak könyvtárakkal, de más közgyűjteményekkel is), hogy finomítani és bővíteni lehessen az elért eredményeket a közös feladat, a jelen webes kultúrájának megőrzésének érdekében. Éppen ezért, a munka során használt megosztott táblázat továbbra is elérhető marad, és örömmel vennénk, ha továbbra is kerülne bele új tétel (ezekről szeretnénk értesítést is kapni, hogy észrevegyük), különösen azokban a témakörökben, amelyekben aránytalanul kevés cím gyűlt össze (l. az értékelő részt). Távlatilag szeretnénk a többi megyei hatókörű városi könyvtárral, valamint a szak- és egyetemi könyvtárakkal is felvenni a kapcsolatot és a mostanihoz hasonló együttműködést kialakítani, ahogyan más, a webarchiválásra nyitott közgyűjteménnyel is, és persze ne feledkezzünk el a határon túli magyar kulturális intézményekről sem!

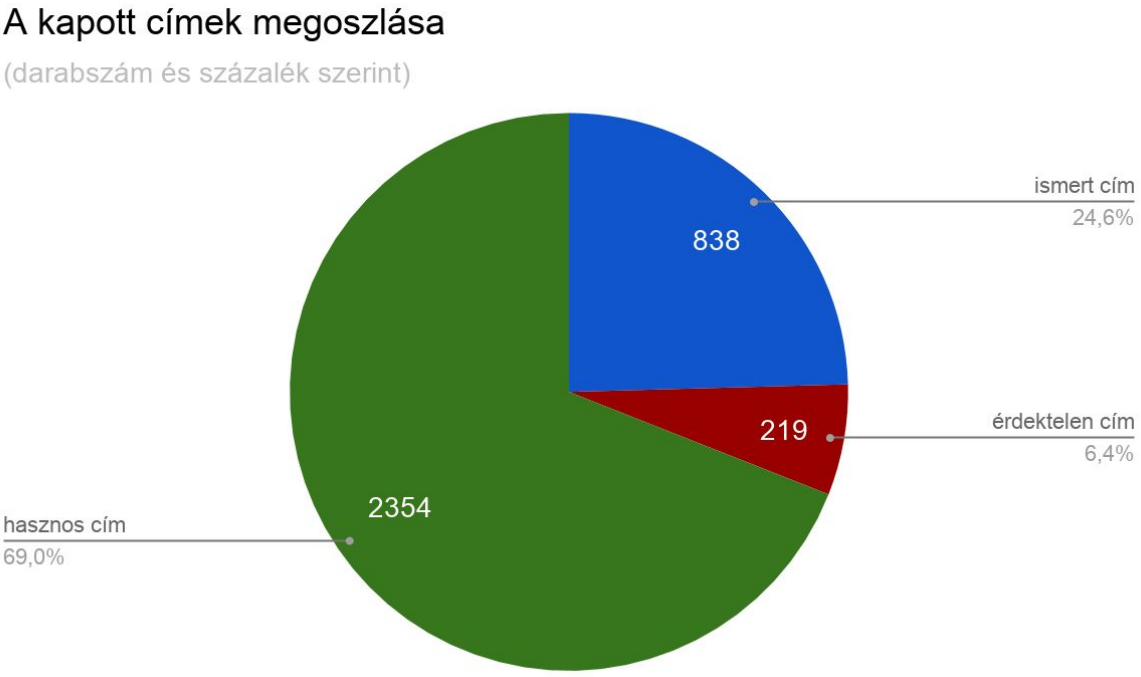
Most következzen számszerüsítve is az elmúlt hónapok munkájának eredménye, de előtte még egy rövid kitérő. A projekt kiírásában a webarchiválás támogatása szerepelt feladatként a nyertesek számára, amit úgy fordítottunk le az egyeztetések során, hogy a közreműködő partnerek megpróbálják feltérképezni a megyei (esetleg régiós) webtartalmat, elsősorban a megadott tematika szerint, majd ezek címeit és bizonyos adatait – a saját felhasználás lehetősége mellett – átadják a Webarchívumnak, kiegészítendő a nemzeti könyvtárban eddig már összegyűjtött állományt. Feltétel volt, hogy csak azok a címek kerüljenek átadásra, melyek még nem szerepelnek a Webarchívum nyilvántartásában (ezt kezdetben egy – sajnos nem naprakész – űrlappal, majd az időközben elkészült új honlap keresőjével lehetett ellenőrizni), és az is, milyen szempontoknak és paramétereknek kell megfelelniük a tételeknek (l. Feladatok a KDS-K-projektben a webarchiválás támogatására – kritériumok, követelmények című, a MIA-l levelező-listára 2020. 03. 13-án kiküldött dokumentum). Emellett még szó volt vállalható egyéb feladatokról is, például az archivált tartalmak és oldalképek ellenőrzéséről, metaadatolásáról, de ezekből – valószínűleg különböző okokból – már semmi nem valósult meg. Kézzelfogható eredményt hozott viszont az, hogy egy könyvtártól kaptunk olyan kapcsolati adatokat is, melyek határon túli webhelyek feltérképezésében lesznek hasznosak.

A kapott tételek darabszáma és százalékos megoszlása

Az OSZK Webarchívum jelenleg 34 ezer élő webhelyet és URL-jeit tart nyilván különböző gyűjteményekben, de bizonyos tematikákhoz tartozók még nem kerültek összegyűjtésre, valamint a közösségi médiával is még éppen csak elkezdtünk foglalkozni. Persze maga az összegyűjtés kifejezés is problémás, mert gyakorlatilag teljeskörű gyűjtemény nem létezik a webhelyek változékonysága, illetve az ismeretszerzés korlátai miatt.

A pályázatban résztvevő könyvtárak nagyságrendileg összesen ennek az állománynak kb. 10%-ra rúgó webhelyet gyűjtöttek össze és ajánlottak a Webarchívum figyelmébe, melyből végül is hasznos címként az eddigi gyűjtemény 7%-val fog gyarapodni, ebből 2% jut a Webarchívum által korábban már feltárt tematikájú gyűjteményekbe, és 5%-nyi kerül a szisztematikusan még nem feltérképezettekbe.

Június végéig – eltekintve néhány duplumtól – összesen 3411 db új tételt (plusz 31 db alternatív URL-t) kaptunk a 11 partnerünktől, melyet egyesével ellenőriztünk ismertség, relevancia és formai szempontok (név és címalak) szerint. Ennek a munkának eredményeként 3491 db besorolásra váró webhely nézett velünk farkasszemet, ugyanis 12. könyvtárként az OSZK is gyarapította saját gyűjteményét 80 db még ismeretlen tétellel (véletlenszerűen talált, pl. kapcsolódó címek feltárásával)… A 3411 db címből az ellenőrzés során 838 db (24,6%) bizonyult általunk már ismertnek, 219 db (6,4%) érdektelennek és 2354 db (69,0%) hasznos, új címnek, amihez persze még hozzáadódik 80 db OSZK gyarapítás is (a bővített darabszám 2,3%-a).



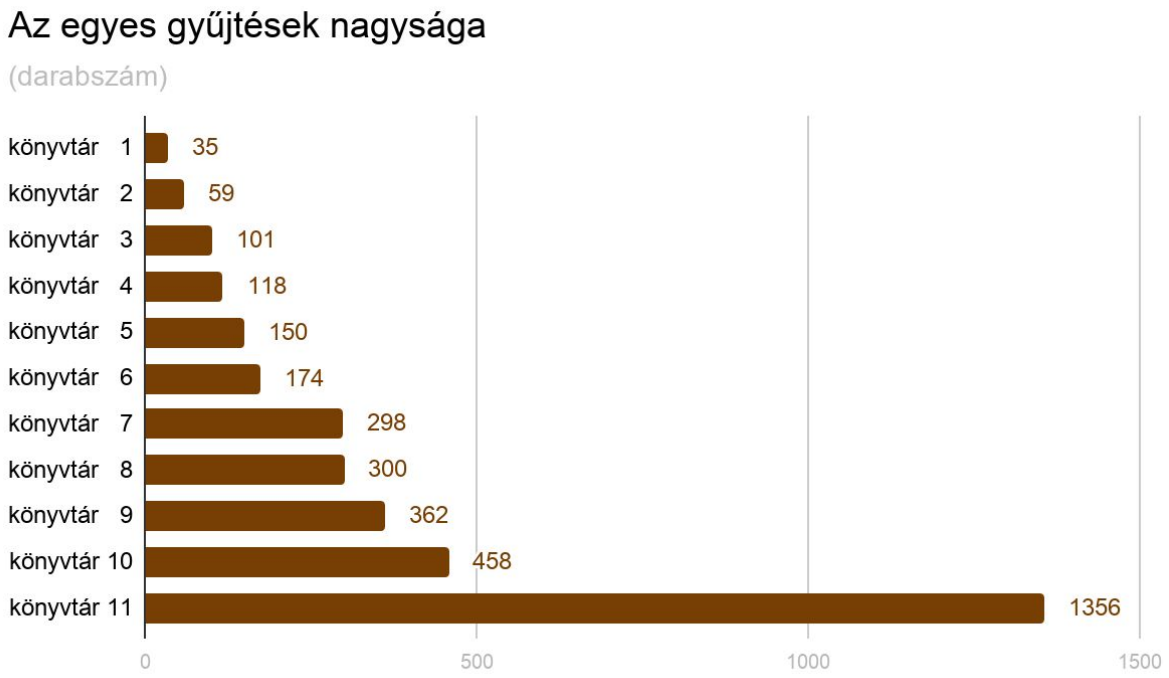
Felmerül a kérdés, miből adódtak az ismert és az érdektelen tételek? Azt minden résztvevő fel tudja mérni, mennyire tartotta magát a kért szempontokhoz (ellenőrzés, gyűjtőkör)... Az ismert címek tekintetében azonban az mindenképpen árnyalja a képet, hogy a Webarchívum a projekt teljes ideje alatt nem tudta biztosítani a naprakész címellenőrzést (csak időközben készült el az új kereső); illetve a véletlenek összjátéka folytán májusra jött létre az oktatási webhelyekről egy új tematikus gyűjtemény az Oktatási Hivataltól kapott lista alapján, emiatt az adatok feldolgozásakor néhány iskola honlapja már az ismert címek közé került. Az érdektelen tételek esetében az lehetett zavaró tényező, hogy az esetleg helyismereti szempontból fontos és nyilvántartandó aloldal a Webarchívum számára már ismert főoldal része, így nem érdekes számára, és ez nem lett kiszűrve adatfeltöltés előtt. Szerencsére voltak olyan könyvtárak is, amelyek már a gyűjtés során megmutatták addigi munkájukat, így azt észrevételezve elkerülhető vagy csökkenthető volt az esetleges hibás gyakorlat folytatása. Mindezek miatt, valamint azért, mert végül is a résztvevők számára új szempontok szerint kellett a gyűjtést végezni, első próbálkozásra nem tekinthető rossznak az eredmény, egy esetleges újabb gyűjtés vagy folyamatos együttműködés esetén várhatóan javulna a hatékonyság.

A gyűjtések nagysága, eloszlása

Nem szeretnénk külön kitérni az egyes résztvevő könyvtárak munkájának értékelésére, se darabszám, se hasznosság tekintetében. Reális értékeléshez sok egyéb szempontot is

figyelembe kellene venni, azok nélkül hamis következtetések születnének egy-egy számszerűsített mutatóból – ezekről az előzetes munkamegbeszéléseken is szó volt.

Számszerűsítve 35, 59, 101, 118, 150, 174, 298, 300, 362, 458 és 1356 tételszámot kaptuk az egyes könyvtáraktól, összesen 3411 db webhely nevét, címét és esetleg egyéb adatokat. A két szélső érték között majdnem 40-szeres a különbség, a gyűjtések zöme a 100 és 500 közötti darabszámból került ki. Átlagnak a 310 tétel tekinthető, ami a két szélső érték kivonásával 224 lesz. Csak érdekességként néztük meg, hogy a legnagyobb gyűjtemény hasznosság tekintetében eltér-e az összesített átlagtól, és meglepődve tapasztaltuk, hogy szinte megegyeznek az arányok: a felvett címeknek ugyanúgy 69,0%-a bizonyult hasznosnak, mint az összesített gyűjtemény esetében, ismert címből viszont kevesebbet, érdektelenből többet kaptunk az összesítés átlagához képest.



A hasznos új címek tematikus megoszlása, besorolása

A projekt keretében – a partner könyvtárak és az OSZK révén – összegyűlt és a Webarchívum által még nem ismert hasznos címeket érdemes kettéválasztani aszerint, hogy általunk korábban már szisztematikusan gyűjtött tematikáról van szó, vagy olyanról, mellyel kapcsolatban még nem volt érdemi címfeltárás. (Mindkét esetben az elsődleges tematika alapján került besorolásra az adott tétel és az alapján szerepel az összesítésekben.)

A Webarchívum által korábban már összegyűjtött tematikák esetében 715 db új címet kaptunk, ez az eddig összegyűjtött 34 ezer webhely kb. 2%-a, a még nem gyűjtött tematikák esetében kapott 1719 db új cím az eddig állomány 5%-ra rúg.

Lássuk témakörök szerint a címek megoszlását az eddig ismert és a kapott új tételek szerint.

Témakör*	Nyilvántartott	Új cím	Százalékban
Egyetemek, főiskolák	3224 db	4 db	0,12%
Elektronikus periodikák	5232 db	53 db	1,01%
Irodalom, irodalomtudomány és -történet	1188 db	14 db	1,17%
Könyv- és egyéb kiadók, kereskedők	1253 db	28 db	2,23%
Kormányzat, önkormányzatok, politikai és civil szervezetek	3609 db	93 db	2,57%
Könyvtárak, levéltárak, múzeumok és galériák	1836 db	43 db	2,34%
Kutatóintézetek, tudományos szervezetek	850 db	8 db	0,94%
Média, sajtó, műsorszórás	805 db	12 db	1,49%
Képző-, előadó-, zene- és filmművészet	7043 db	107 db	1,52%
Kulturális intézmények, művelődési házak, rendezvényhelyszínek	660 db	38 db	5,75%
Közoktatás és egyéb képzések	5132 db	294 db	5,73%
Vallások, hitrendszerek, egyházak	2219 db	21 db	0,95%

** Bizonyos esetekben nem tematikus, hanem műfaji besorolást alkalmazunk, ilyen esetekben ez az elsődleges szempont. Így kerültek külön gyűjteménybe az e-periodikák és a közösségi tartalmak (pl. Facebook-oldak, Instagram-fiókok, stb.) is. A táblázat adatai ezt tükrözik, így pl. egy múzeum Twitter-csatornája nem a közgyűjteményi webhelyek közé kerül.*

Érdekes látni, hogy a 12 csoportból 7-ben az átlag alatt maradt a gyarapodás, háromnál kicsivel, de két csoportban messze átlag feletti a bővülés. A kulturális intézmények, rendezvényhelyszínek átlag feletti bővülése esetében vélhetően visszaigazolódik az az előzetes koncepció, hogy a helyismeret előnyt jelent lokális tartalmak gyűjtésének esetében. Az oktatási webhelyek esetében a kiugró arányban valószínűleg szerepet játszik az is, hogy

az a legfrissebb gyűjtemény, és még nem sikerült olyan szinten feltárni a vonatkozó tartalmat, mint a régebbi gyűjteményeknél (pl. a szakképzési centrumok esetében csak a központi intézmény szerepel a kapott nyilvántartásban, a tagiskolák nem).

A Webarchívum által eddig szisztematikusan még nem gyűjtött tematikák esetében a következő méretű gyűjtések születtek. (Ezeknél a gyűjteményeknél nem vettük figyelembe, hogy esetleg már hány címet tartunk nyilván bennük.)

Témakör*	Új cím
Bölcsész- és társadalomtudományok	2 db
Egészségügy, szociális szféra	32 db
Életmód, szabadidő, hobbi	22 db
Facebook oldalak	596 db
- egyéb közösségi oldalak	13 db
Idegenforgalom, vendéglátás	477 db
Ipar, mezőgazdaság, munkaügy	141 db
Sport, testkultúra	311 db
Szolgáltatás, kereskedelem, közlekedés, marketing	43 db
Természet- és műszaki tudományok	39 db
Történelem, hely- és családtörténet	43 db
* *A besorolás elve megegyezik az előző táblázatban leírtakkal.*	

Nem meglepő módon a legnagyobb csoport a Facebook (bár különböző mértékben lett felvéve a gyűjteményekbe), amit az idegenforgalom-vendéglátás és a sport követ. Érdekesen árnyalja a rangsort, hogy az idegenforgalom-vendéglátás kategória lett a legtöbbször megadva másodlagos besorolásnak, és bár ezt csak indokolt esetben alkalmaztuk, jól mutatja, mely ágazatnak van leginkább kapcsolódása más területekkel.

Az is látszik a számokból, hogy még jócskán vannak fel nem tárt webhelyek, bőven van tere a további munkának, és nem csak a Facebook esetében (az önmagában is hihetetlenül nagy és növekvő súlyú). Az ország lakosságának ugyan kisebbik felét fedik le azok a megyék, melyek könyvtárai most részt vettek ebben projektben, de bizonyos témakörökbe gyanúsan kevés tétel került. Ha arra gondolunk, hogy csak állami fenntartásban 91 kórház van, de ezek mellett számtalan szakrendelő, magánkórház és -rendelő, valamint szociális intézmény van az országban, akkor a kapott 31 cím elég kevésnek tűnik… Láthatóan a sport tematika jobban megmozgatta a gyűjtést végző kollégákat, talán a koronavírus járvány miatti kényszerű otthonlét okán is. Az idegenforgalom-vendéglátás ugyan nagyobb számmal szerepelt az összesítésben, de ez a tematika önmagában is hatalmas lenne, ahogy a szolgáltatás-kereskedelem vagy az életmód-szabadidő-hobbi kategória is, amikből viszont megint nagyon kevés címet kaptunk, és sokkal több helytörténettel foglalkozó webhelyet is vártunk.

Talán azt külön nem is érdemes kiemelni, hogy a helyismeret előnye visszaköszön a bővülésben, hisz pont ez volt a cél, azt viszont már igen, hogy ennek köszönhetően kiemelt gyűjtőköri webhelyeket is megismertünk, pl. eddig rejtőző periodikákat vagy kulturális intézményeket.

Munka után – munka előtt...

A projekt során kapott webhely címeket felvesszük tematikus gyűjteményeinkbe és negyedéves ütemezéssel mentéseket készítünk róluk. Ezen kívül az összes tétel már bekerült a június végén indított magyar webtér aratásba, mely során a tematikus gyűjteményekben szereplő címeknél nagyságrendileg bővebb körről készítünk azoknál kisebb mélységű pillanatfelvételt. Sajnos a közösségi média webhelyei esetében nem ilyen jó a helyzet, mert azok archiválása más jellegű technológiát igényel, és erre jelenleg nincs igazán kiforrott automatikus eljárás, ezért a legfontosabb tételek esetében kézzel készítünk egyedi mentéseket, a többi címet csak nyilvántartjuk.

A KDS-projekt keretében végzett közös munkáról az immár rendszeres őszi workshopunkon is szeretnénk számot adni. A rendezvény tervezett témája az intézményi együttműködés lesz, és az előadások mellett – melyre a mostani partnereink közül is szeretnénk felkérni előadót, aki a saját és intézménye szemszögéből értékelné a közös munkát – nagyobb hangsúlyt kapnának a gyakorlati szempontok is.

Szolgálati közleményként szeretnénk még megemlíteni, hogy a kulturális törvény májusi módosítása a nemzeti könyvtár feladatává tette a webtartalom megőrzését (a részleteket rendelet fogja szabályozni), megteremtve a jogi hátterét munkánknak. Júniusban pedig életbe lépett az OSZK új szervezeti szabályzata, mely önálló osztályt hozott létre a webarchiválás feladatainak ellátására.

Sokféle számadat szerepelt ebben az összefoglalóban… Azonban, ha meg szeretnénk vonni az együttműködés mérlegét, mégsem a Webarchívumban lévő gyűjtemény számszerű gyarapodását emelnénk ki, hanem azt, hogy több nagy, megyei hatókörű városi könyvtár és sok könyvtáros kolléga kapott ízelítőt a webarchiválás feladataiból, nehézségeiből, és láthatóan sokukban alakult ki érdeklődés és elköteleződés az ügy iránt, ami remélhetően majd visszaköszön a napi munka során is.

Köszönjük szépen minden résztvevő könyvtárnak és kollégának az együttműködést, a közös munkát – amit nem elköszönéssel szeretnénk lezárni...!