



Networkshop 2019

2019. április 23-26.
Széchenyi István Egyetem, Győr





INNOVÁCIÓS ÉS TECHNOLÓGIAI
MINISZTERIUM



Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Networkshop

2019. április 23–26. Széchenyi István Egyetem, Győr
konferencia előadásainak közleményei

ISBN 978-615-00-7260-9

DOI: [10.31915/NWS.2019](https://doi.org/10.31915/NWS.2019)

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével
Budapest
2019

TARTALOMJEGYZÉK

Előszó	5
SMART TECHNOLOGIÁK, AI, IOT	
Berki Borbála: Humán kognitív jellemzők szerepe a virtuális terek használata során.....	7
Nacsa János - Hajós Mátyás - Komáromi Gergely - Horváth Gergely - Pataki Balázs: Gömbcsap szerelése ember-robot kollaborációban	13
Paniti Imre – Nacsa János – Abai Kristóf: Robotos szerelési kísérletek háromujjas megfogóval	22
TARTALOMSZOLGÁLTATÁSOK: KÖNYVTÁRAK, LEVÉLTÁRAK, MÚZEUMOK	
Albert Ágota: GDPR és a köznevelési intézmények – a GDPR bevezetésének 13+1 lépése a köznevelési intézményekben	31
Bobák Barbara – Gábori Kovács József: Kézírásfelismerés Arany János levelein	38
Haász Antal: Áttérés a Primo New User Interface használatára a Magyar Tudományos Akadémia Könyvtár és Információs Központban	45
Holl András – Bilicsi Erika: Nyílt publikálási szoftverek és platformok	54
Hubay Miklós – Gyuricza Andrea: Új könyvtári fogalmak gyakorló könyvtárosok szemszögéből	61
Juhász Attila – Szabó Dóra: Parazita folyóiratok, a tudományos világ élősködői	68
Kovács Béla Lóránt: Egészségügyi informatikai adatbázisban való szöveges keresés mesterséges intelligenciával támogatott szemantikus keresővel	73
Kovács László – Micsik András: MTA SZTAKI DSD – 25 éve a digitális könyvtárak szolgálatában	80
Kómár Éva: A múzeumi adatrepresentáció és közzététel formái a Közgyműteményi Digitalizálási Stratégia keretében	87
Lencsés Ákos: A nyílt hozzáférés kérdése a hivatalos statisztikában	96
Lendvay Miklós: Entitások és kapcsolódások a könyvtári rendszerekben és azokon túl	104
Muzs Kriszina – Molnár Tamás - Hoczopán Szabolcs: Open Access pályázati rendszer technikai megvalósítása és a szerzők támogatása a Szegedi Tudományegyetemen	114
Nagy Dóra – Nagy Gyula: Haladó funkciók és innovatív fejlesztések az EPrints és Omeka szoftverek körében	121
Nagy Gyula – Nagy Dóra – Sándor Ákos: Tömeges adatkonverzió és rugalmas export-import lehetőségek az EPrints, OJS és Omeka szoftverek körében	131
Németh Tibor: Magánvállalkozás és közszolgálat: A biográfiai index műfaji megújítása a Magyar Életrajzi Kézikönyvben	141
Sütő Péter: Az elektronikus szakirodalmi információforrások hazai lelőhelyadatbázisa, a COMPASS	150
Szepesi Judit: E-Learning rendszer és elektronikus tananyagfejlesztés a Könyvtári Intézetben	157
Ungváry Rudolf: Besorolási, szabványosított, normatív vagy „autorizált”	163
KÖZNEVELÉS, FELSŐOKTATÁS, ELEKTRONIKUS TANULÁSI KÖRNYEZETEK	
Antal Péter: Digitalizáció és sport: az IKT kihívásai a testnevelő képzésben	172
Csernai Zoltán: Az informatikai gondolkodás (computational thinking) fogalmi keretei	181
Leitold Ferenc: A felhasználói viselkedés, mint információbiztonsági kockázat becslése	189

DR. BÁLINT LAJOS (1946–2019) EMLÉKÉRE

ELŐSZÓ

Tisztelt Olvasó!

A huszonnyolcadik Networkshop konferenciakötetét ajánljuk szíves figyelmébe. Bár a hazai kutatási és felsőoktatási informatikai hálózatot építő, működtető és felhasználó szakmai közösség konferenciái hosszú múltra tekintenek vissza, ez a második alkalom, hogy szerkesztett, lektorált tanulmányokból kötetet jelentettünk meg. Minden tanulmány DOI azonosítót kapott, és bibliográfiai adataikat az MTMT-ben is rögzítettük. A kötet csak digitális formában kerül terjesztésre – megtalálható lesz az MTA Könyvtár és Információs Központ Open Conference Systems rendszerében, és a REAL-ban – de várhatólag más hazai repozitóriumokban is.

2019-es konferencia 24 előadása jelenik meg ebben a kötetben – több, mint a tavalyiban, de kevesebb, mint ahány előadás szerepelt a programban. Köszönjük az Információs és Technológiai Minisztérium támogatását, amellyel a kiadást segítette.

Budapest, 2019

A szerkesztők

Smart technológiák, AI, IoT

Humán kognitív jellemzők szerepe a virtuális terek használata során

Berki Borbála

Multidiszciplináris Műszaki Tudományi Doktori Iskola

Széchenyi István Egyetem

berki.borbala@sze.hu

Role of human cognitive characteristics in the use of virtual spaces Virtual realities are used by human beings; thus, their cognitive characteristics are closely related to the user experience. This paper shows the contributing human characteristics and the results of a connecting research about human spatial abilities and the use of virtual realities. For measuring the spatial abilities, the Corsi-test and the mental rotation task were used. Based on the data of 31 people, no significant relationship was found between the measured spatial abilities and the performance on the VR task or the navigational experience. Thus, individual spatial memory and mental rotation skills do not affect the use of this type of desktop VR.

Keywords: desktop VR, human cognition, MaxWhere, virtual reality

Bevezetés

A virtuális valóságokra gyakran egy technológiai termékként gondolunk, amelynél a felhasználói élményt a különböző újítások és fejlesztések határozzák meg. Azonban mivel ezeket az eszközöket emberek használják, ezért elengedhetetlen a humán jellemzők figyelembevétele. Jelen esetben itt nem klasszikus ergonómiai szempontokat értem, mint például, hogy fizikailag kényelmesen kezelhető legyen az eszköz, és nem is egy tipikus szoftverergonómiai szempontot, hanem hogy az emberi gondolkodási folyamatokhoz hogyan kapcsolódik egy virtuális tér. Ehhez egy mérést is bemutatok, amely a MaxWhere virtuális valóságban zajlott. Ez egy ún. asztali virtuális valóság, amely a 3D térben elhelyezett okostáblákon jelenít meg weblapokat, dokumentumokat vagy egyéb fájlokat.

1. Humán kognitív jellemzők és VR

A kognitív kifejezés a latin *cognosco* igéből származik, jelentése megért, felfog. A megismerési és gondolkodási tevékenységeket jelenti. Ide tartozik az észlelés, figyelem, végrehajtó funkciók, emlékezeti folyamatok, illetve egyéb magasabb rendű gondolkodási feladatok, mint a tervezés, döntéshozás, problémamegoldás. A virtuális terek alkalmazása során a különböző kognitív funkciók segítenek megérteni a felhasználónak, hogy mi is az a szimuláció, amit látnak, és hogy hogyan tudják használni a teret. Virtuális valóság megismerésekor az alábbi kérdések merülhetnek fel a felhasználóban: elsőként talán az, hogy: mi az, amit látok? Hol vagyok a virtuális valóságon belül, illetve merre tartok, hová megyek? Majd ezután a többi lehetséges cselekvés is sorra kerül. Ehhez elengedhetetlen a pontos észlelés és a figyelem. A virtuális térben való navigálás a perceptuális-



motoros koordináció megfelelő működésén alapszik. Emellett a munkamemória, a már meglévő tudásbázis is szerepet játszik a virtuális valóság.¹

Virtuális valóságok esetében leggyakrabban a téri képességek vizsgálata kerül előtérbe a kognitív jellemzők közül, hiszen a virtuális valóság térszerűségét a kognitív funkcióknak köszönhetően tudja megtapasztalni a felhasználó. A téri képességek a kognitív funkciók és adottságok egy csoportja, amelyek meghatározók a téri-vizuális információk feldolgozásában. Illetve az olyan problémák megoldása során, ahol téri jellegű információk manipulációjára van szükség.

Avirtuális valóságok számos tekintetben különböznek: eltérők lehetnek a navigációs lehetőségek és eszközök, a tér elrendezései, tárgyai, egyszóval komplexsége is különféle lehet; valamint a tér célja, felhasználási módja is sokféle lehet. Ezeket az eltéréseket figyelembe véve, talán nem is olyan meglepő, hogy eltérők a kutatási eredmények a tekintetben, hogy az erősebb vagy a gyengébb téri képességekkel rendelkező felhasználók profitálnak többet a virtuális terek alkalmazásával. Egyes eredmények szerint a téri képességeknek van egy minimális szintje, amely szükséges a résztvevők jó teljesítményéhez a kísérleti feladatban, de ezen szint fölött az eltérő téri képességek nem magyarázzák az eltérő teljesítményeket.²

Más kutatások szerint anatómiai modellek tanulásakor a többféle irányból való bemutatás nem segítette, hanem tovább nehezítette a gyengébb téri képességű résztvevők tanulását.³ Más kutatási eredmények azonban ennek az ellenkezőjét állítják, vagyis, hogy a rosszabb téri adottságokkal rendelkezők jutottak előnyhöz, ha a virtuális teret használták a tanulás során.^{4,5}

2. Asztali virtuális valóság

A virtuális valóság (VR) nemcsak piaci, hanem tudományos szempontból is egyre nagyobb figyelmet kap az utóbbi években. Legtöbb embernek a virtuális valóság szó hallatán a népszerű VR szemüveges megoldások jutnak eszébe, azonban a virtuális valóság megtapasztalásához egy laptop is elegendő lehet. Ebben az esetben kétdimenziós kijelzőn jelenik meg a háromdimenziós virtuális tér. Az ilyen típusú virtuális valóságokat *asztali VR*-nak (angolul: desktop-VR), vagy nem belemerülő (angolul: non-immersive) virtuális valóságnak nevezik, mivel teljes 360° helyett, csak a képernyő keretein belül jelenik meg a szimuláció.

- 1 Wickens, C. D., Baker P.: Cognitive Issues in Virtual Reality. In Barfield W., Furness T. A. (szerk.), *Virtual environments and advanced interface design*. New York, NY, US: Oxford University Press. 1995, 514-541.
- 2 Modjeska, D., Chignell, M.: Individual differences in exploration using desktop VR, in *Journal of the American Society for Information Science and Technology*, 2003/54/3. szám, 216-228.
- 3 Levinson, A. J., Weaver, B., Garside, S., McGinn, H., Norman, G. R.: Virtual reality and brain anatomy: a randomised trial of elearning instructional designs, in *Medical Education*, 41(5), 495-501.
- 4 Lee, E. A., Wong, K. W.: Learning with Desktop Virtual Reality: Low Spatial Ability Learners Are More Positively Affected, in *Computers & Education*, 2014/79. szám, 49-58.
- 5 Jang, S., Vitale, J. M., Jyung, R. W., Black, J. B.: Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment, in *Computers & Education*, 2017/106. szám, 150-165.

Az asztali virtuális valóságok előnye az alacsonyabb eszközigeny és hogy nem okoznak szimulációs rosszullétet. Könnyebben kezelhetők, mivel a navigációhoz már ismert eszközöket (egér, billentyűzet) kell használni. Emiatt az oktatás területén is népszerűek ezek a virtuális valóságok, hiszen egy új és látványos felületet biztosít a tudás átadására. Magyarországon például a Széchenyi István Egyetem is alkalmazza már az oktatás során a MaxWhere⁶ asztali virtuális valóságot. Egy-egy oktatási térben a klasszikus szövegből és képekből álló tananyag mellett, helyet kapnak a kapcsolódó videó- és hanganyagok, gyakorló feladatok, vagy akár teljes 3D-modellek is. Továbbá virtuális laboratóriumok is vannak egy-egy ilyen térben, amelyet a hallgatók bármikor elérhetnek, akár otthonról is.⁷

Tudományos vizsgálatokban is alkalmazták már a MaxWhere virtuális valóságot. Ez alapján hatékonyabban oszthatóak meg segítségével a digitális tananyagok, mint a klasszikus e-learning rendszerekkel (e-mail, Moodle), illetve összetettebb munkafolyamatokat gyorsabban hajtanak végre a felhasználók.⁸ A gyorsaság háttérében az állhat, hogy azonos digitális munkafolyamat végrehajtásához kevesebb felhasználói beavatkozás, illetve gépi művelet szükséges a virtuális térben, mint klasszikus kétdimenziós interfészekben.⁹ Az oktatási területen további előny, hogy a közvetlenül 3D térbe beillesztett képekre több felhasználó emlékezett, mint a beillesztett weblapokon megjelenő azonos képekre.¹⁰

3. Térei képességek és virtuális valóság kapcsolatának mérése¹¹

3.1. Virtuális tér és a kapcsolódó feladatok

A teljes kísérlet a MaxWhere virtuális valóságban futott, amely képes megjeleníteni weblapokat, dokumentumokat, képeket, illetve hangfájlokat a háromdimenziós virtuális térben elhelyezett, úgynevezett okostáblán. Az okostábla egy 4:3 arányú kijelzőnek felel meg, amelyre a korábban felsorolt tartalmak behívhatóak.

A vizsgálat a Glassy elnevezésű térben zajlott, amelyet prezentációs, illetve online kiállítási térnek terveztek. Egy modern, üveg és látszóbeton felületekből álló háromszintes, nyitott épület oldalfalain van elhelyezve 40 okostábla (1. ábra). A résztvevők a legfelső szintről indulnak, ahol olvashatók az instrukciók, valamint itt található meg az a weblap, amely végigvezeti őket a feladaton, és ahová a válaszaikat kell beírniuk. A többi okostáblán fényképeket helyeztünk el (forrás: unsplash.com),

6 MaxWhere, hozzáférés: 2019.06.25. <https://www.maxwhere.com/>

7 Budai T., Kuczmann M.: Towards a Modern, Integrated Virtual Laboratory System, in *Acta Polytechnica Hungarica*, 2018/15/3. szám. 191-204.

8 Lampert B., Pongrácz A., Sipos J., Vehrer A., Horváth I.: MaxWhere VR-Learning Improves Effectiveness over Classical Tools of E-learning, in *Acta Polytechnica Hungarica*, 2018/15/3. szám, 125-147.

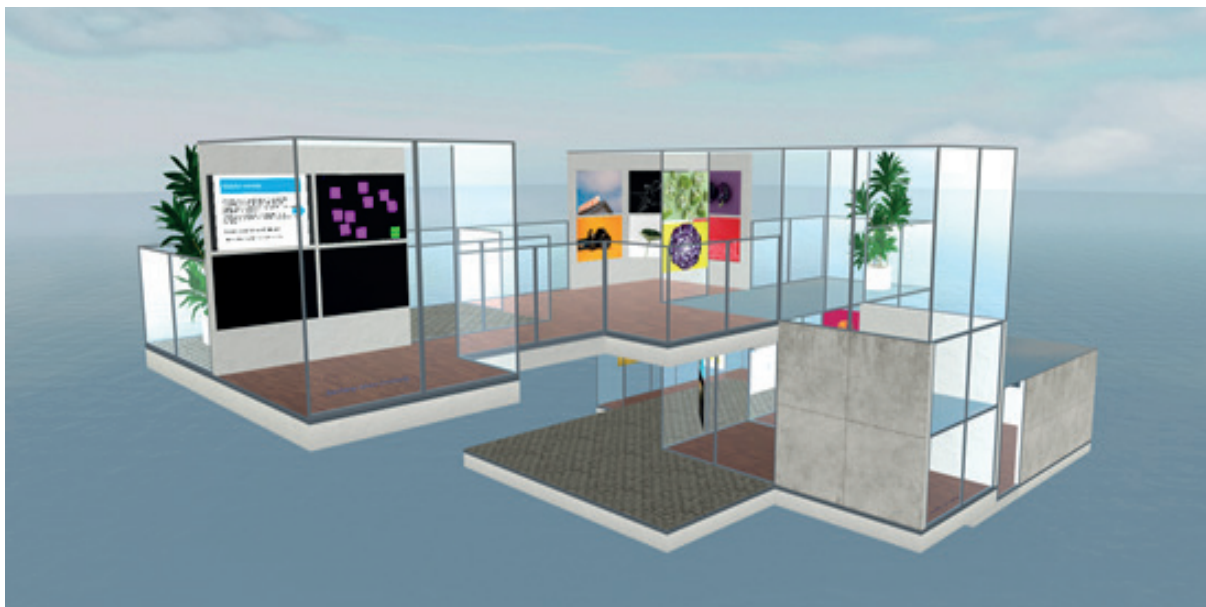
9 Horváth I., Sudár A.: Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information, in *Acta Polytechnica Hungarica*, 2018/15/3. szám, 149-173.

10 Berki B.: Better Memory Performance for Images in MaxWhere 3D VR Space than in Website, in *Proceedings of the 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018)*, 281-284.

11 Berki B.: Kognitív jellemzők a 3D virtuális terek hatékonyságának háttérében, in *Új Nemzeti Kiválóság Program 2018/2019 Tanulmánykötet*, 2019, 107-115.



amelyeken különböző tárgyak láthatóak semleges háttér előtt, néhány esetben feliratok vagy számok. Emellett két okostáblán egy-egy pdf dokumentum volt. A feladat megoldásához a térben kellett mozogniuk, és a digitális tartalmakkal, azok elrendezésével kapcsolatos kérdésekre válaszolniuk. A feladatokkal összesen 12 pontot szerezhettek. A térben való hatékony feladatvégzést a megszerzett pontok száma, valamint az ehhez szükséges idő jelzi. A válaszokat a legfelső szinten elhelyezett kérdőívben kellett rögzíteniük, ahová a TAB billentyű megnyomásával visszaugorhattak.



1. ábra: A kutatásban alkalmazott virtuális tér

3.2. Térei képességmérő eljárások a kutatásban

A Corsi-teszt¹² a téri információk rövid távú kódolását és megtartását mérő eljárás, amelynek a számítógépes verzióját alkalmaztuk. A képernyőn kilenc négyzet látható, amelyek közül néhány egymás után eltérő színben villan fel. A résztvevő feladata, hogy a felvillanás sorrendjében rákattintson a megfelelő négyzetekre. A felvillanó négyzetek száma egyre növekszik a feladat során, amennyiben a kitöltő helyesen válaszolt. Hibázás esetén újra azonos hosszúságú szekvenciát kell megjegyeznie, ha a második sem sikerül, akkor vége a feladatnak, és az utolsó sikeresen megjegyzett szekvencia hossza adja meg a Corsi-terjedelmet.

A mentális forgatás feladat¹³ során egy formának az elforgatott verzióját kell felismerni, két lehetséges ábra közül. A jelen kutatásban egy számítógépes verziót használtam, kétdimenziós ingerekkel. Először öt gyakorlófeladatot oldottak meg a résztvevők, majd a teszt során tíz ábrát láttak.

¹² Corsi, P. M.: *Human memory and the medial temporal region of the brain*, 1972, Doktori értekezés: McGill University (Kanada).

¹³ Shepard, R. N., Metzler, J.: Mental Rotation of Three-Dimensional Objects, in *Science*, 1971/171/3972. szám, 701-703.

3.3. Navigációs élmény mérése

A résztvevőknek értékelniük kellett a virtuális térben való navigációt, azt, hogy mennyire mozgott magabiztosan a térben, mennyire érezte automatikusnak, természetesnek az irányítást. Ehhez tízes skálán kellett kifejezni az egyetértésüket öt kapcsolódó állítással.

3.4. Résztvevők és mérés menete

A vizsgálatban 31 fő vett részt (átlagos életkor 20,5 év; szórás: 3,4). A résztvevők között 25 férfi és 6 nő volt.

A vizsgálat vezetője először röviden szóban ismertette az instrukciókat, amelyeket a résztvevők a virtuális térben is láthattak. A feladatok a virtuális térben egy weboldalon jelentek meg és ott is kellett megadni a válaszokat. Amennyiben a résztvevők kattintással jelezték, hogy részt kívánnak venni a mérésben, akkor elsőként a Corsi feladatot oldották meg, majd a mentális forgatást. Ezután már nem volt szükséges, hogy az okostábla teljes méretben jelenjen meg. Ezután következtek a virtuális térben végzendő feladatok, majd a navigáció szubjektív értékelése és egyéb általános kérdések következtek.

3.5. Mérés eredményei

A téri képességek (Corsi-teszt, mentális forgatás) és a hatékonyság (megszerzett pontok) kapcsolatát Spearman-féle rangkorrelációval vizsgáltam. A hipotézisemnek megfelelően a megszerzett pontok nem mutattak szoros együttjárást sem a Corsi-terjedelemmel ($r_s(28)=0.09$, $p=0.635$), sem a mentális forgatás eredményeivel ($r_s(29)=-0.174$, $p=0.351$). Vagyis nem volt szignifikáns kapcsolat a különböző téri képességek és a 3D virtuális térben végzett feladatok között. A navigációs teljesítmény és a hatékonyság kapcsolatát a Spearman-féle rangkorrelációval vizsgáltam. Azonban hipotézisemmel ellentétben, nem volt szignifikáns kapcsolat a két változó között ($r_s(29)=0.068$, $p=0.718$).

4. Összefoglalás

A fentebb bemutatott kutatás a virtuális térben való feladatvégzés és a kognitív jellemzők közül a téri képességek kapcsolatának feltárása volt. A vizsgált két téri képességmérő teszt (Corsi-feladat, mentális forgatás) eredményei és a kísérleti feladaton elért pontszám teljesítménye között nem volt statisztikailag szignifikáns kapcsolat, amelyből arra lehet következtetni, hogy az egyéni téri képességek nem befolyásolják a virtuális térben való feladatvégzést. Ezen eredmények alapján a MaxWhere VR használata előnyös, mivel a térben való hatékony feladatvégzéshez nem szükségesek kiemelkedő téri képességek.

A humán kognitív jellemzők közül tehát a téri képességeknek nincsen kiemelt szerepe a MaxWhere virtuális valóság használata során.

Köszönetnyilvánítás

A kutatás az Emberi Erőforrások Minisztériuma ÚNKP-18-3 kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült.



Bibliográfia

Berki B.: Better Memory Performance for Images in MaxWhere 3D VR Space than in Website, in Proceedings of the 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018), 281-284.

Berki B.: Kognitív jellemzők a 3D virtuális terek hatékonyságának hátterében, in *Új Nemzeti Kiválóság Program 2018/2019 Tanulmánykötet*, 2019, 107-115.

Budai T., Kuczmann M.: Towards a Modern, Integrated Virtual Laboratory System, in *Acta Polytechnica Hungarica*, 2018/15/3. szám. 191-204.

Corsi, P. M.: *Human memory and the medial temporal region of the brain*, 1972, Doktori értekezés: McGill University (Kanada).

Horváth I., Sudár A.: Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information, in *Acta Polytechnica Hungarica*, 2018/15/3. szám, 149-173.

Lampert B., Pongrácz A., Sipos J., Vehrer A., Horváth I.: MaxWhere VR-Learning Improves Effectiveness over Classical Tools of E-learning, in *Acta Polytechnica Hungarica*, 2018/15/3. szám, 125-147.

MaxWhere, hozzáférés: 2019.06.25. <https://www.maxwhere.com/>

Jang, S., Vitale, J. M., Jyung, R. W., Black, J. B.: [Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment](#), in *Computers & Education*, 2017/106. szám, 150-165.

Lee, E. A., Wong, K. W.: Learning with Desktop Virtual Reality: Low Spatial Ability Learners Are More Positively Affected, in *Computers & Education*, 2014/79. szám, 49-58.

Levinson, A. J., Weaver, B., Garside, S., McGinn, H., Norman, G. R.: Virtual reality and brain anatomy: a randomised trial of elearning instructional designs, in *Medical Education*, 41(5), 495-501.

Modjeska, D., Chignell, M.: Individual differences in exploration using desktop VR, in *Journal of the American Society for Information Science and Technology*, 2003/54/3. szám, 216-228.

Shepard, R. N., Metzler, J.: Mental Rotation of Three-Dimensional Objects, in *Science*, 1971/171/3972. szám, 701-703.

Wickens, C. D., Baker P.: Cognitive Issues in Virtual Reality. In Barfield W., Furness T. A. (szerk.), *Virtual environments and advanced interface design*. New York, NY, US: Oxford University Press. 1995, 514-541.

Gömbcsap szerelése ember-robot kollaborációban

Nacsa János ^{1,2}

^{1:} *Mérnöki és Üzleti Intelligencia Kutatólaboratórium, SZTAKI*

nacsa.janos@sztaki.hu

^{2:} *Járműipari Kutatóközpont, Széchenyi István Egyetem*

Hajós Mátyás, Komáromi Gergely, Horváth Gergely

Mérnöki és Üzleti Intelligencia Kutatólaboratórium, SZTAKI

[hajos.matyas](mailto:hajos.matyas@sztaki.hu) | [komaromi.gergely](mailto:komaromi.gergely@sztaki.hu) | horvath.gergely@sztaki.hu

Pataki Balázs

Elosztott Rendszerek Osztály, SZTAKI

pataki.balazs@sztaki.hu

Assembly of ball valve in human–robot collaboration: A ball valve assembly task, as an example, was analysed and solved with a collaborative way, when a cobot and a human operator can work together. A human assistance system (HMIC) was introduced to lead and help the operator to process the different assembly steps.

Bevezetés

Az Ipar 4.0 koncepció megjelenése radikálisan megváltoztatta az egész társadalomban az ipari termelésről alkotott képet. A lehetőségek és ígéretek mögötti valóban elérhető valóság kutatására és a hazai fejlesztések támogatására jött létre a Stratégiai K+F műhelyek kiválósága pályázat keretében az Ipar 4.0 kutatási és innovációs kiválósági központ, melynek megvalósulási helyszíne a SZTAKI győri telephelye, amely a Széchenyi Egyetem területén működik. A kiválósági központban a megcélzott eredmények három pontban foglalhatók össze:

- Az I4.0 elveinek, a kiber-fizikai gyártórendszer paradigma háttér-technológiáinak kutatása.
- Egyetemi és ipari oktatás, tesztkörnyezet létrehozása ötletek kipróbálására, problémák feltárására.
- I4.0 szakértői támogatást igénylő ipari megbízások vállalása és teljesítése.

A háromféle és egymást kiegészítve erősítő eredményt az alábbi kutatási területekre összpontosítva reméljük elérni:

- Szituáció-tudatos, erőforrás-hatékony és robusztus termelés-tervezés és irányítás.
- Kooperatív és adaptív termelési és logisztikai hálózatok tervezése és működtetése.
- Robusztus kooperatív kontrol kiber-fizikai rendszerekben.
- I4.0 megoldások az energia-hatékony és fenntartható gyártás támogatására.
- Számítási felhő alapú termelésinformatikai szolgáltatások.
- I4.0 megoldásokat használó, demonstrációs mini gyártó- és logisztikai rendszer létrehozása.



Jelen dolgozat az utolsó pontként említett mintarendszer egyik első eredményét ismerteti. Pontosabban a mintarendszer három fókuszából (kiber-fizikai gyártó rendszer, digitális iker és kollaboratív robotos szerelés) a legutolsóban született megoldást mutatja be: egy általánosan használt alkatrész, az egy colos gömbcsap szerelésének kollaboratív megoldását.

1. Kollaboratív robotos szerelés

A robotos szerelés és általában az ipari robotok használata olyan környezetben valósul meg a mai napig, ahol a gyorsan mozgó robot egy körültekintően lehatárolt térben dolgozik, ahova ember nem léphet be, amikor a robot mozog. Hagyományosan mechanikusan lezárt tereket képeznek (pl. fémketrecek) vagy speciális mozgásérzékelőkkel (pl. lézervégzőkkel) biztosítják, hogy a mozgó robot és az ember ne kerüljön közel egymáshoz.

Ugyanakkor egyre több olyan feladat kerül megfogalmazásra, ahol a robot és az ember közös munkája sok előnnyel kecsegtet. Pl. egy nehezebb alkatrész pontos elhelyezése, majd rögzítése során jó, ha egy robot tartja az alkatrészt, az ember pedig finom pici mozgásokkal tudja a helyére pozicionálni. Általában fogalmazódik meg az igény az emberi rugalmasság és problémamegoldó képesség kombinálására a robot precizitásával, kitartásával és akár erejével.

A lehetőség megteremtése azonban rögtön sokféle problémát felvet, amit röviden úgy lehetne jellemezni, hogyan lehet biztonságos, félelmek nélküli munkakörnyezetet megvalósítani. Részletezve ez a robottal dolgozó ember szempontjából azt jelenti, hogy:

- a robot olyan pályán mozog, amire az ember számít: „tudom, hogy mit fog csinálni”
- a robot felismeri az ember tevékenységét: „tudom, hogy érzékel engem”

Ez a két egyszerűnek tűnő elvárás megvalósítandó műszaki feladattá átfogalmazva komoly kihívásokat jelent, melyek közül a legfontosabbak az alábbiak:

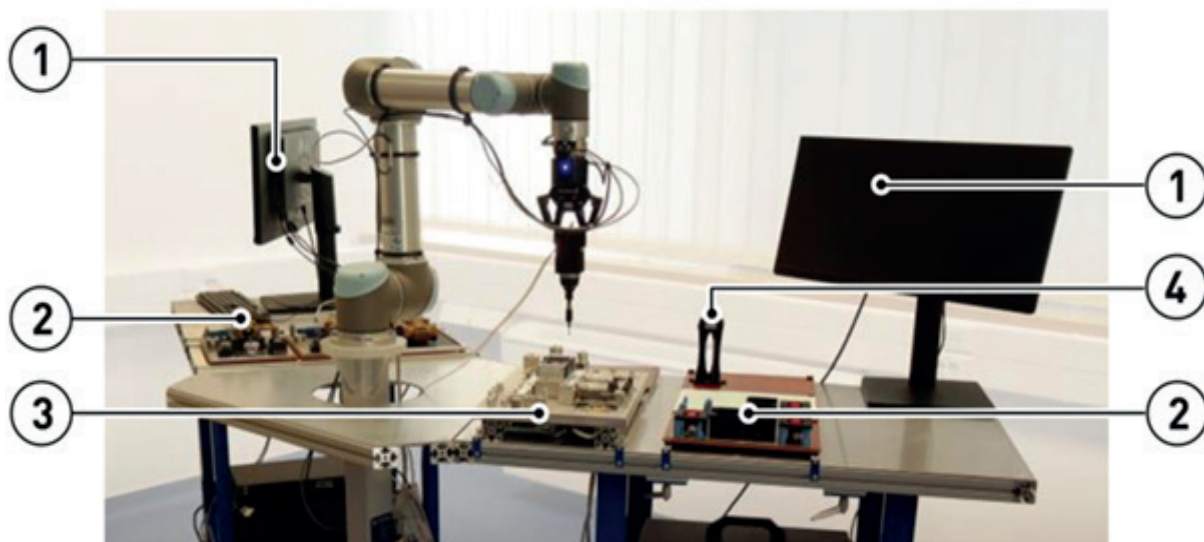
- intuitív ember-robot interfész,
- érzékelés dinamikusan változó környezetben,
- adaptivitás a robot vezérlésében,
- interoperabilitás,
- elegendően pontos modellezés és szimuláció.

Érdemes megemlíteni, hogy ezek az elvárások nagyon sokféle szabványban (EN ISO 12100, EN ISO 13949-1/2, EN 60204-1, EN 62061, EN ISO 10218, ISO/PDTS 15066, ISO 10075-1, ISO 26000 stb.) is megfogalmazódnak, melyeket meg lehetne nehéz maradéktalanul teljesíteni [1],[2].

2. A kollaboratív szerelési környezet

A szerelési környezet (1. ábra) központi eleme a robot, mely egy mobil állványon áll. Az állvány súlya olyan, hogy biztosítja, hogy a robot bázisa ne mozdulhasson el. A robot ill. az állvány köré több szerelő asztal helyezhető el, amely biztosítja hogy (1) egyidőben több helyen levő palettáról lehessen szerelni vagy (2) többféle szerelési feladathoz szükséges munkakörnyezetet lehessen kialakítani. Az asztalok

lábain olyan kerekerek vannak, amelyeket fel lehet emelni és akkor az asztal fixen megáll, ráadásul az asztalok egymáshoz és a robotállványhoz oldható kötéssel vannak összeépítve. A kialakítás előnye, hogy egyszerűen átépíthető, egy robot környezete könnyedén megváltoztatható, de kb. egy nap alatt a teljes I4.0 labor munkahelyeinek az elhelyezése átkonfigurálható.



1. ábra A gömbcsap szerelés robotos környezete

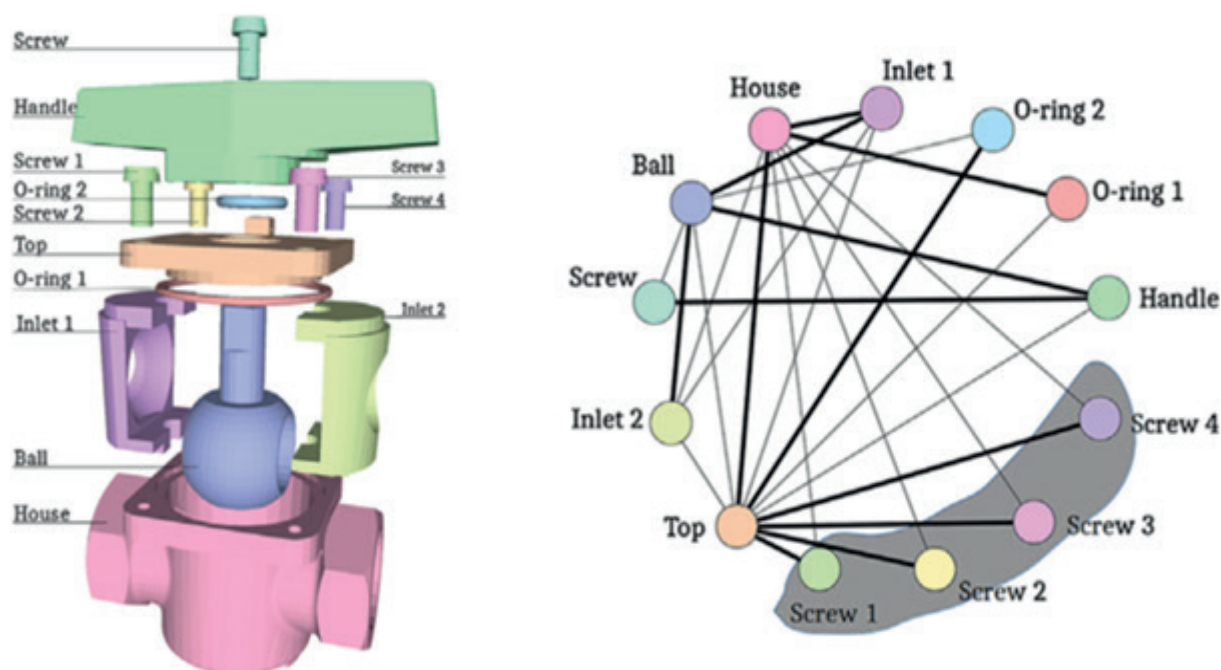
Az asztalok alatti tálcákra helyezkednek el a szerelési környezet IT erőforrásai a robotvezérlő és egy számítógép, amely biztosítja az ember és a robot valamint a teljes rendszer és a laborszintű vezérlés MESS [4] közötti kapcsolatot. Az emberrel sokféle módon képes a rendszer kommunikálni (4. fejezet), de különösen a fejlesztéskor a hagyományos monitorok (1) az elsődlegesek. A szerelendő munkadarabok ún. palettákra (2) találhatók, melyeken az alkatrészek strukturáltan, szerelésre előkészítve találhatók. A palettákat mobilrobot (AGV) vagy operátor szállítja a munkahelyre és lyuk-csap illesztéssel kerülnek pontos pozícióba vagy a robot vagy az ember közreműködésével. A szerelési készüléket (3) kézzel kell átállítani (megfelelő ülékek ki/be szerelésével, pozicionálásával), hogy alkalmas legyen pl. az egy colos gömbcsap szerelésére. A legtöbb szerelési lépés a robot megfogójával végrehajtható, kivéve a csavarozás, amikor egy csavarozó szerszámot kell a robotnak megfognia. A speciálisan tervezett szerszám egy rögzített állványon (4) található, innen veszi fel a robot, amikor a csavarozás következik.

3. Az egy colos gömbcsap szerelési modellje

A gömbcsap, amely alapvetően egy egyszerű tárgy, összesen 13 alkatrészből áll, melyek sok szempontból különböznek: műanyag - fém, alaktartó - alakváltozó; speciális elem - tömeg alkatrész stb. A 2. ábra bal felén a gömbcsap egy ún. „robbantott alkatrészsrajza látható”, míg a jobb oldalon a gráf mutatja az egyes alkatrészek kapcsolódását (vastag vonalak). Látható az ábrán, hogy a négy csavar egymással felcserélhető, hogy melyiket melyik menetbe csavarjuk be a fedélre. Elméleti módszerek is léteznek [5] az összeállított CAD modellből közel

automatikusan generálni a szerelési sorrendtervet, a gömbcsap esetében ez a módszer ellenőrzésre szolgált és ugyanazt eredményezte, mint a szerzők által kialakított sorrend. Fontos megjegyezni, hogy a robotos szerelés nem lett volna lehetséges, ha a robot megfogójának a pofájába nem került volna egy speciális profilú elem, melynek segítségével a robot megbízhatóan fel tudta venni a kisebbik O gyűrűt. Az eredeti – univerzálisabb – elem ezzel nem boldogult.

A szerelés lépéseinek a programozásánál több olyan mikro operációt is be kellett építeni a robot programba, amelyet az elméleti modellből nem lehetséges meghatározni. Pl. a fröccsöntött műanyag tömítések méretének szórása miatt a fedélbe helyezett tömítésekkel körülvelt gömböt kétféleképpen is „meg kellett nyomni”, hogy biztosan belekerüljön a tömítés pereme a fedélen erre a célra kialakított vátatba. A kézi szerelések is azt mutatták, hogy ugyanezzel a problémával az ember is gyakran találkozott.



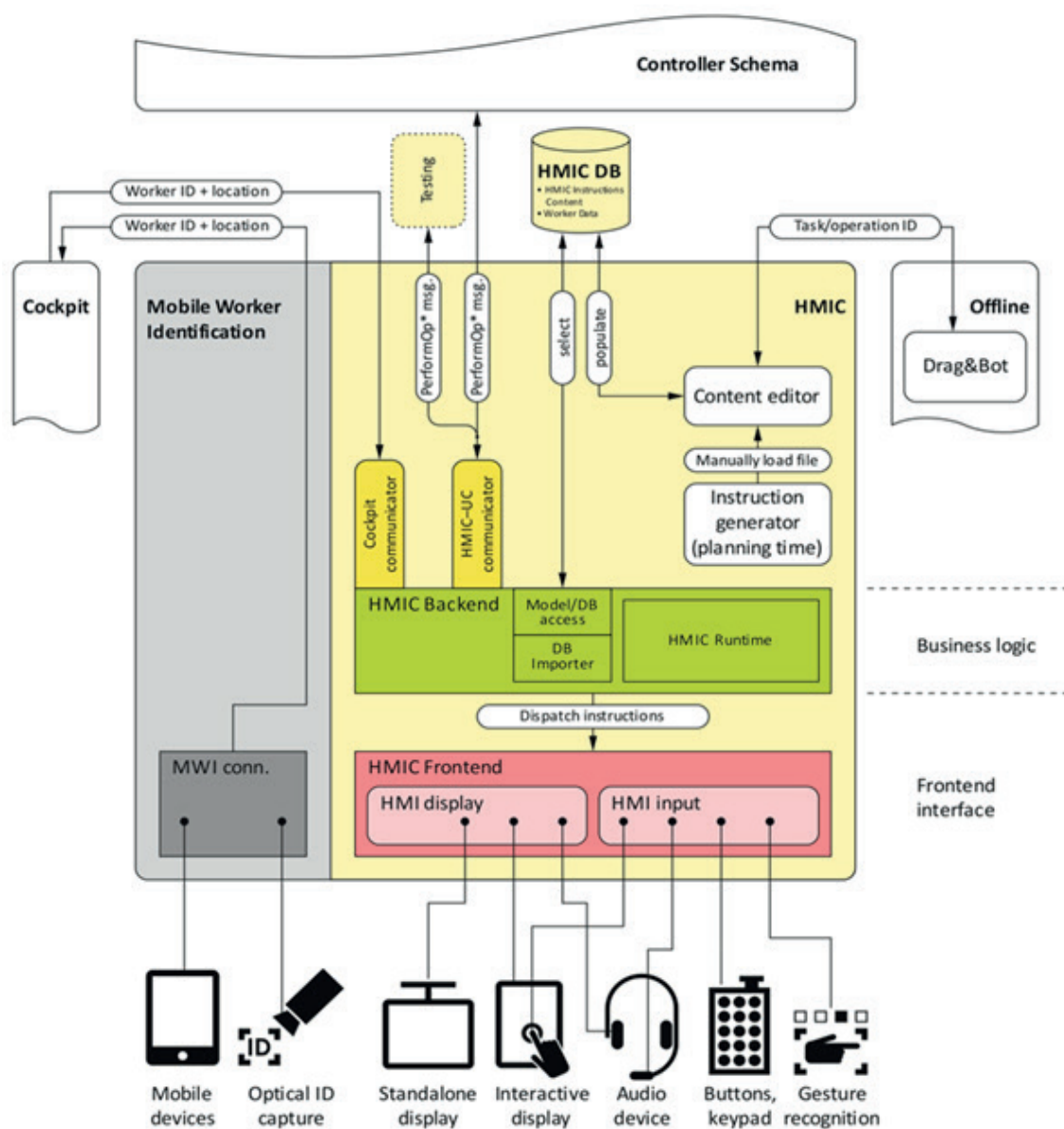
2. ábra A gömbcsap alkatrészei és szerelési kapcsolatok

4. Multi-modális kommunikáció a robot és az ember között

Ipari környezetben valamilyen IT eszköz (pl. számítógép) vagy gép (pl. robot) és az ember közötti kommunikációt hagyományosan monitor és billentyűzet/egér eszközökkel valósítják meg. A multimodális kommunikáció arra szolgál, hogy kiszélesítse a lehetséges kommunikációs csatornákat, hogy mindig azon a csatornán és abban a formában kommunikáljon az ember és a gép, amely az adott feladathoz és környezethez, sőt az adott emberhez a leginkább megfelelő. Rögtön a vizuális megjelenítés formája lehet többféle – például szöveg, kép, videó, virtuális valóság, kiterjesztett valóság stb. – de más módon is közölhetünk információkat az emberrel (beszédszerű hang, speciális fényjelzés, hangeffektusok, rezgés a telefonon stb.) Még nagyobb a skála, hogy az ember hogyan küldhet üzeneteket: a billentyű,

egér, a napjainkban igen elterjedt érintőképernyő mellett szóban elmondott parancsokkal, különféle nyomógombokkal (pl. vészstop), de akár testhelyzetével, kar vagy kézmozdulataival.

Ugyanilyen fontos a kommunikáció tartalma is, ami sokkal többet jelent, mint pl. az üzenet nyelve. Más információra van szüksége egy gyakorlott szakembernek, mint egy kezdőnek vagy egy frissen a feladathoz állított munkásnak. Egy fiatalabb ember általában jobban érti a vizuális jelzéseket, míg az idősebbek inkább a szöveges magyarázatokhoz vannak hozzászokva. Van, aki igényli, hogy sok ellenőrzési pont legyen, mert ez megnyugtatja, míg másokat ezek inkább idegesítenek. Olyan eszközre van szükség, amelyben ez a sokféle kommunikációs csatorna és igény egyszerűen konfigurálva megvalósítható.



3. ábra Sematikus ábra a HMIC architektúrájáról



5. HMIC

Szerelés közben a felhasználót utasításokkal kívánjuk ellátni, de nem szeretnénk korlátozni magunkat abban, hogy erre milyen eszközt használjunk. Az előző fejezetben bemutatott eszközök egységes kezelése érdekében kidolgozásra került a Human Machine Interface Controller (HMIC) [3]. Elvárás, hogy a kommunikáció kétirányú legyen, tehát felhasználói beavatkozásra is legyen lehetőség. Mivel nincs szükség arra, hogy a felhasználó által használt eszköz komplex logikát implementáljon, ezért logikus döntésnek bizonyult a szerver-kliens megvalósítás és a népszerű és széles körben elterjedt sztenderdek használata (például HTML5). Ez lehetőséget biztosít arra, hogy a megjelenítő eszközök flexibilisen változtathatók legyenek, illetve tetszőleges új eszközt használhassunk. A rendszer eredetileg egy komplex rendszer részeként készült, így fel lett készítve külső komponensekkel való kommunikációra, szabványosított interfészen keresztül. Ennek részleteit a 3. ábra szemlélteti.

6. Kollaboratív gömbcsap szerelés állapotgépi leírása

A szerelés sok apró lépése öt komplex műveletbe lett összevonva. Ezek a lépések, bár különböző komplexitásúak, a tapasztalatok szerint minden operátor számára érthető feladatokat jelentenek:

- A gömb és a tömitések összeállítása.
- A szükséges két O gyűrű felhelyezése a fedélre.
- A ház összeállítása, amikor először a fedél közepén levő lyukba kell helyezni az első lépésben összeállított gömböt a tömitésekkel együtt, majd az egészre rányomni a házat.
- A fogantyú ráhelyezése a házra.
- A fedél és a ház összezsavarozása, valamint a fogantyú csavarozása.

Az első két művelet a strukturált palettán történik, a harmadik a szerelési készülék egy külön területén, melyet megfelelő ülékek szolgálnak. Illetve a művelet végén a házat megfordítva a készülék satujába kerül a munkadarab. Itt kerül rá a negyedik lépésben a fogantyú illetve történik a csavarozás.

Mindegyik szerelési műveletet a kollaboratív környezetben az operátor is és a robot is végre tudja hajtani, ugyanakkor a műveletek közötti ellenőrzést azonban csak az ember, amely feladatok elsősorban a robot műveletei miatt szükségesek, de fel kell készülni képzetlen és így gyakorlatlan operátor támogatására is. Ezekből a megfontolásokból alakult ki az alábbi állapotgépi leírás, ahol soronként kell végrehajtani a műveleteket:

Feladat	Emberi művelet	Robot művelet
Kiindulási feltételek ellenőrzése	01_C	
A gömb és a tömitések összeállítása	02_M	02_R
A gömb tömités ellenőrzése	02_C	
O gyűrűk fedélbe helyezése	03_M	03_R
Fedélben levő O gyűrűk ellenőrzése	03_C	
Ház összeállítása	04_M	04_R
Az összeállított ház ellenőrzése	04_C	
Fogantyú felhelyezése	05_M	05_R
Az fogantyús ház ellenőrzése	05_C	
Csavarozás	06_M	06_R
A kész gömbcsap ellenőrzése	06_C	
Hiba	00_X	

Az állapotleírásban az utolsó sor kivételt jelent, hiszen ide akkor ugrik a vezérlés, ha valahol hiba történt. Ekkor a félig kész alkatrészt eltávolítva a szerelési készülékből és palettát cserélve egy új munkadarab szerelése kezdhető meg.

7. Az operátor támogatása a HMIC környezetben

Az állapotleírás nyomán készülhetett el az operátor munkáját segítő rendszer, amely – valós ipari tesztek (Volvo, Aciturri) tapasztalatai [6] alapján a szöveges információ mellett videót vagy képet tartalmaz. Az ellenőrzési pontokon az operátor két képet lát a szöveg mellett: az alkatrészek palettáját és a szerelési készüléket, mindkettőt abban az állapotban, ahogy az adott ponton ki kell nézniük. Az emberi szerelés választása esetén a feladat ismertetése mellett egy többször is megtekinthető videó mutatja a tennivalót. A robotos szerelés választásakor látható, hogy a robot hogyan végzi el a szerelési lépést (egy korábbi felvétel), majd a szerelés végeztével a rendszer tovább lép a következő – ellenőrzési – lépésre (4. ábra).



4. ábra A 02_R művelet a HMIC-ben

Köszönetnyilvánítás

A bemutatott kutatás „Ipar 4.0 kutatási és innovációs kiválósági központ” projekt (GINOP-2.3.2-15-2016-00002) keretében valósult meg, és az EU Horizon 2020-as EPIC projektje (No. 739592) is támogatta, mivel bekerült a projekt esettanulmányai közé.

Összegzés

Kollaboratív szerelési munkakörnyezetben, amely az Ipar 4.0 kutatások egyik fókuszja, egy kísérleti rendszer került ismertetésre, ahol az egy colos gömbcsapot robot és/vagy ember szereli össze 5 lépésben. Röviden bemutatásra került az együttműködést biztosító HMIC rendszer is.

Hivatkozások

- [1] Zsolt Kemény, Richárd Beregi, János Nacsa, Csaba Kardos, Dániel Horváth: Example of a problem-to-course life cycle in layout and process planning at the MTA SZTAKI learning factories; 9th Conference on Learning Factories 2019
- [2] L. Wang, R. Gao, J. Váncza, J. Krüger, X.V. Wang, S. Makris, G. Chryssolouris: Symbiotic human-robot collaborative assembly; CIRP Annals -- Manufacturing Technology, Vol. 68. No. 2. (2019)
- [3] Cs. Kardos, Zs. Kemény, A. Kovács, B. E. Pataki, J. Váncza: [Context-dependent multimodal communication in human-robot collaboration](#); 51st CIRP Conference on Manufacturing Systems, Procedia CIRP 72 (2018) p. 15-20

- [4] Pedone Gianfranco, Beregi Richárd, Erdős Gábor: Manufacturing Execution and Support System (MESS) for Future Intelligent Factory Architecture (FIFA), Functional Specification, Kutatási jelentés (IPAR 4.0 Kutatási és Innovációs Kiválósági Központ, GINOP-2.3.2-15-2016-00002), SZTAKI, 2018
- [5] Kardos, Cs., Kovács, A., Váncza, J.: Decomposition approach to optimal feature-based assembly planning. CIRP Annals - Manufacturing Technology, 66(1), 2017
- [6] SYMBIO-TIC Final Video
http://www.symbio-tic.eu/index.php?option=com_k2&view=item&id=734:symbio-tic-final-video&Itemid=33

Robotos szerelési kísérletek háromujjas megfogóval

Paniti Imre^{1,2}, Nacsá János^{1,2}, Abai Kristóf^{1,2}

¹ Termelésinformatikai és Termelésirányítási Kiválósági Központ (EPIC),
Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI),

² Járműipari Kutatóközpont (JKK), Széchenyi István Egyetem (SZE)
paniti.imre@sztaki.hu, nacsajanos@sztaki.hu, abai.kristof@sztaki.hu

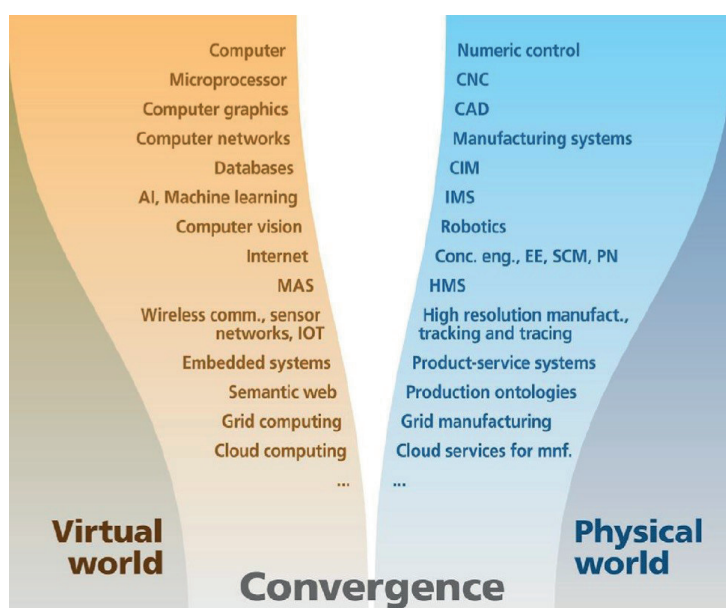
Robotic assembly experiments with three-finger gripper

The article introduces a collaborative workspace built around the Industry 4.0 Sample System with a Collaborative Robot (a so called Cobot) using a 3-finger gripper. It describes in detail the developments that allow independent movement of the 3 fingers. It shows what kind of force sensors are used in the system and the capabilities of them. The Virtual Reality (VR) model of the system is also presented, where the model follows the movements of the system in real-time by continuously reading the Kinect sensor (RGB camera and depth sensor) input and the joint coordinates of the robot. Finally, three demonstration applications are described.

Keywords: Robotised assembly, Industry 4.0, VR modelling

Bevezetés

Az ipar folyamatos fejlődésének köszönhetően az elterjedt és széles körben alkalmazott technológiák kereteit lassan kinőtték az egyre növekvő és szerteágazó igények. Ezek a változások olyan jelentősek, hogy új ipari korszaknak tekinthető ezek bevezetése, így joggal mondhatjuk, hogy napjainkban is zajlik a negyedik ipari forradalom, vagy ismertebb szóhasználatnál élve, az Ipar 4.0 korát éljük. Ez az átalakulás jól szemléltethető az 1. ábrán látható virtuális és fizikai világ konvergenciájával.



1. ábra: A virtuális és fizikai világ konvergenciája [1]

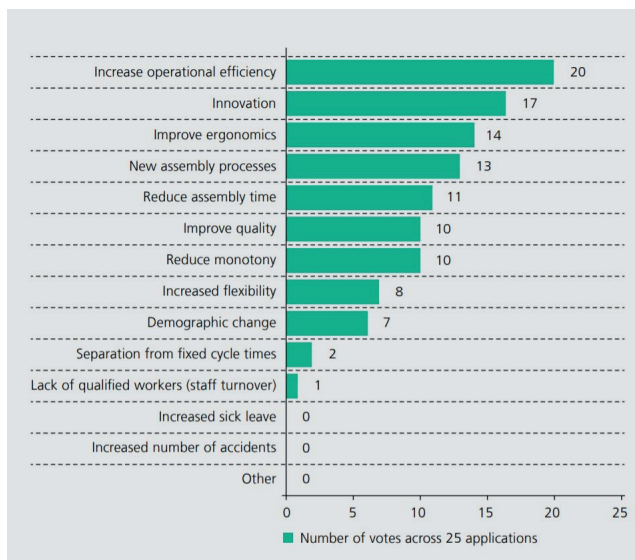
Mivel rengeteg kérdés merül fel az új technológiákkal kapcsolatban, és számos még megoldásra váró feladat áll előttünk, szükségét láttuk egy olyan laboratórium létrehozásának, melyben megépíthető egy olyan kis ipari környezet, egy mintarendszer, melyben lehetőség van az újítások kipróbálására, tesztelésére, összetettebb rendszerek tervezésére és implementálására, illetve lehetőség van új felmerülő problémák megoldására és a megoldások bemutatására is. Ezen felül a rendszer az oktatásban is jelentős szerepet vállalhat, lehetőséget nyújt arra, hogy a hallgatók, már az egyetemi tanulmányaik alatt megismerhessék a fejlődési irányokat, találkozzanak az új technológiákkal.

A 4. Ipari forradalom magába foglalja a robotok fejlődését is, így a hagyományos ipari robotok mellett megjelentek a könnyűsúlyú kollaboratív (együttműködő) robotok, amelyek mozgásától már nem kell megvédeni az embereket különböző ipari biztonságtechnikai eszközökkel, mivel felépítésükből és vezérlésükből adódóan nem árthatnak különösebben az azokkal kapcsolatba kerülőknek.

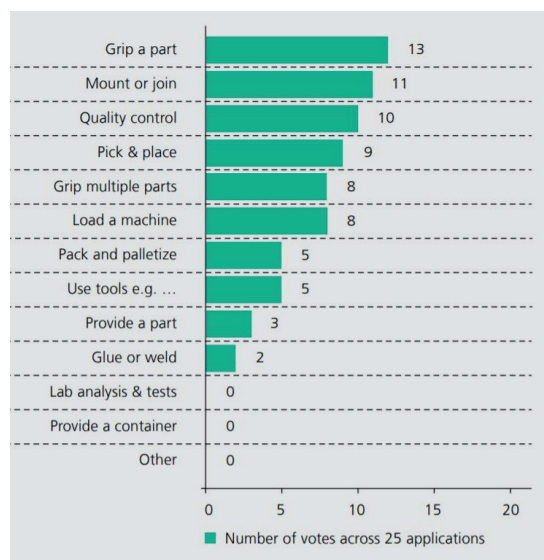
A Fraunhofer Institute for Industrial Engineering (FhG IAO) készített egy tanulmányt [2], melyben arra mutat rá, hogy az ember-robot együttműködés hatékonyan alkalmazható, az automatizált és tisztán humán munkaerőt foglalkoztató termelésben is. A tanulmány célja továbbá az volt, hogy megmutassa a kis- és középvállalatoknak, hogy a könnyűsúlyú robotok alkalmazása számukra is megtérülő beruházás lehet, növelheti a termelékenységet még abban az esetben is, ha kis sorozatnagyságokat gyártanak nagyszámú termékváltozat mellett. Az eddigi nagy beruházást igénylő automatizálás most a relatív olcsó könnyűsúlyú robotoknak köszönhetően nemcsak a nagyvállalatok privilégiuma.

A felmérésben feltárták, hogy a cégek miért választották az adott alkalmazás tervezésekor, a könnyűsúlyú robotok alkalmazását, az ember-robot együttműködést (2. ábra). A megkérdezett vállalatok 80%-a számított arra, hogy ezzel a beruházással hatékonyabb lesz a termelés. Sok cég innovációs céllal fogott bele a kollaboratív munkavégzésbe, és a cégek több mint felénél szempont volt az ergonómia növelése, ezzel a dolgozók munkájának kényelmesebbé és biztonságosabbá tétele. Ezen kívül még sok más indokot felsoroltak a gyártók, mint például az összeszerelési idő csökkentése, monotonitás csökkentése, rugalmasság növelése.

A 3. ábrán látható, hogy a felmérésben szereplő alkalmazásokban milyen típusú művelet végrehajtására alkalmazták a robotokat.



2. ábra: könnyűsúlyú robotok alkalmazásának okai [2]



3. ábra: felmérésben szereplő robotok alkalmazásának típusai [2]

A felmérésből látszik, hogy a robotok nagy százalékban valamilyen megfogó segítségével alkatrész manipulációt végeznek. Ehhez általában vákuumos megfogót, vagy kétujjas megfogót alkalmaznak.

Kobotos szerelésben alkalmazott kétujjas megfogók és korlátaik

Az iparban használatos kétujjas megfogóknak a költsége sok esetben egy nagyságrendbe eshet a robot költségével, továbbá ahhoz, hogy a megfogó egy adott termék megfogására alkalmas legyen megfogópofa tervezése és gyártása is szükséges, amely sok esetben nem egyszerű, gyakran költséges feladat és a termék esetleges változása, vagy egy másik termékre való átállás másik megfogópofa alkalmazását igényli. Ennek ára bizonyos esetekben minimálisra csökkenthető, amennyiben a megfogópofa 3D nyomtatás segítségével előállítható. Előfordulhat, hogy egyes alkalmazásoknál komplexebb megfogást kell alkalmazni, ami már a két ujjas megfogóval nem kivitelezhető. Ezekben az esetekben jöhetnek szóba a háromujjas megfogók.

Háromujjas megfogó alkalmazási módjai és vezérlésének továbbfejlesztése

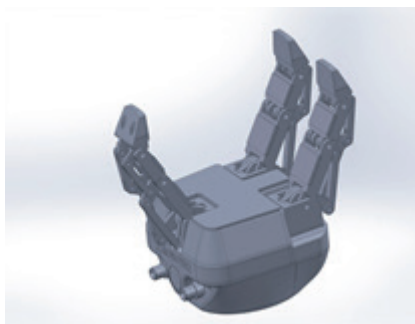
A 4. ábra bal oldalán a "pinch" módban a megfogó 2 ujjasként működik, mivel az egyező oldalon levő "B" és "C" ujjja összezárva együtt mozog. A következő mód az "scissor", amikor pont ezt az összezáró-szétnyíló képességet használjuk egy tárgy felcsippentésére. A harmadik "wide" módban a "B" és "C" ujjak legyezőszerűen szétnyílván széles megfogást biztosítanak hosszabb tárgyak biztos megfogására. A bal szélső "normal" megfogás esetén a három ujj az összezáródás síkjában párhuzamosan mozog és a tárgy relatív helyzetétől függően rá is fordulnak az ujjpercek a tárgyra.



4. ábra: A Robotiq háromujjas megfogójának fogástípusai

Ilyen típusú megfogót alkalmazott négy csapat (köztük a győztes is) a DARPA 2013-as robotikai versenyén [3].

A LinkageDesigner [4] a Wolfram Mathematica egy add-in csomagja, amely mechanizmusok tervezésre, modellezésre, valamint szimulációra alkalmazható. 2006-óta a Wolfram Research terméklistáján is szerepel. Ennek a programcsomagnak a használatával képesek vagyunk egész ipari cellák modelljeinek előállítására, segítségével létrehozhatjuk a cella digitális ikerpárját. Ahhoz, hogy el tudjuk készíteni a megfogó „Linkage” modelljét, pontos geometriai modellre volt szükség. Mivel a gyártó csak befoglaló geometriát ad közre, így „Reverse engineering” módszerrel elkészítettük a megfogó pontos geometriai modelljét (5. ábra), így a Linkage skeleton modellje már pontos geometriai paraméterekkel is feltölthetők.



5. ábra: A háromujjas megfogó 3D modellje

Az így kapott megfogó modell integrálható egy robotos cella modelljébe, mely mint digitális ikerként pl. az ApertusVR segítségével megjeleníthető.

Az ApertusVR [5] egy MIT licenszű, nyílt forráskódú, platformfüggetlen programozói könyvtár, amit az SZTAKI megosztott virtuális valóság csoportja fejleszt, többek között azzal a céllal, hogy az iparban alkalmazott merev, vizualizációs célokra használt szoftvereket kiváltsa. Külső szenzorok jeleinek beágyazása és megjelenítése sokkal egyszerűbb ezzel a rendszerrel, mint a hagyományos, többnyire 3D-játékok fejlesztésére használt programokkal. Virtuális, kiterjesztett és kevert valóság integrálására egyaránt használható, így a robotos környezetünkben is kiválóan tudjuk alkalmazni a mélységi kamerából nyert adatok és a 3D modell „összefésülésénél”. Skálázhatósága és elosztottsága révén könnyen illeszthető már



meglévő alkalmazásokhoz (mint például az IBM Node-RED [6] nevű programjához), egy új alapvetően vizuális absztrakciós réteget képezve a hardverek és a szoftverek között.

Az előbb említett Node-RED-et kifejezetten IoT alkalmazások gyors prototípus-fejlesztéséhez találták ki, ahol különböző típusú rendszereket kell összekapcsolni.

Mi az ApertusVR könyvtárat kapcsoltuk össze a magas szintű robotvezérlésünkkel (az URSZTAKI-val), amiben a 3 ujjas megfogó is helyet kapott.

Az URSZTAKI keretrendszer létrehozásának több célja is volt.

Egyrészt egy olyan felület létrehozása, ami lehetővé teszi a robotok moduláris programozását, hogy a robot-programnyelv-függetlenítés mellett a későbbiekben magasabb szintű vezérlőhöz (pl. MES-hez) lehessen csatlakoztatni a rendszert.

Másrészt olyan keretrendszer kidolgozása, amelyben könnyen lehet létrehozni új alprogramokat (feladatokat) és ezeket egyszerűen el lehet juttatni további robotokra.

Harmadrészt olyan modularitás kidolgozása, amellyel könnyedén kezelhetőek a különböző felszereltségű robotok.

Ezen célok elérése érdekében a program script fájlokra épül, amelyek beágyazódnak a robotprogramba, majd onnan néhány függvény hívásával indulnak. Ezáltal a módosításokhoz csak szöveges fájlokat kell átírni, azok frissítése könnyen automatizálható és verziójuk követhető. A megvalósítás esetünkben UR robotkarokkal történt.

A Robotiq 3 ujjas megfogója kifejezetten támogatja UR kobotokra való felszerelését, mind hardveresen, mind a kobot script nyelvbe ágyazódó speciális utasításkészlettel. A részletes dokumentáció azonban megmutatta, hogy az UR script szoftver nem használja ki a megfogó összes képességét (4. ábra), csak egy limitált elemkészlet került implementálásra. Ebből az okból kifolyólag saját fejlesztéssel a scriptet bővítettük, mely már lehetővé teszi az egyes ujjak független mozgatót mind a négy mozgási irányba ("A", "B" és "C" ujjak valamint a "B" és "C" ujj egymás felé való "S" mozgását). A saját fejlesztésű script mozgásonként megadja a művelet után a fenti négy pontos pozíciót, valamint azt, hogy egy megfogási művelet során melyik ujj érte el a programozott értéket, vagy tárgyat érzékelve korábban állt-e meg. Ezzel a bővített scripttel fut az URSZTAKI.

A csukló-koordinátáinak a robotvezérlő egységétől TCP/IP protokollt és egy belső hálózatot használva jutnak el a 3D-s megjelenítőhöz, ami szinte valós időben tudja követni a robot mozgását, egyfajta digitális ikert, vagy más néven árnyékot képezve (6. ábra).



6. ábra: UR10-es kobotos környezet és a digitális iker megjelenítése a háttérben

Példák bemutatása

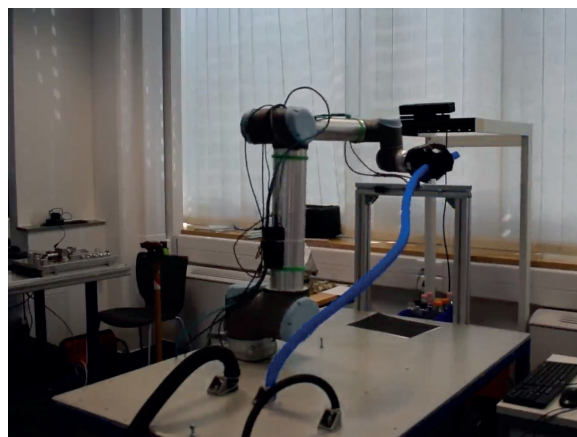
Ebben a részben három példán keresztül szeretnénk bemutatni az eddigi fejlesztéseinket.

1. példa: 3D szkennelés robottal – 7. ábra

A kollaboratív munkakörnyezethez véleményünk szerint hozzátartozik, hogy az emberek és a kobotok osztozzanak a munkaeszközökön is. Ez csak akkor lehetséges, ha mindkét fél meg tudja fogni a tárgyalt eszközt. Az embereknek szánt termékek esetében az ergonómia már alap terméktervezési követelmény, de amennyiben azt szeretnénk, hogy a kobot is meg tudja fogni ugyan azt az eszközt, akkor gyártásnál erre is figyelni kell, vagy ki kell egészíteniünk egy adapterrel a terméket, aminek a csatlakozását biztosítani kell. Esetünkben az utóbbit választottuk, hogy egy állványra is rögzíthető kézi szkennert a kobot mozgatni tudjon. Az optimális szkennelés céljából a TCP (Tool Center Point) értékét (ami a robot „munkapontja”, egy relatív pozíció az utolsó csukló végéhez viszonyítva) a szkennelési távolságnak megfelelően kellett beállítani, amit az általunk használt Artec Space Spider technikai specifikációja [7] határozott meg.



7. ábra: UR10-es kobotos szkennelés



8. ábra: Rugalmas cső manipuláció UR10-es kobottal

2. példa: Rugalmas csövek csatlakoztatása – 8. ábra

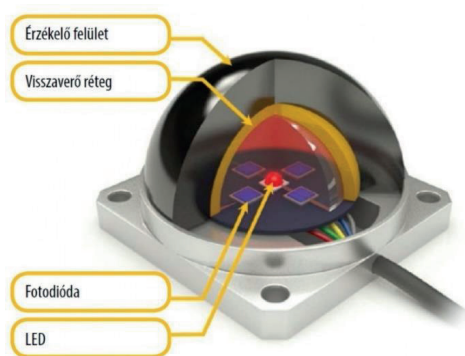
Az iparban egy nem szokványos és kihívásokkal teli feladat a rugalmas csövek manipulációja, ami egy háromujjas megfogóval stabilabb megvalósítást eredményezhet. Esetünkben nem csak a megfogó normál megfogását alkalmaztuk, de a megfogási detektálást is. A robot megfelelő mozgáspályájának programozása történhet kizárólag az egyik rögzített csővég, a csőhossz és a csatlakozási pozíció ismeretében, vagy a később akár az ArtiMinds RPS [8] és egy erő-nyomatékmérő cella segítségével.

3. Példa: Műanyag palack felemelése – 9(c) ábra

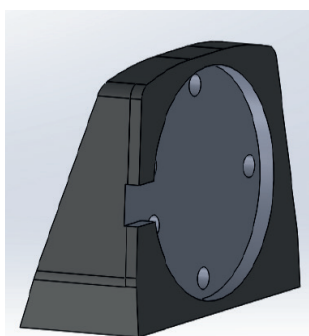
Robotos környezetben sok olyan feladat létezik, ahol valamilyen kis mechanikai teherbírással rendelkező, törékeny vagy könnyen deformálható tárgyat kell manipulálni a robot segítségével. Már tárgyaltuk, hogy milyen sok megfogási módot alkalmazhatunk, azonban egyiknél sem tudunk valós idejű és pontos visszajelzést kapni a megfogott testről.

Ahhoz, hogy a megfogónkat minél összetettebb és egy egyszerű megfogó képességeit meghaladó feladatok megoldására is képessé tegyük, az ujjak utolsó tagjára (az ún. „Distal Phalanx” részre) egy-egy erőmérő szenzort – 9(a) ábra – rögzítettünk 3D nyomtatott adapter – 9(b) ábra – segítségével [9], így pontosan mérhetjük, hogy a megfogási pontokban mekkora erőhatás éri a megfogott tárgyat. A kísérletek várakozásunkkal ellentétben azt mutatták, hogy a pontos erőhatás mérése mellett nem becsülhető jól a felemelt tárgy tömege, mert az ismétléses mérések nagy szórást mutattak.

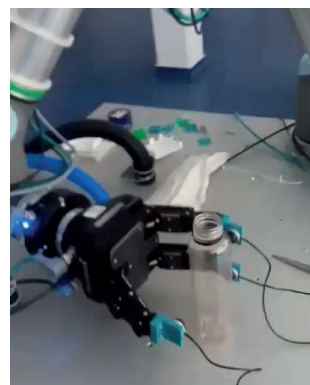
Ugyanakkor sikeres volt a feladat abból a szempontból, hogy ezzel a rendszerrel képesek vagyunk felemelni egy üres PET palackot, annak megroppantása nélkül, amit abban az esetben is megtart a robot, ha azt emelés közben megtöltjük vízzel.



(a)[10]



(b)



(c)

9. ábra: Erőmérő szenzorokkal kiegészített megfogó alkalmazása

Hasonló kísérleteket végzett KABOLI kollégáival együtt [11].

Összegzés

Az ismertetett környezet alkalmas arra, hogy gyorsan újabb példaalkalmazásokat fejlesszünk, az Ipar 4.0 elveinek megfelelően az alkalmazások digitális ikerpárját is létrehozva, valamint egyedi feladatokra is egységesített környezetben dolgozhassunk. A három példa mutatja a 3 ujjas megfogó sokrétű kihasználhatóságát. Ugyanakkor a tapasztalatunk azt mutatja, hogy a megfogó sokféle üzemmódjának előnyeit jelentősen korlátozza a megfogó relatív nagysága a robotkarhoz viszonyítva. Sok alkalmazás, amely felmerült, a megfogó mérete és kábeligénye miatt (3 csatlakozása van) nem valósítható meg. Az ujjakra szerelt erőszensorok helyett a közeljövőben a kar és a megfogó közé illesztett hagyományos 6 tengelyes erőszensorral kívánjuk a kísérleteket folytatni.

Köszönetnyilvánítás

A bemutatott kutatást a „Felsőoktatási Intézményi Kiválósági Program – Digitális ipari technológiák kutatása a Széchenyi István Egyetemen” projekt (20523-3/2018/FEKUTSTRAT) támogatta. A kutatáshoz szükséges infrastruktúra pedig az „Ipar 4.0 kutatási és innovációs kiválósági központ” projekt (GINOP-2.3.2-15-2016-00002) keretében valósult meg.

Irodalomjegyzék

- [1] Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W. and Ueda, K., 2016. Cyber-physical systems in manufacturing. *Cirp Annals*, 65(2), pp.621-641.
- [2] Bauer, W., Bender, M., Braun, M., Rally, P. and Scholtz, O., 2016. Lightweight robots in manual assembly—best to start simply. *Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart*.
<https://www.produktionsmanagement.iao.fraunhofer.de/content/dam/produktionsmanagement/de/documents/LBR/Studie-Leichtbauroboter-Fraunhofer-IAO-2016-EN.pdf>, utolsó letöltés: 2019.10.16.
- [3] <https://blog.robotiq.com/bid/69213/DARPA-Robotic-Challenge-2013-Won-Using-Robotiq-Adaptive-Grippers>, utolsó letöltés: 2019.10.16.
- [4] www.linkagedesigner.com/, utolsó letöltés: 2019.10.16.
- [5] <http://apertusvr.org/>, utolsó letöltés: 2019.10.16.
- [6] <https://nodered.org/>, utolsó letöltés: 2019.10.16.
- [7] <https://www.artec3d.com/portable-3d-scanners/artec-spider#specifications>, utolsó letöltés: 2019.10.16.
- [8] <https://www.artiminds.com/>, utolsó letöltés: 2019.10.16.
- [9] Bencze Kristóf: Erőszensor alkalmazása robotos környezetben, Diplomaterv, BME GTT, 2018
- [10] <http://www.techmonitor.hu/piacmonitor/elemezes-velemen/az-ero-legyen-veled-20161001>, utolsó letöltés: 2019.10.16.
- [11] Kabolí, M., Yao, K. and Cheng, G., 2016, November. Tactile-based manipulation of deformable objects with dynamic center of mass. In 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), IEEE, pp. 752-757.

Tartalomszolgáltatások: könyvtárak, levéltárak, múzeumok

GDPR és a köznevelési intézmények – a GDPR bevezetésének 13+1 lépése a köznevelési intézményekben

Albert Ágota
adatvédelmi tisztviselő
dralbertagota@gdprszakszeruen.hu

GDPR in the institutions of public education

The topic of this presentation is the steps of the formation of an adequate data protection of the GDPR instructions in the institutions of public education and the discussion of the most important problems, which are particularly related to these institutions: how can we start becoming GDPR-compatible? What are the greatest obstacles (besides the issue of how to take pictures) and „how we can put out the fire”? How does an institution of public education differ from a „simple” member of any market?

Since the 25th of May, 2018, these institutions – independent from their maintainer – are being regulated by the GDPR, which means a great challenge to them. Apart from the vast amount of data that has to be processed daily, they face situations like

- the processing of data is regulated by a number of sector laws (law on educational register, law on public education, Public Servants' Act, Labour Code, law on taxes, etc.)
- the special composition of the data subjects (children, parents, employees, other data subjects)
- as data controllers they have complex social network (maintainer, local authority, community of parents, student organizations, churches, supervising organs, different authorities, business partners, foreign partner institutions etc.), through which significant transfers of personal data happen
- they can process data not only as data controllers but also as data processors, (e.g.: secondary school entrance exams, organizing high level „érettségi” etc.) and
- the supervision of employees and students requires a different approach from schools than from a traditional member of the market (using the followings (e.g.): register of working hours, video monitoring system, smart devices and internet in the possession of institutions).

Keywords: GDPR, public education, data protection

Hogyan válhat egy intézmény GDPR-kompatibilissé?

1. lépés

Tudatosítsuk, a GDPR igen is vonatkozik az intézményünkre. Ha homokba dugjuk a fejünket és azt mondjuk hogy nem, attól még igen. Sőt, mindenért az adatkezelő, azaz az intézményvezető a felelős egyszemélyben. Azonban nem elég, ha csak az igazgató érzi úgy, a nyakába szakadt még egy teljesítendő feladat, hanem az



iskola minden alkalmazottjának tudnia kell, ha adatot kezel, azt a GDPR előírásainak megfelelően kell tennie. Vonatkozik ez az osztályfőnökökre, a rendszergazdára, az iskolapszichológusra és a portásra is, mindenkire, aki valamilyen módon személyes adattal kapcsolatba kerül. Természetesen más adatokhoz fér hozzá egy gazdaságis és másokhoz a könyvtáros, éppen ezért fontos, hogy mindenki tisztában legyen a jogosultságával és felelősségi körével.

2. lépés

Vegyünk igénybe szakembert. A rendelet alkotói úgy gondolták, az iskolák mint közérdek alapján közfeladatot végrehajtó intézmények az adatvédelem szempontjából „veszélyes üzemek”, ezért kötelező adatvédelmi tisztviselőt (a szakzsargonban DPO-t) kinevezniük. Ez a fenntartó kötelessége, a kinevezés elmaradása GDPR-ellenes. Az adatvédelem speciális ismereteket követel, amelyek nem lesznek meg attól, hogy például egy pedagógusra rátestáljuk ezt a feladatot. Sőt, előtanulmányok nélkül egy-két napos intenzív GDPR kurzus is maximum annyira elég, hogy az áldozatban tudatosuljon, mennyi mindent nem tud. Olyan személy sem lehet DPO, aki munkaköréből adódóan felelős az adatvédelmi szabályozás betartásáért (pl. az intézmény vezetője).

Az adatvédelmi tisztviselő csak tanácsokat adhat, a döntés az adatkezelőé, ahogy a felelősség is.

3. lépés

Térképezzük fel, milyen személyes adatokat kezel az iskolánk. Ahogy az egészségügy, úgy a köznevelés is elképzelhetetlen adathegyek nélkül.

Az adatok feltérképezéséhez elkerülhetetlen az olyan alapfogalmak megismerése, pl. mi a személyes adat és mit jelentenek a személyes adat különleges kategóriái, hiszen ezek pontos behatárolása nélkül elsikkadhatnak igen fontos adatkezelések. Személyes adat egy osztályzat? Igen. És különleges személyes adat az, hogy Bence vagy Hanna mogoróallergiás? Igen. És az személyes adat, hogy az ötödikesek közül a magas vörös kölyök intőt kapott? Igen, amennyiben az ötödikesek között annyi magas vörös kölyök van, hogy már ennyi információval is beazonosítható az adott diák.

Aharmadik lépés tehát fellelni az intézményben a személyes adatokat, beazonosítani az adatkezeléseket és olyan adattérképeket gyártani, amelyek alapján tisztázható egy-egy adat életútja, azaz

- melyik adat honnan és hogyan érkezett be (szülőtől, tanulótól, fenntartótól stb.),
- ki kezeli azt házon belül (titkárság, osztályfőnök, gazdaságis kolléga stb.),
- hol és hogyan tároljuk (papíron, elektronikusan vagy mindkettőn),
- hogyan, hova és mi alapján továbbítjuk (lásd a köznevelési törvény felsorolását) és
- mikor kell leselejtezni (megsemmisíteni) azt.

4. lépés

Tisztítsuk meg az adatvagyonunkat. A felesleges munka elkerülése érdekében el kell döntenünk, hogy mely adatot kezelhetjük tovább és melyet nem. Főbb szempontok:

- **az adott adat kezelése megfelel az alapelveknek?** Jogszerűen, tisztességesen és átláthatóan kezeljük?
- **van jogalapunk az adott személyes adat kezelésére?** Ha nincs, akkor vagy teljesíteni kell a hivatkozott jogalaphoz szükséges feltételt (pl. be kell szerezni a hozzájárulást), vagy kíméletlenül ki kell selejtezni (meg kell semmisíteni) azt. A „mindig így szoktuk” jogalapot a GDPR nem ismeri, a hozzájárulás pedig – a közhiedelemmel ellentétben – nem univerzálisan használható jogalap, mivel az bármikor visszavonható, illetve egyes helyzetekben (pl. alá- és fölérendeltségi viszonyban, megkérdőjelezhető önkéntesség esetén) nem alkalmazható.
- **időben szavatol még az adott adat?** A vonatkozó jogszabályok (pl. 20/2012. (VIII. 31.) EMMI rendelet 1. sz. melléklete), illetve az intézményünk irattári terve tájékoztatást nyújt, hogy melyik adat meddig őrizhető. Ha lejárt a megőrzési idő, a kezelés nem jogszerű.
- **szükség van az adott adatkezelésre? Van konkrét cél** vagy csak raktározzuk az adatot, majd csak jó lesz valamire? Ha nincs cél, az adat kezelése nem jogszerű.
- **az adatkezelés céljának eléréséhez feltétlen szükség van az adott adatra?** A GDPR elkötelezett híve az adattakarékosságnak, azaz nem támogatja olyan adat kezelést, amely nem szükséges az adott cél eléréséhez. A felesleges adatoktól szintén meg kell szabadulnunk.
- **helyes, pontos az adat?** Ha nem az, akkor kijavítható, helyesbíthető? Ha nem, döntenünk kell a sorsáról. A pontatlan, hibás személyes adat is személyes adat.
- **megfelelő biztonságot biztosítunk az adatnak?**

Minél nagyobb és minél régebbi az intézményünk, annál nagyobb a feltérképezés és kiválogatás munkaigénye, ráadásul nem ez az a feladat, amit gyakornokokra felelősen rá lehetne bízni.

Ha lekarcsúsítottuk az adatvagyonunkat, jöhet az **adattérképek újrarajzolása**. Ezek már alkalmas kiinduló pontjai az iskola adatkezelési nyilvántartásának, amely a GDPR alapján kötelező, és amelyet egy adatvédelmi hatósági ellenőrzés során – az elszámoltathatóság elvének megfelelően – be kell tudni mutatnunk. Ahogy az iskola leltározza a számítógépeit, labdáit meg csontvázait, úgy az adatkezeléseit is leltároznia kell.

5. lépés

Határozzuk meg az egyes adatkezelések követelményeit. Az adatkezelések döntő többségében jogszabályok határozzák meg, hogy milyen adatot kezelhet, hova továbbíthat az iskola. Az iskolák közérdekű feladataik végrehajtásához szükséges adatkezeléseknél a jogalap a GDPR 6. cikk (1) bekezdés e) pontjában meghatározott közérdek lesz és ez a jogalap – a felügyeleti hatóság álláspontja szerint – elnyeli a



további adatkezelési jogalapokat, valamint be kell tartani a GDPR alapelveit is.

Vannak olyan adatkezelések, amelyek bonyolultabbak, mint egy átlag munkáltató adatkezelései. Egy iskolában pl. sokkal nehezebb a munkahelyi és magánhasználatot elkülöníteni – decemberben a pedagógus használatában lévő számítógépről keresett beigli recept a karácsonyi vacsorához kell vagy a hagyományokkal foglalkozó osztályfőnöki órára? A cicalány007@ e-mail cím a titkos barátnő címe vagy egy diák anyukájáé, netalán magáé a diáké? Talán nincs még egy olyan munkáltató, ahol a magánszférát olyan nehéz elválasztani a munkahelyi kötelezettségek teljesítésétől, mint egy átlagos iskola.

Lehet még „problémás” adatkezeléseket sorolni, pl. a „normál” körülmények között általában jogos érdekre hivatkozó adatkezelésekhez milyen jogalap használható az iskolákban? Például:

- az intézményben felszerelt kamerák felvételeivel,
- a diákok internethasználatának ellenőrzésével,
- az iskolai rendezvényen készült tömegfelvételekkel kapcsolatos adatkezelés, amelyeken nemcsak a ballagó diák van, hanem közeli és távoli rokonsága nullától száz éves korig, és amely képek a modern technikának köszönhetően bármikor portréfotó minőségre nagyíthatók.

Az adattovábbítások vizsgálata is szükséges, hiszen az iskola nemcsak jogszabály alapján továbbít adatokat harmadik személyeknek, hanem pl. testvériskolai kapcsolattartás vagy versenyekre jelentkezés keretében is. Sőt, iskolai kórustalálkozót szervezhetnek pl. Kinában, amely harmadik országnak számít, szállást pedig oda is foglalni kell sok-sok személyes adat megadásával. Ráadásul az adattovábbítást nyilván kell tartanunk, különben hogyan fogjuk tudni megmondani az érintettnek, hogy mikor, hova, kinek, mi alapján és milyen garanciákkal továbbítottuk a személyes adatait?

Kétségeink eldöntésében segítséget nyújthatnak a NAIH anyagai.¹

6. lépés

Igazítsuk a belső szabályzatainkat és utasításainkat a GDPR előírásaihoz. Az iskoláknak kötelező adatvédelmi szabályzatot készíteniük, a GDPR azonban a konkrét tartalmat az adatkezelőre bízta.

Mi legyen benne? Amit fontosnak tartunk, pl.:

- adatkezelések részletes leírása, különösen a speciális eseteké (pl. tanulmányi és tantestületi kirándulással, iskolaújsággal, intézményi honlappal kapcsolatos adatkezelések, dokumentum másolatok kezelése stb.)
- adatfeldolgozókkal kapcsolatos rendelkezések, illetve az olyan adatkezelések leírása, amelyek esetében az iskolánk adatfeldolgozó (pl. központi felmérések, tanulmányi versenyek stb.)

¹ <https://www.naih.hu/adatvedelmi-allasfoglalasok,-jelentesek.html>

- érintetti jogok leírása és a jogok érvényesítésének módja (az érintetti tájékoztatás formái és kötelező tartalma, érintetti kérelmekre válaszadásnak, a hozzájárulás visszavonásának, az adatkezelés ellen tiltakozásnak, törlési kérelemnek adminisztrációja stb.)
- kötelező, illetve elszámoltathatósági szempontból fontos nyilvántartások tartalma (pl. az adatkezelések, adattovábbítások és adatvédelmi incidensek nyilvántartása stb.)
- adatvédelmi incidens esetén követendő protokoll előre elkészített nyomtatványokkal, nehogy a nagy kapkodásban kimaradjon valami,
- olyan speciális adatkezelések, mint osztálycsoport létrehozása közösségi oldalon, oktatással kapcsolatos appok használata (pl. Kahoot!), tanárok-diákok saját IoT-eszközeinek használata stb.
- beépített adatvédelem érvényesítéséhez szükséges eljárások (pl. adatok biztonságos tárolásának, továbbításának módja, jogosultságok kiosztásának elvei stb.)
- az adatvédelmi tisztviselő feladatai és hatásköre.
- Az adatvédelmi és adatbiztonsági szabályzaton kívül szükségünk van még egyéb dokumentumokra is, pl.
- speciális szabályzatok megalkotása (kameraműködtetési, beléptetési, informatikai, információ biztonsági, image policy stb.)
- érdekmérlegelési tesztek végzésének eljárásrendje, kockázatfelmérés és kezelés módszerei stb.

7. lépés

Készítsük el az adatkezelési tájékoztatókat. A GDPR sarkalatos pontja az érintettek tájékoztatása, ennek elmaradása komoly szankciókat vonhat maga után annyi kedvezménnyel, hogy költségvetési intézmény esetén a maximális közigazgatási bírság 20 millió forint.

Milyen tájékoztatókra van szükség? Célszerű minimum három adatkezelési tájékoztatót készíteni, egyet a diákoknak, egyet a szülőknek, egyet pedig az alkalmazottaknak. A fenntartónak gondolnia kell még olyan egyéb érintettekre is, mint pl. a szerződéses partnerek, nekik is kell készíteni tájékoztatót.

Az általános adatkezelési tájékoztató általában az iskola honlapján kap helyet, de mindenképpen kell nyomtatott verzió is, amely az intézmény titkárságán bármikor elérhető. A speciális, egy-egy adatkezelésre vonatkozó tájékoztatók helyet kaphatnak pl. a beiratkozási és hozzájárulást kérő nyomtatványokon is, így később nincs vita a tekintetben, hogy megtörtént-e az érintettek szükséges tájékoztatása.

8. lépés

Ültessük át a gyakorlatba a gyerekekkel kapcsolatos eltéréseket. Pl. mi legyen a 14-18 év közötti gyerekekkel illetve akkor, ha nagykorú lesz a diák? Ki férhet hozzá a tanuló adataihoz, ki gyakorolja a szülői felügyeleti jogot? A szülőnek mindenről joga van tudni, vagy adott esetben a szülővel szemben is védeni kell a gyermek jogait? Elmondhatja-e az iskola, hogy informatika órán milyen témákra keresett



rá a tinédzser a neten, vagy jobb ezekről hallgatni? Milyen plusz garanciákat kell beépíteni az adatkezelésekbe akkor, ha az érintett 18 éven aluli?

A GDPR alapján a diákoknak szóló adatkezelési tájékoztatót úgy kell megfogalmazni, hogy korosztályuknak megfelelő mértékben ők is megérthessék az abban foglaltakat és legyenek tisztában pl. azzal, hogy nem kell nekik olyan fotót eltűnniük az iskola honlapján, amelyen nagyon nem tetszenek maguknak. A tájékoztató készülhet sok oldalas írásos verzióban, de érdemes megfontolni a videofilm vagy képregény formátumot is.

9. lépés

Tegyük azért, hogy az érintettek gyakorolhassák jogaikat, betartva a jogszabályban foglalt határidőt. Nem szabad elfelejtenünk, hogy a szülők speciális érintettek egyrészt saját jogon, másrészt gyermekeik miatt, frusztrációjuk pedig az egeket verheti pl. azért, mert a fiuk vagy lányuk egyfolytában a fülüket rágja, milyen előnytelen kép van róluk az iskola honlapján és legyenek szívesek, intézkedjenek a kép eltávolítása érdekében. Éppen ezért célszerű az érintetti kérelmek megfelelő rendezése érdekében a belső adatkezelési szabályzatban részletesen rögzíteni az eljárásrendet és minden esetben bevonni az adatvédelmi tisztviselőt is.

10. lépés

Alakítsuk ki az adatvédelmi incidens protokollunkat – és nemcsak a leírása fontos, hanem a begyakorlása is, akár próbaincidents keretében. A 72 órás hatósági bejelentési határidő nagyon rövid és minden egyes bejelentés egyben önfeljelentés is, amely vizsgálatot von maga után. Az incidenskezelés pedig nem annyi, hogy szétárjuk a kezünket, ez most így sikerült, máskor majd biztos gondosabbak leszünk és kicsit nehezebben lopják majd el a tanáriból a nyitott ablakon keresztül azt a laptopot, amely titkosítás nélkül tartalmazott adatot több száz gyerekről és szüleikről.

11. lépés

Ne felejtsük el a beépített adatvédelemről. A GDPR összetolja az adatkezelést és az adatvédelmet, az adatkezelőnek pedig mindent meg kell tennie annak érdekében, hogy az érintettek a lehető legnagyobb biztonságban tudják a személyes adataikat. Az adatbiztonság fontos része pl. a számítógépes rendszer védelme, de ide tartoznak olyan triviális dolgok is, mint a jogosultságok kiosztása, a jelszavakkal kapcsolatos előírások betartatása, a rendszeres biztonsági mentés, a tiszta asztal politika, a szekrények és fiókok zárása, a leselejtezett papírok megsemmisítése stb. Ajánlatos megismerkednünk olyan technikákkal is, mint az álnevesítés, az anonimizálás és a titkosítás és figyelembe kell venni mindezeket már az adatkezelések megtervezése folyamán is.

12. lépés

Oktatás és bevezetés. A polcra feltett 500 oldalnyi szabályzat még nem jelenti azt, hogy az iskolánk megfelel a GDPR követelményeinek. Az adatvédelmi kultúrának be kell épülnie a mindennapokba, és nemcsak beépülnie kell, hanem bizonyítanunk

is tudni kell, hogy a rendeletnek megfelelően jártunk el. Ehhez szükséges az alkalmazottak képzése, hiszen nem egy személyben az igazgató kezeli az adatokat, hanem minden egyes tanár, pedagógiai asszisztens és még a portás is. Ez nem azt jelenti, hogy mindenkinek tudnia kell mindent, hanem azt, hogy a saját területén mindenki sajátítsa el a szükséges ismereteket, és ne csak bemagolja, hanem alkalmazza is az előírásokat. Az elvileg és általában képzéseknek semmi értelme, csakis annak, ha az alkalmazottak munkaköri szintjükhöz kapcsolódva szereznek értelmezhető tudást.

13. lépés

Nemzetközi kapcsolatok. Az iskolának szerteágazó nemzetközi kapcsolatai lehetnek, és nemcsak a nemzetiségi iskolának, hanem a két tanítási nyelvűeknek, nyelvtagozatosoknak és a „sima” intézményeknek is. A nemzetközi kapcsolatok elengedhetetlen része az adattovábbítás külföldi illetőségű adatkezelő vagy nemzetközi szervezet számára, akár személyes adatok, akár különleges adatok formájában. A GDPR alapján az Európai Gazdasági Térség országai belföldnek számítanak és vannak olyan országok is, amelyek a Bizottság véleménye szerint garantálják az uniós védelemnek megfelelő védelmi szintet, ilyen pl. Svájc, Japán és Uruguay, míg olyan országok, mint az Egyesült Államok, Ukrajna vagy Szerbia harmadik országnak számítanak. Ezekbe az országokba irányuló adattovábbítás esetén speciális feltételeknek kell teljesülnie, amelyek fennállását az iskolának kell bizonyítania akkor is, ha csak Kárpátaljára rukkan ki egy pár napra a 10/a. osztály.

13 + 1. lépés

Kövessünk nyomon az eseményeket és reagáljunk, ha szükséges. Kétszer ugyanabba a folyóba nem lehet belelépni és ez igaz az adatvédelemre is. Még ha az uniós rendelet maga nem is változik gyakran, a hazai ágazati jogszabályok bármikor módosulhatnak, ahogy várhatóan a jogértelmezés is idővel egyre kifinomultabb lesz. Nemcsak arra van szükség, hogy az iskola ellenőrizze a saját adatvédelmi gyakorlatát és folyamatosan az újabb és újabb igényekhez igazítsa azt, hanem arra is, hogy a jogértelmezésnek megfelelő irányvonalat kövesse. Ez az a terület, ahol szintén elengedhetetlenül szükséges az adatvédelmi tisztviselő közreműködése.

Kézírásfelismerés Arany János levelein

Bobák Barbara

*Magyar Tudományos Akadémia Bölcsészettudományi Kutatóközpont
Irodalomtudományi Intézet*
bobak.barbara@btk.mta.hu

Gábori Kovács József¹

*Magyar Tudományos Akadémia Bölcsészettudományi Kutatóközpont
Irodalomtudományi Intézet*
kovacs.jozsef@btk.mta.hu

Arany János 1865–1869 között titoknoki, majd 1870–1877 között főtitkári minőségben segítette az MTA munkáját. Az író tehát csaknem másfél évtizeden keresztül látott el fontos tudományszervezési feladatokat.² Életművének kutatása ezért elképzelhetetlen hivatali iratainak, levelezésének feltárása és publikálása nélkül. E dokumentumok vizsgálata kiegészítheti olvasmányainak listáját, tisztázhatja, hogy milyen mértékben vett részt az Akadémia kiadványainak elkészítésében, valamint segítséget nyújthat annak felismerésében, hogy a hivatali foglalatosság milyen, eddig nem ismert kapcsolatokat teremthetett az ő költészetével. Továbbá ezek a dokumentumok kulcsszerepet játszhatnak az Akadémia szervezeti rendszerének leírásában, az 1870-es átszervezés okainak, az Akadémia hazai és külföldi intézményes kapcsolatainak feltárásában; reprezentációs szerepének, illetve tudományszervezésben elfoglalt pozíciójának meghatározásában. Az intézmény sokféle ágazó tudományszervezési feladatait alapvetően meghatározta az a tény, hogy az Akadémia számos tudományterület képviselőit – nyelv- és széptudományok, bölcséleti tudományok, törvénytudományok, történeti tudományok, matematikai tudományok és természettudományok – fogta össze a korszakban. E tudományterületek külső – hazai és külföldi – intézményekkel és kutatókkal való kapcsolattartása pedig rendszerint a mindenkori titoknak, illetve az általa vezetett Titoknoki, később Főtitkári Hivatal közvetítésével történt, így Arany János hivatali levelezésének kiadása az említett tudományterületek és a hozzájuk a korszakban sorolt tudományágak történetének, fejlődésének, hazai és külföldi intézmény- és kapcsolatrendszerének vizsgálatához is nélkülözhetetlen adatokat nyújthat.

1 Gábori Kovács József a tanulmány írásának ideje alatt MTA Prémium Posztdoktori Ösztöndíj támogatásban részesült.

2 Gergely Pál. „Bevezetés”. In: Arany János. „Hivatali iratok 2.: Akadémiai évek (1859–77)”. szerk., jegyz. Gergely Pál. Bp. Akadémiai K. (1964) (Arany János összes művei, 14): 602–604. ; Arany akadémiai feladatairól lásd még Gergely Pál. „Arany János és az Akadémia.” Bp. MTA Irodalomtudományi Intézete – Akadémiai K. (1957). (Irodalomtörténeti füzetek, 11.) ; Keresztury Dezső. „Csak hangköre más: Arany János 1857–1882.” Bp. Szépirodalmi K. (1987): 468–480. ; Voinovich Géza. „Arany János életrajza 1860–1882.” Bp. Magyar Tudományos Akadémia (1938): 205–214.

Korábban, az *Arany János összes művei* című kritikai kiadás XIV. kötetének összeállítása során Gergely Pál elkészítette Arany akadémiai iratainak jegyzékét. Ő körülbelül 2500 dokumentumot sorolt fel, ám ennek csak kisebb részét – 717 – tette közzé.³ Ez a tény és a kötet megjelenése óta eltelt több mint ötven év eredményei szükségessé tették Gergely jegyzékének felülvizsgálatát és kiegészítését. Főként mivel a kötet összeállítása során a szerkesztő szinte csak az Arany által írt hivatali iratok közül válogatott, míg az Akadémia titoknokához írott levelek közül csak a legfontosabbak kerültek a gyűjteménybe.⁴

Mindezeket szem előtt tartva vált indokolttá, hogy Arany János hivatali levelezésére essen a választás a gépi kézírásfelismerés és átírás tesztelésére, valamint fejlesztésére. A papíralapú dokumentumok digitalizálásának fontossága kétségszű. A tartós megőrzés, az egyszerűbb tárolhatóság, a könnyebb és sokrétűbb feldolgozhatóság mind indokolja azt. Függetlenül attól, hogy a humán tudományok területén munkálkodó kutatók a mai napig print kiadásokat, továbbá kéziratokat használnak munkájukhoz, egyre több írásos anyag rendelkezik elektronikus verzióval is, az újonnan születő művek jelentős hányada pedig már eleve digitális formában jön létre. A technika fejlődésével és térhódításával tehát megváltozott a tudásanyag megszerzésének – és létrehozásának – a módja, megváltoztak az információhoz való hozzáférésre vonatkozó igények, ami szintén a digitalizálás elkerülhetetlensége felé mutat.

Egy szöveges dokumentum digitalizálásának legfőbb lépései a szkennelés és a karakterfelismerés, legyen szó akár nyomtatott, akár kézzel írt szövegről. Nyomtatott szövegek esetén az optikai karakterfelismerés (OCR), írott szövegek esetén pedig a kézírásfelismerés (HTR) technológiáját alkalmazzuk.

Az OCR

A digitalizálási folyamat első lépése a szkennelés, amikor a papíralapú dokumentumokról jó minőségű, magas pontsűrűségű (min. 600 dpi⁵) szkennelt képek készülnek. Ezt követően a képeken valamilyen OCR program kerül futtatásra, amellyel további információt tudunk kinyerni a szkennelt anyagból. Az optikai karakterfelismerés (Optical Character Recognition – OCR) olyan technológia, amely felismerhetővé teszi a képen szereplő karaktereket, számokat és központosítási jeleket. A folyamatban nehézséget jelentenek azok a *zaj*nak nevezett elemek, amelyek egy nyomtatott dokumentumban előfordulhatnak, például egy folt vagy gyűrődés a papíron, homályos háttér, tinta-foltok stb. A karakterfelismerő teljesítményét tanulással lehet fejleszteni, így idővel képessé válik olyan karakterek és minták azonosítására is, amelyekkel korábban nem találkozott.⁶

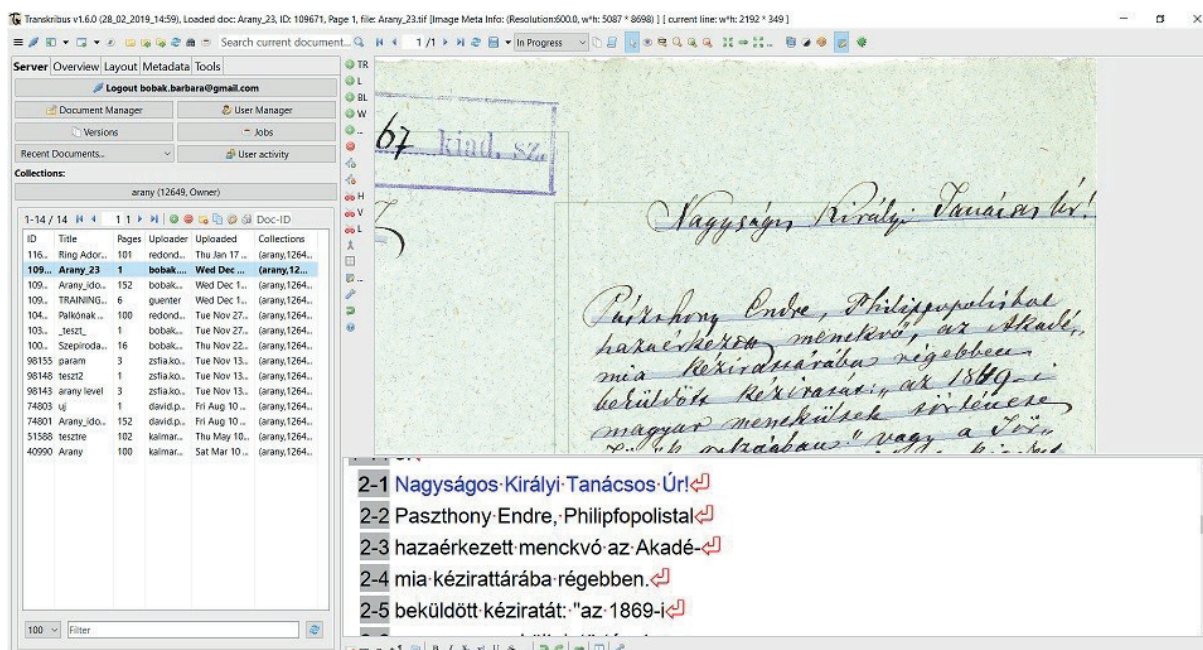
3 Arany János. „Hivatali iratok...”

4 Lásd pl. A Magyar Földhitelintézet pénzügyi osztálya t. c. Igazgatóságának a MTA elnöke. In: Arany János. „Hivatali iratok...”: 589.

5 dots per inch – pont per hüvelyk

6 Optikai karakterfelismerés. Hozzáférés 2019. 03. 10.

<http://szovegbanyaszat.tyotex.hu/content/PDF/ch+ocr.pdf>



1. kép A Transkribus felhasználói felülete

Az OCR alkalmazása által tehát a képekből szövegfájl formátumok jönnek létre, amelyek szerkeszthetők és kereshetők, vagyis alkalmasak a feldolgozásra.

A HTR

Ha a digitalizálni kívánt dokumentum nem nyomtatott, hanem kéziratos, abban az esetben kézírásfelismerő (Handwritten Text Recognition – HTR) programra van szükség. A cél ugyanaz, mint a nyomtatott szövegek esetében, azaz a kéziratról készült szkennelt képfájlból feldolgozásra alkalmas szövegfájl formátum létrehozása. A folyamatot itt is nehezítik mindazok a zajelemek, amik a nyomtatott dokumentumokban is előfordulhatnak, kézírásnál azonban a személyiségjegyekkel is számolni kell, hiszen ahány ember, annyiféle kézírás létezik. Az OCR technológiájához hasonlóan a HTR is tanulási fázison megy keresztül, ami alkalmassá teszi a szöveget alkotó karakterek felismerésére, később pedig új minták azonosítására is.

Az OCR és a HTR térhódításával olyan számítógépes alkalmazások és online szolgáltatások is megjelentek, amelyek képesek mindkét technológiát alkalmazni a feltöltött dokumentumokon.

Transkribus⁷

A Transkribus online szolgáltatás, amellyel nyomtatott szövegek és kéziratok felismertetését és átírását végezhetjük, jelentős mennyiségű munkaórát megspórolva. 2016-ban az Innsbruck Egyetem Digitalizálás és Digitális Archiválás

7 Transkribus. Hozzáférés: 2019. 03. 10. <https://transkribus.eu/Transkribus/>

kutatócsoportja (DEA)⁸ hozta létre a READ⁹ projekt részeként, és azóta is folyamatosan fejlesztés alatt áll. A szolgáltatás, melynek jelentős része nyílt forráskódú, mindenki számára ingyenesen elérhető. A Transkribus, a szolgáltatás honlapjáról egy regisztrációt követően letölthető és telepíthető, használatához azonban internetelérés szükséges.

Az alkalmazás felülete rendkívül felhasználóbarát, logikusan felépített és tagolt, az egyes funkciók könnyen elérhetők. A kialakításnál láthatóan fontos szempont volt, hogy azok is könnyen boldogulhassanak vele, akik a technológiára kevésbé fogékonyak. A könnyen kezelhetőség célkitűzését az is bizonyítja, hogy a szolgáltatás wiki oldalán számos, részletesen leírt és képekkel illusztrált .pdf, valamint videó formátumú használati útmutató tölthető le angol és német nyelven.¹⁰

Első modell	Második modell
fő .lf-én - re. oldy F. ezennele Tidlobb zklá sz. zoalaááédáéb eMI. MM865... 186 186j éé aa	5. 730 1867 kiad. sz. 720 Toldy J. ezennel millobb iejellilölböz t i ibi. 1é851771 MMAÉL. 7. 1885 136 /1867 levélt. szám. G 99t175 d. rérééj8 tá. 3
Nagysags Krályi Taneilcr aszthoy Endre, Phdifyopolisbol hazaészezett. menckvő, az Akadé- mia kezrdttárába régebben. beküldött keziratati , "az 1869-i magyar menekültet története Török országban vagy a For-	Nagyságos Királyi Tanácsos Úr! aszthony Endre, Philipgopolisból hazaérkezett, ménckvő, az Akadé- mia kézirddttárába régebben beküldött kéziratát: "az 1869-i magyar menekültek története Török országban" vagy a Tor-
ténelmi Bizottsag által kiadat- m vagy, högy ő maga kiadhasa mg s ez autal sányorú helyzetzen valamit könyéthessenem, el nm- fogadás esetere visszakerven. a folyó hó 9-én tartott ostályüles határozatából szerencsém van. e dolyozatot aztenni a mondott Bizott sághozs véleményyada vegon .A menmngrevea jedig elő re lavható volaa, högy e kez	ténelmi Bizottság által kiadat- m vagy, hogy ő maga kiadhasa s ez által sanyára helyzetzen valamit könyethessen, el nem, fogadás esetere visszak érvén; a folyó hó 9-én tartott osztályáles határozatából szerincsem vann e dolgozatot áttenni a mondott Bizott- sághozg véleményyada algyosy. A nyenyren pedig elő

2. kép Az első és a második modellel felismertetett levélrészlet

8 Digitalisierung & Elektronische Archivierung (DEA). Hozzáférés: 2019. 03. 10.

<https://www.uibk.ac.at/germanistik/einrichtungen/dea.html>

9 Recognition and Enrichment of Archival Documents. Hozzáférés: 2019. 03. 10.

<https://read.transkribus.eu/about/>

10 Transkribus – How to Guides. Hozzáférés: 2019. 03. 10.

https://transkribus.eu/wiki/index.php/How_to_Guides



Az átíráshoz szánt dokumentumokat *kollekciókba* kell gyűjteni, az egyes kollekciók így számos dokumentumot tartalmazhatnak, mennyiségükre nincs megkötés.

A Transkribus – lehetőségeiből és funkcióiból adódóan – tartalmaz optikai karakterfelismerő és kézirásfelismerő modult is. Egy feltöltött dokumentum feldolgozásának első lépése a megfelelő szegmentálás, azaz a képfájlon lévő szöveg sorainak a beazonosítása és kijelölése. Megadható, hogy a dokumentum egészen vagy csak bizonyos oldalain menjen végbe a szegmentálás. A végeredmény szabad

	Token	Karakter
Első modell	128/200	276/954
	64%	29%

Második modell	48/200	116/954
	24%	12%

1. táblázat A modellek tesztelésének hibaszázalékai

kézzel javítható, vagy teljes egészében pótolható, amennyiben az automatikus azonosítás nem ment végbe megfelelően. Következő lépésként a kijelölt szegmentumokon futtatható az OCR vagy HTR, a feltöltött dokumentum jellegétől függően. Kézirásfelismerés esetén szükséges egy *modell* kiválasztása is, amely az adott nyelv karakterkészletének sajátosságainak felismerésére lett létrehozva, és ami tanítható további feltöltött és hibátlanul átírt dokumentumok révén, ezzel növelve a hatékonyságát. A Transkribus angol és német nyelvre tartalmaz beépített modellt, azonban kérhető bármely más nyelvre is, amennyiben megfelelő mennyiségű (min. 15 000 átírt szó)¹¹ és minőségű átírt dokumentum áll rendelkezésre. Az átírt szöveg végül számos formátumban exportálható: kétrétegű PDF, TEI, DOCX, TXT, stb.

A korpusz

Arany János hivatali iratainak összegyűjtése és feldolgozása az Arany János munkái kritikai kiadáshoz kapcsolódva 2014 szeptemberében vette kezdetét. A munka során eddig hol két, hol egy munkatárs tevékenysége révén kb. 2000 dokumentum betűhív, nyers átírata készült el, nagyjából 2400 oldal terjedelemben, részben .docx formátumban, részben a LyX dokumentum előkészítő rendszer kritikai kiadás készítésére alkalmassá tett, Hegedüs Béla által fejlesztett verziójában. Eközben befejeződött a hivatali iratok túlnyomórészt az MTA Könyvtár és Információs Központ Kézirattárában található anyagának szkennelése is, amely kb. 9200 dokumentumról nagyjából 30000 felvételre eredményezett. A Transkribus teszteléséhez ebből a jelentős mennyiségű anyagból került felhasználásra – ezen szöveg megírásáig – 200 oldalnyi kézirat az Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar Digitális Bölcsészeti Központ kezdeményezésére és együttműködésében.

¹¹ Transkribus – Questions and Answers. Hozzáférés: 2019. 03. 10.
https://transkribus.eu/wiki/index.php/Questions_and_Answers

Tesztelés

A teszteléshez két lépésben, 100-100 oldalnyi kézzel átírt – pontosabban a HTR eredményén javított – levelet biztosítottunk a szolgáltatás fejlesztőinek, a magyar nyelvre vonatkozó modell létrehozására és tanítására. Bár a Transkribus hivatalos leírása szerint már kb. 75 oldalnyi szöveg is elegendő lehet a tanulási fázishoz, a magyar nyelv esetében az első 100 oldalt követően csekély sikerrel tudott a modell boldogulni a kézírással. Teljesítménye összességében körülbelül 30%-ra volt tehető, ha nem vesszük figyelembe például a leveleken szereplő pecsétek szövegeit. Ezeket a modell szintén próbálta kiolvasni, gyakorlatilag eredménytelenül, amiben közrejátszott a pecsétek elhelyezkedése és árnyalata is.

A második körben biztosított további 100 oldalnyi átírással azonban számottevően javult a modell teljesítménye. Nagyságrendileg csupán egy hibásan átírt szó fordult elő soronként. Fontos ugyanakkor megjegyezni, hogy mindegyik alkalommal azonos személytől származó kéziratok átírataival ment végbe a fejlesztés. A cél viszont az, hogy a modell képessé váljon bármilyen magyar nyelvű kézírás felismerésére a lehető legkisebb hibázási aránnyal. Ennélfogva a tesztelés következő szakaszában egy másik személytől származó, de időben ugyanakkor keletkezett kéziratok átíratát szolgáltatjuk a tanulási fázishoz.

Az 1. táblázat a két modell teljesítménybeli különbségeit mutatja számszerű és százalékos formában is. Egy tetszőlegesen kiválasztott levélen le lett futtatva mindkét modell, a végeredményül kapott átíratokból pedig véletlenszerűen ki lett emelve ugyanaz a 200 db token¹². A táblázatból kiolvasható, hogy a modellek mennyi hibát ejtettek a tokenek szintjén, valamint a tokeneken belül a karakterek szintjén. Míg az első modell a mintának több mint felét (64%) elrontotta, a karaktereknek pedig majd harmadát (29%) nem tudta felismerni, addig a második modellnél ezek a hibaértékek valamivel több mint felére csökkentek. Fontos azonban megjegyezni, hogy ezek az arányok nagyon kis méretű mintát jelképeznek, nagyobb méretű, vagy akár a teljes átíratok összehasonlítása ezeket az arányokat valamelyest módosíthatják.

Konklúzió

Az eddig összegyűjtött 9200 hivatali irat kiadása, nyomtatott kötetenként 700 dokumentummal számolva, 13-14 kötetben valósulhatna meg, ami a kritikai kiadások elkészítésének szokásos időtartamával – 8-10 év – kalkulálva, minimum 104 munkaévet venne igénybe. Ezt az időt pedig lerövidíthetné egy kéziratátíró-program használata, még akkor is, ha a program által elkészített nyers átíratok ellenőrzése, a hivatalos átírási elveknek való megfeleltetése, valamint a kritikai és magyarázó jegyzetek elkészítése továbbra is a textológusok feladata lenne. A kézzel írt karaktereket tehát némi betanítás után nagy hatásfokkal felismerni képes program jelentős segítséget nyújthat a szövegkiadások elkészítésében, ám ez főként akkor lehet igaz, ha a szövegkiadó az eredeti írásképet leginkább megőrző betűhű

¹² Szóköztől szóközig tartó karaktersorozat



átiratok készítésére törekszik, illetve, ha az átíró program felhasználói felületén a gép által készített nyers átíraton a kívánt változtatások rögtön el is végezhetők, azaz az említett szövegszerkesztő alkalmas kritikai szöveg készítésére is.

Mindazonáltal a jelenleg – ingyenesen – elérhető online kézírásfelismerő és átíró szolgáltatások közül a Transkribus bizonyára a leghasznosabbnak a szövegfeldolgozás területén. Szabadon hozzáférhető, a lényeges eszközöket magába foglaló, nagyfokú teljesítményjavulásra képes szolgáltatás, amellyel még úgy is jelentős mennyiségű munkaórát tud megspórolni egy kutató, hogy az automatikus átírás végeredménye továbbra is ellenőrzést igényel. Felhasználóbarát kialakításának köszönhetően pedig egyedülálló a maga kategóriájában, szemben azokkal a szövegfeldolgozó alkalmazásokkal, amelyek informatikai és/vagy programozási háttértudást igényelnek a működtetésükhöz, valamint grafikus felülettel sem rendelkeznek.

Bibliográfia

Arany János. „Hivatali iratok 2.: Akadémiai évek (1859–77)”. szerk., jegyz. Gergely Pál. Bp. Akadémiai K. (1964) (Arany János összes művei, 14.)

Digitalisierung & Elektronische Archivierung (DEA). Hozzáférés: 2019. 03. 10.
<https://www.uibk.ac.at/germanistik/einrichtungen/dea.html>

Gergely Pál. „Arany János és az Akadémia.” Bp. MTA Irodalomtudományi Intézete – Akadémiai K. (1957). (Irodalomtörténeti Füzetek, 11.)

Keresztury Dezső. „Csak hangköre más: Arany János 1857–1882.” Bp. Szépirodalmi K. (1987)

Optikai karakterfelismerés. Hozzáférés 2019. 03. 10.
<http://szovegbanyaszat.tydotex.hu/content/PDF/ch+ocr.pdf>

Recognition and Enrichment of Archival Documents. Hozzáférés: 2019. 03. 10.
<https://read.transkribus.eu/about/>

Transkribus. Hozzáférés: 2019. 03. 10. <https://transkribus.eu/Transkribus/>

Transkribus – How to Guides. Hozzáférés: 2019. 03. 10.
https://transkribus.eu/wiki/index.php/How_to_Guides

Transkribus – Questions and Answers. Hozzáférés: 2019. 03. 10.
https://transkribus.eu/wiki/index.php/Questions_and_Answers

Voinovich Géza. „Arany János életrajza 1860–1882.” Bp. Magyar Tudományos Akadémia (1938)

Áttérés a Primo New User Interface használatára a Magyar Tudományos Akadémia Könyvtár és Információs Központban

Haász Antal

Magyar Tudományos Akadémia Könyvtár és Információs Központ

Szakinformatikai Osztály

haasz.antal@konyvtar.mta.hu

After the almost one year test-period, we have started to use the Primo NEW UI at the beginning of this year. The NEW UI is based on the modern web technologies. The study is about the main steps of migration: customizing of the opening page, and operational changes to the new library services.

Keywords: Primo, discovery system, web technology, API

Bevezetés

A Primo New UI bevezetésére – mintegy egy éves előkészítő fázis után – 2019 februárjában került sor az MTA KIK-ben.

A Primo használata nem volt előzmény nélküli könyvtárunkban, hiszen annak „classic” felülete már 2015 novembere óta elérhető volt felhasználóink számára. A rendszer általános jellemzőiről illetve a 2015-ös bevezetés tanulságairól már egy korábbi tanulmányban beszámoltunk,¹ ezért ezekről itt most nem szólunk.

A továbbiakban részletesen ismertetjük az új verzióra (New UI) történő átállás meghatározó pontjait. Bemutatjuk a nyitóoldal – modern webes technológiákon alapuló – egyedi konfigurálásának főbb lépéseit, valamint azokat a működéssel kapcsolatos fejlesztéseket, amelyekkel a hatékonyabb rendszerhasználatot igyekeztünk elősegíteni.

1. A nyitóoldal kialakítása

A munka kezdetekor az Ex Libris által biztosított, standard, minimális szerkezeti- és dizájn- elemeket tartalmazó nyitóoldal állt rendelkezésünkre. A kitűzött cél az volt, hogy ezt a kezdetleges és „semleges” felületet egy egyedi, az intézményre jellemző szín- és formavilággal, valamint a felhasználói igényeket eredményesen kiszolgáló, dinamikusan működő szolgáltatásokkal töltsük meg.

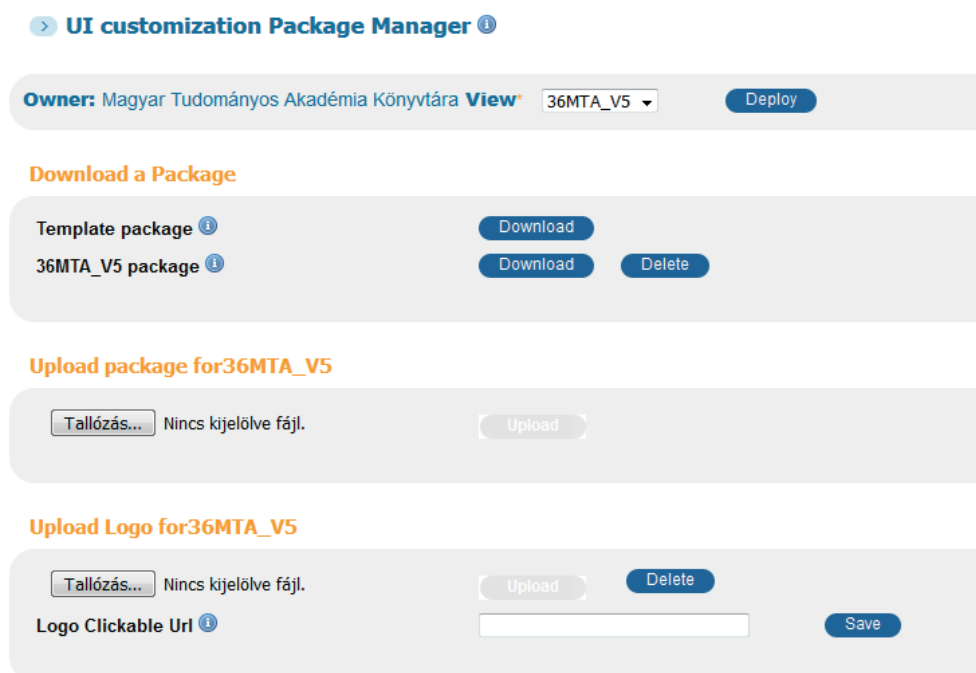
A munka elsősorban a szakinformatikusok és – a felhasználói szokásokat és igényeket leginkább ismerő – olvasószolgálati munkatársak együttműködésében folyt. Folyamatos volt az egyeztetés a két osztály munkatársai között: ötletbörzék,

¹ Haász, Antal. 2017. "A Primo használata a Magyar Tudományos Akadémia Könyvtár és Információs Központban." In *Hagyományok és kihívások V. Múlt és Jövő*, 167–178. doi: https://doi.org/10.21862/HagyKihiv_2016.167



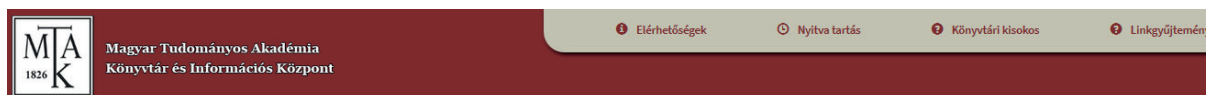
látványtervek elkészítése és véleményezése, hibalisták összeállítása, majd azok alapján történő javítások jellemezték ezt az időszakot.

A nyitóoldal konfigurálására a Primo Back Office felületén elérhető Customization Manager ad lehetőséget.



1. ábra A Customization Manager a Primo Back Office-ban

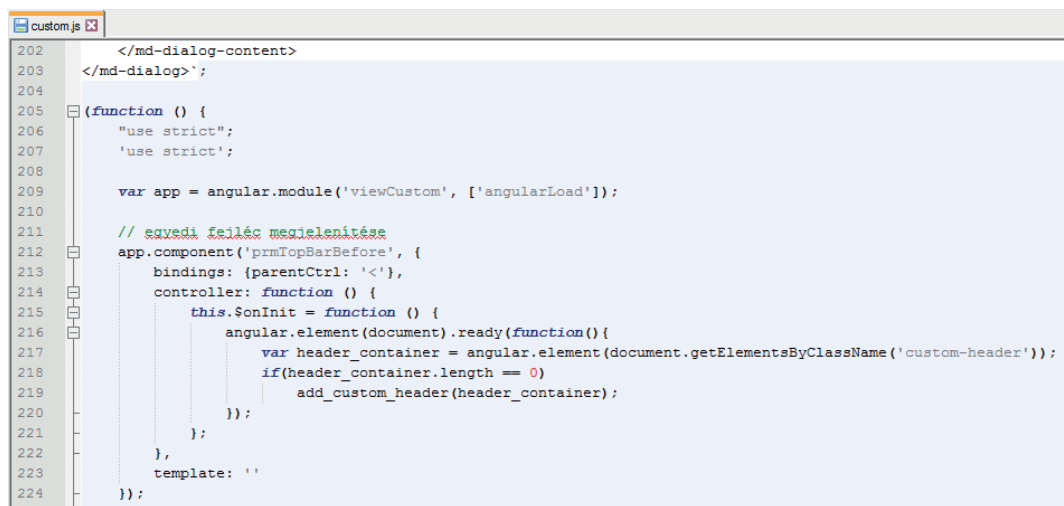
A Customization Manager-rel lehet az egyedi szerkesztéshez szükséges konfigurációs fájlokat és stíluslapokat letölteni, illetve a módosítás után azokat visszatölteni. Az úgynevezett Customization Package négy mappát tartalmaz. A HTML mappában érhetőek el nyelvenként a törzsoldal szerkezeti elemei (pl.: home_hu_HU.html), a CSS mappában található custom1.css fájl használatával lehet az egyedi szín- és formavilágot kialakítani. Az IMG mappába szükséges feltölteni a nyitóoldalon, illetve keresési felületeken használni kívánt képi elemeket (pl. library-logo.png), továbbá a JS mappában található custom.js fájlban lehet JavaScript kódokat rögzíteni bonyolultabb webes megoldások alkalmazására. Elsőként az egyedi fejléc kialakításához fogtunk hozzá.



2. ábra A Primo egyedi fejléce az MTA KIK-ben

A munkát megkönnyítette, hogy a GitHub-on egy számunkra megfelelő menü-szerkezetet leíró alap-kódkészlet volt elérhető.²

A letöltött kód-részletet a custom.js fájlban kellett beilleszteni. Ugyanitt kellett létrehozni – html-kóddal – a megjelenítendő menüpontokat is. A könyvtári logo feltöltésével (IMG mappa) majd a könyvtári honlappal korreláló színvilág kialakításával (custom1.css) véglegesítettük az új verzió egyedi fejlécét.



```
202 </md-dialog-content>
203 </md-dialog>;
204
205 (function () {
206     "use strict";
207     'use strict';
208
209     var app = angular.module('viewCustom', ['angularLoad']);
210
211     // egyedi fejléc megjelenítése
212     app.component('prmTopBarBefore', {
213         bindings: {parentCtrl: '<'},
214         controller: function () {
215             this.$onInit = function () {
216                 angular.element(document).ready(function() {
217                     var header_container = angular.element(document.getElementsByClassName('custom-header'));
218                     if(header_container.length == 0)
219                         add_custom_header(header_container);
220                 });
221             };
222         },
223         template: ''
224     });
```

3. ábra A GitHub-ról letöltött kódrészlet a custom.js fájlban

Fontos szerepet szántunk az oldalon a hírblokknak, hiszen felhasználóink ezen keresztül tájékozódhatnak a könyvtárral kapcsolatos aktualitásokról.³

A „classic” felületre éveken keresztül „manuálisan”, egyenként megszerkesztve tettük fel a híreket. Ennek kiváltására kerestünk egy olyan új módszert, amely a meglévő hírforrások (könyvtári honlap, könyvtári hírlevél) alapján automatikus hírmegjelenést eredményez. A lehetőségek áttekintése után a megoldást a Primo és a könyvtári hírlevél (WordPress) közötti RESTful API-n alapuló kommunikációban találtuk meg. A működés alapfeltétele volt, hogy a WordPress oldalon telepítve és engedélyezve legyen a REST API. A WordPress oldal kezelőjének továbbá feladata, hogy a Primo-n is megjeleníteni kívánt hírt egy meghatározott kategóriába (Primo) sorolja. A működés során a Primo-ból egy úgynevezett HTTP GET kérés indul a Hírlevél felé, amely úgy van megírva, hogy az utolsó három, Primo-kategóriába sorolt hírt vegye át, majd jelenítse meg a kezdőoldal hír szekciójában. Mindezek eredményeképpen a Primo-ban folyamatos az automatikus hírátvétel a könyvtár hírleveléből.

² Hozzáférés: 2019.06.24.,

<https://github.com/pdxlibrary/Custom-Header-for-Primo-Explore-Primo-NUI->

³ Mivel a könyvtári honlapon (<https://konyvtar.mta.hu/>) is folyamatos a könyvtárral kapcsolatos híradás, ezért megegyezés született arról, hogy a Primo-n elsősorban a katalógussal, az új könyvtári beszerzésekkel, adatbázisokkal kapcsolatos híreket szerepeltetjük, illetve csak a fontosabb nyitva tartási (pl. nyári szünet) információkat közöljük.



Hírek, aktualitások

Keleti Gyűjteményünk nyári zárva tartása

A Keleti Gyűjtemény vízvezetékcsere miatt 2019. július 4-től (csütörtöktől) zárva lesz. Nyitás augusztus 20. után, a pontos dátumot később tudjuk közzétenni. A kellemetlenségért elnézést kérünk, olvasóinknak, látogatóinknak kellemes nyarat kívánunk!

[Tovább a hírhez ...](#)

Új beszerzésű külföldi könyveink 2019/5.

2019/5-ös lista

[Tovább a hírhez ...](#)

4. ábra Hírmegjelenítés a Primo-n

Könyvtárunkban folyóiratokra és különgyűjteményi dokumentumokra előkészítési igényeket lehet leadni. Igyekeztünk az igénylés folyamatát minél inkább leegyszerűsíteni. Ennek érdekében könnyen kezelhető űrlapokat terveztünk, amelyek révén olvasóink már a nyitóoldaltól intézhetik lefoglalási kéréseiket.

Az űrlap kitöltése és elküldése után, a Primo HTTP POST kérést küld egy másik URL-re (work.mta.hu) JSON formátumban. A megszólított szerveren a kérés hatására lefutnak a megfelelő php-scriptek, amelyek feldolgozzák a beérkező adatokat és e-mail értesítéseket küldenek ki egyrészt az olvasószolgálati munkatársaknak az új kérésről, másrészt a felhasználónak arról, hogy a kérés teljesítése folyamatban van.

Kézirattár – Dokumentum-előkészítés igénylőlap BEZÁR

Olvasó neve:

Olvasójegy/kölcsönzőjegy száma:

Olvasó email címe:

Előkészítés határideje  19/07/02 

Raktári jelzet, leírás (szerző/cím/év):

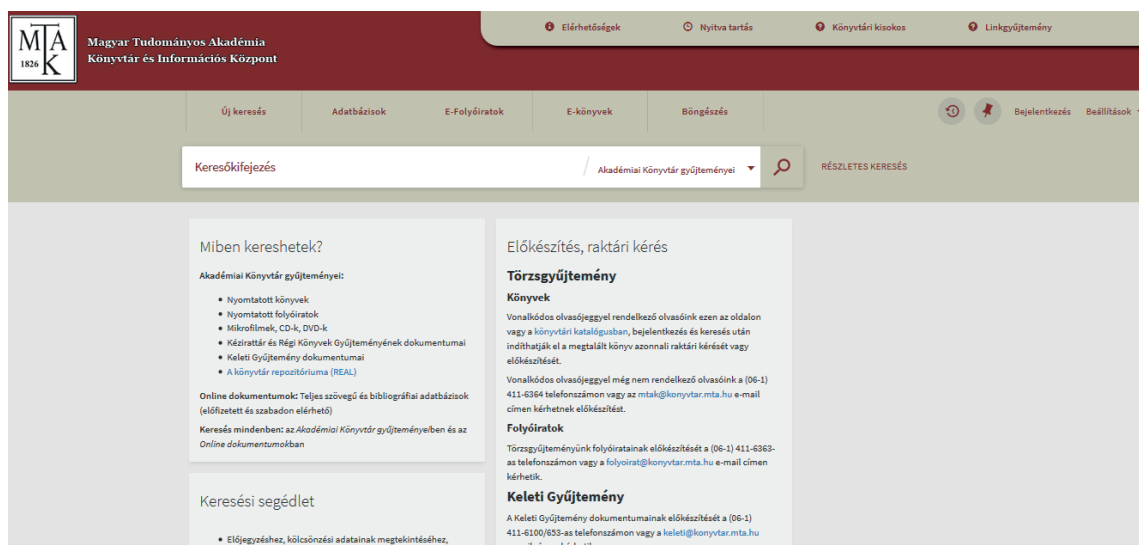
☐ Az űrlap kitöltése és elküldése előtt elolvastam és tudomásul vettem a Kézirattár és Régi Könyvek Gyűjteménye használati szabályzatát.

KÜLD

A folytatáshoz el kell fogadnia a használati feltételeket.

5. ábra Kézirattári dokumentum-igénylő űrlap

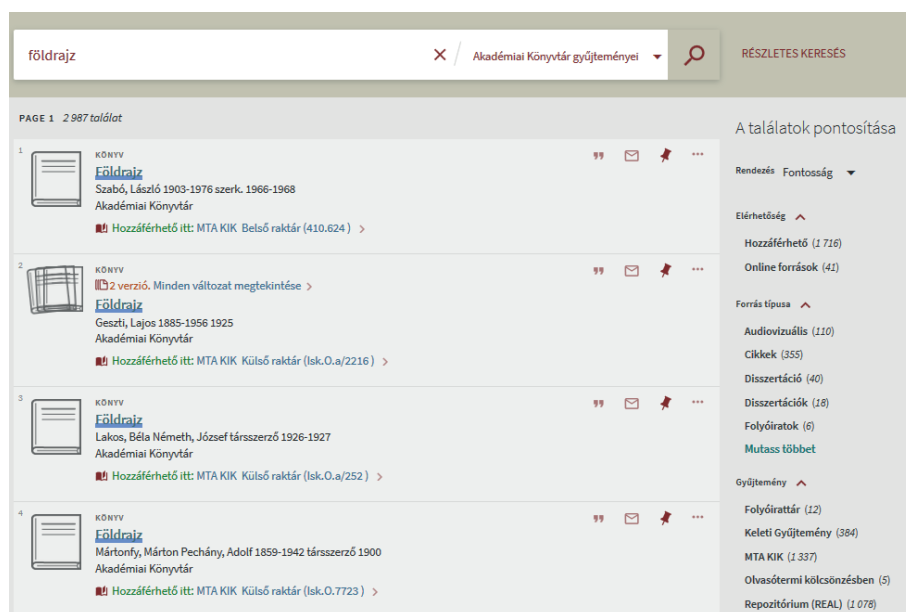
A fent bemutatott fejlesztések eredményeképpen megszületett a nyitóoldal azon verziója, ami most is – néhány kisebb módosítással – használatban van.



6. ábra Részlet a Primo nyitóoldalából

2. A keresés és az adatmegjelenítés fejlesztése

A Primo – mint a discovery rendszerek általában – lehetővé teszi a könyvtár teljes (hagyományos, elektronikus, illetve digitális) gyűjteményének egyidejű lekeresését. Az egyszerű (egymező) keresés mellett lehetőségünk van részletes keresési adatok megadására, valamint böngészésre is. Kereshetünk csak valamely rész adatbázisban (scope)⁴ vagy választhatjuk az összes gyűjteményben való keresést.



7. ábra A kereső oldal, a találati halmaz és a facetták a Primo-ban

⁴ Az MTA KIK Primo-jában két keresőcsoportot hoztunk létre: az Akadémiai Könyvtár gyűjteményei (ebben benne vannak a repozitóriumi dokumentumok is), valamint az Online dokumentumok csoportokat.



A keresés lefolytatása után a Primo úgynevezett facetták alkalmazásával a találatokat rendezi, csoportosítja, ami – különösen a nagyszámú találati lista esetén – jelentősen növelheti a keresés hatékonyságát.

Tapasztalataink alapján a felhasználóink számára fontos szempont a dokumentum típusa alapján történő szűkítési lehetőség. Ezért a New UI elindításakor egy új, a dokumentum típusára utaló facetta bevezetését határoztuk el, amelyet „Forrás típusa” megjelöléssel használunk.

Könyvtári integrált rendszerünkben (ALEPH)⁵ 2017 márciusában tértünk át egy új dokumentumtípológia⁶ használatára. Az addig használt 7 formátum mellett, azok differenciált bontásával további 23 formátumot kezdtünk el használni.

Az új facetta működését az átdolgozott dokumentumtípológia alapján kívántuk felépíteni.

Aleph-ben alapértelmezetten az FMT-mezőben jelenik meg egy adott bibliográfiai-rekord vonatkozásában a típus-információ. Emellett a rendszer lehetőséget kínál arra is, hogy ugyanezt az információt egy másik, úgynevezett TYP virtuális mezőben is tároljuk. A TYP mező a rekordok LDR, illetve 008-as mezőinek értéke alapján épül fel.⁷

A TYP-mező a rendszer-beállítások alapján az Aleph-rekordok publikálásakor átkerül a Primo-ba, így az MARC21/PNX⁸ mezőmegfeleltetés alapjává tehető.

8. ábra Mezőmegfeleltetés az LV (Levelezés) dokumentumtípusra

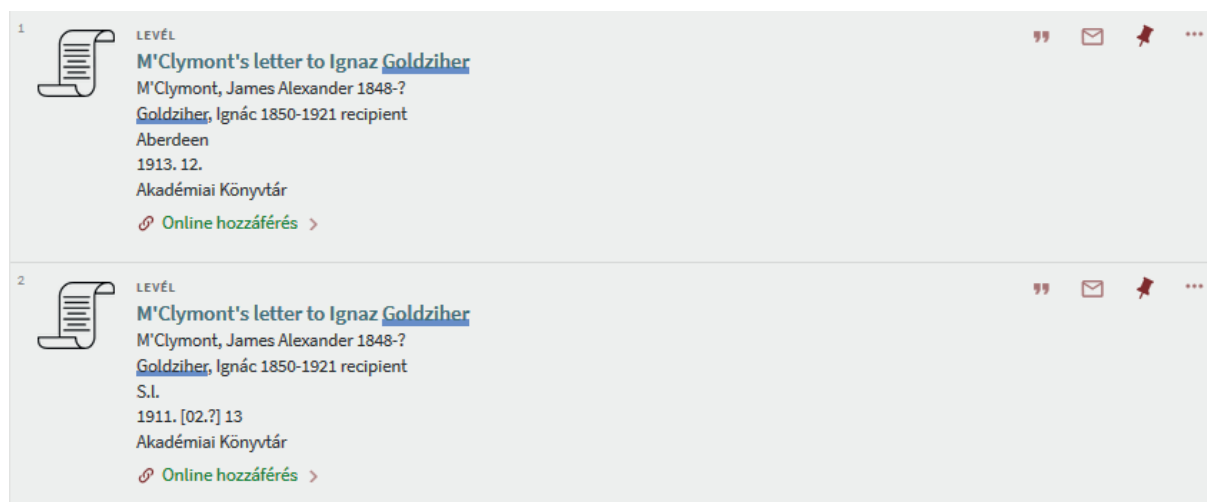
5 Jelenleg használt GUI, illetve Server verzió: 23.2.1

6 Erről lásd bővebben: Gyuricza, Andrea és Haász, Antal. 2018. "Dokumentumtípológia." In *A MARC21 szerinti katalogizálás bevezetése az MTA Könyvtár és Információs Központban*, 75–80.

7 A mező-értékek összerendelése és ezáltal az egyes típusok meghatározása az Aleph-ben a `tab_type_config.hun` táblában történik.

8 Primo Normalized XML

Minden típushoz ikon is rendelhető, amely a szöveges megjelölés mellett vizuálisan is megjeleníti az adott tétel formátumát.



9. ábra Dokumentumtípus-megjelenítés a Primo-ban

A leggyakrabban használt típusokra már elvégeztük a megfeleltetéseket, így azok már láthatóak a Primo-ban. Terveink szerint a teljes dokumentumtípológia átültetésre fog kerülni a Primo-ban.

Az adatmegjelenítés kapcsán mind a rövid, mind a részletes nézetben módosításokat hajtottunk végre.

A találati listában zavaróan egymásra torlódtak a mű adatai, megnehezítve az áttekintést. A probléma megoldása érdekében minden adatcsoportot külön sorba helyeztünk, amit úgy értünk el, hogy Back Office vonatkozó részében mezőhatárolónak (delimiter) a
 html-tag-et vettük fel.

A Részletek fülön jó néhány, tájékoztatási szempontból fontos adattartalom nem volt elérhető. Ezért – konzultálva az olvasószolgálatos és feldolgozó kollégákkal – megkezdtük ezen adatok pótlását a felületen, amelyet szintén a már fent említett mezőmegfeleltetéssel végeztünk. A munka során megjegyzés (például 500, 502, 506, 535, 540), illetve folyóiratok állományadatait megjelenítő mezőket (876, 877) tettünk láthatóvá itt.

Az adatmegjelenítés konfigurálása nem lezárt folyamat. Az igényeknek megfelelően bármikor vehetünk ki-, illetve tehetünk be adatcsoportot a felhasználói felületekre.

3. Raktári kérés a Primo-ban

A Primo-ban az előjegyzés funkció használatával valósítottuk meg az online raktári kérés lehetőségét. A klasszikus értelemben vett előjegyzési funkció⁹

⁹ Kölcsönzésben lévő dokumentumra – általában külön díj ellenében – leadható foglalás. A dokumentum visszaérkezéséről a könyvtár értesíti a foglalást kezdeményező olvasót és megadott ideig félreteszi számára a kért dokumentumot.



használata könyvtári gyakorlatunkban nincs (és nem is volt) jelen (csak raktárban lévő dokumentumra lehet előjegyzést tenni, kölcsönzésben lévőre nem). Ezért esetünkben helyesebb inkább előkészítési vagy lefoglalási rendszerről beszélni. Ennek megfelelően kellett a Primo-ban is kialakítani az előjegyzési rendszert, módosítva az eredeti beállításokat.

A kölcsönzésben lévő dokumentumok előjegyzésének blokkolását Aleph beállításokkal tudtuk megoldani a Primo-ban is.¹⁰ Egy paraméter-tábla egyik ellenőrzési procedúrájával¹¹ – a megjelölt olvasói státuszok esetében – a kölcsönzéseket tartós kölcsönzéssé lehet minősíteni, amelyekre nem lehet előjegyzést tenni. Ebben a táblában az összes olvasói státuszt felsoroltuk, így gyakorlatilag – a célnak megfelelően – az összes kölcsönzött dokumentum vonatkozásában blokkoltuk az előjegyzési lehetőséget.

Ezek után a megfelelő előkészítési űrlap kialakítása következett. Az átvétel helyének kiválasztása, az érdeklődés maximális dátumának megadása, illetve a helyszám (amennyiben a kérő az olvasóteremben tartózkodik) kitöltése után véglegesíthető (kérés küldése) az űrlap. Opcionálisan megjegyzési szöveg is csatolható a kéréshez. Az igénylés teljesítése egy-két munkanapot vesz igénybe (attól függően, hogy külső vagy belső raktári kérésről van szó).

10. ábra Az előjegyzési űrlap a Primo-ban

¹⁰ A Primo a kölcsönzési – ezen belül az előjegyzési – funkciók vonatkozásában a könyvtári rendszer (esetünkben az Aleph) beállításait használja.

¹¹ Ez a tab_hold_request tábla és a check_hold_request_w procedúra.

Az olvasóknak szánt tájékoztató szövegeket egységesítettük, minden képernyőképen a „Raktári kérés” megnevezéssel jelöltük az előjegyzési funkciót. Abban az esetben, ha az adott műre nem lehet előjegyzést tenni, igyekeztünk minél lényegre törőbben tájékoztatni a felhasználóinkat a továbblépés lehetőségéről („Ez a dokumentum valószínűleg különgyűjteményben található. Forduljon a könyvtárhoz.”).

4. Összegzés

A tesztelési időszak, valamint a közel fél éves éles működés tapasztalatai alapján megállapítható, hogy az új verzió – elsősorban rezponzivitásából eredően – sokkal több lehetőséget rejt a felhasználóbarát felületek kialakítása terén. Fontos jellemzője az egyedi konfigurálhatóság, amely számos speciális szolgáltatás elindítását teszi lehetővé.

A rendszerben rejlő lehetőségek kiaknázása érdekében elengedhetetlen a dokumentáció áttekintése, a könyvtári rendszerrel való kapcsolatok feltérképezése.

Végezetül szeretném megköszönni Lovas Bálint informatikus kollégámnak a nyitóoldal kialakításában végzett kitartó és ötletgazdag munkáját, valamint azokat az értékes információkat, amikkel segítette a tanulmány megírását.

Nyílt publikálási szoftverek és platformok

Holl András

MTA Könyvtár és Információs Központ

holl.andras@konyvtar.mta.hu

ORCID: 0000-0002-6873-3425

Bilicsi Erika

MTA Könyvtár és Információs Központ

bilicsi.erika@konyvtar.mta.hu

ORCID: 0000-0003-2651-2955

Open publication software and platforms

The use of open journal publishing software (mainly the Open Journal Systems and its „siblings”) is spreading in Hungary. There are individual installations and platforms which serve several journals. We survey the present usage, and deal with the role of such software platforms for libraries and academia. With the availability of open source, free content management systems developed for scholarly communication, which offer compliance with standard aggregation protocols, small, independent scholarly journals might become more visible.

Keywords: open access, open source, Open Journal Systems (OJS)

Bevezető

Tanulmányunkban a hazánkban is egyre gyakrabban alkalmazott,¹ nyílt forráskódú szoftvereken alapuló, szabványos protokollokat használó, jellemzően nyílt hozzáférésű közleményeket publikáló platformokkal foglalkozunk, elsősorban könyvtárosi szempontból.

Figyelmünket az Open Journal Systems-re² (OJS) összpontosítjuk, ám kitérünk „testvéreire” – az Open Conference Systems-re³ (OCS) és az Open Monograph Press-re⁴ (OMP) – és „rokonaira” is. Kidomborítjuk azt a szerepet, amit ezek a rendszerek az információ-keresésben és a bibliográfiák készítésében játszhatnak. Példákat mutatunk be a DOI-regisztráció, a Magyar Tudományos Művek Tára (MTMT) számára történő adatátadás és a repozitóriumi kapcsolat megvalósításáról.

Nem csupán a nagy kiadóvállalatok jelentetnek meg tudományos műveket – fontos közlemények jelennek meg nem kereskedelmi alapon működő szervezetek (tudományos társaságok, egyetemek, kutatóintézetek kiadói vagy független kiadók) gondozásában. 2017-ben az MTMT szerint a hazai szerzők cikkeinek közel

1 A Magyarországon működő OJS-ben megjelenő folyóiratok listája, Hozzáférés: 2019.07.02. <http://openaccess.mtak.hu/index.php/103-openjournalsystems/155-ojs-hazai-kiadvanyok>

2 Open Journal Systems, Hozzáférés: 2019.07.02. <https://pkp.sfu.ca/ojs/>

3 Open Conference Systems, Hozzáférés: 2019.07.02. <https://pkp.sfu.ca/ocs/>

4 Open Monograph Press, Hozzáférés: 2019.07.02. <https://pkp.sfu.ca/omp/>

harmada olyan folyóiratban jelent meg, amelyet nem követ sem a Web of Science (WoS), sem a Scopus. A WoS vagy Scopus indexelés, jelentsük ki, nem jelent feltétlenül minőségi ítéletet: a nem világnyelven közlő, nem kellő rendszerességgel megjelenő, kis példányszámú, inkább csak helyi jelentőségű lapok is közölnek értékes cikkeket. A nemzeti vagy regionális jelentőségű kutatások esetében nehéz a közleményeket nemzetközi folyóiratokban elhelyezni.

Az ismertetett nyílt forráskódú platformok – bár alkalmazzák ezeket egyes kiadványvállalatok⁵ is – jellemzően kis, független (indie) folyóiratokat szolgálnak ki. Az indie kiadványoknak mind a hosszú távú túlélése veszélyeztetett, mind a láthatósága korlátozott (gyakorta a méretük, nyelvük miatt nem kerülnek be a nagy, kereskedelmi alapon működő bibliográfiai adatbázisokba). Különös fontossága van az előbbieknél fogva a repozitóriumi archiválásuknak és a közös keresőkbe való indexelésüknek.

A független folyóiratok kiadói tanszékek, intézetek, tudományos társaságok, helyi kiadók. Ezek a múltban nem élvezhették a nagy, kereskedelmi indexelő adatbázisok jótéteményeit. Cikkeiket csak nemzeti bibliográfiai és archiváló rendszerek – mint a mi MATARKA-nk vagy EPA-nk – vagy szűk szakterületeket lefedő bibliográfiák dolgozták fel. Az ezekben a cikkekben megjelenő hivatkozások feldolgozása azonban már meghaladja a lehetőségeket – mint azt az e téren történt hazai próbálkozások kudarca bizonyítja.

A WoS és a Scopus eredetileg kézi erővel, később már egyre inkább gépileg dolgozta fel az indexelt folyóiratokban megjelenő cikkek adatait. Ez jelentős befektetést igényelt, amit a szolgáltatások nemzetközi értékesítése tett kifizetődővé (magas haszonkulcs, azaz magas árak mellett).

Ma már azonban az informatika lehetővé teszi, hogy a kis, független folyóiratok analitikus feltárása és a bennük közölt idézetek (többségének) feldolgozása is megtörténhessen – automatikusan. Ehhez arra van szükség, hogy a cikkek adatbázis alapú informatikai rendszerben kerüljenek feldolgozásra és a hivatkozások DOI azonosítókkal legyenek ellátva. A tudományos kommunikáció modern technikai infrastruktúrájának működtetése (keresők/aggregátorok üzemeltetése, hivatkozások gyűjtése) is pénzbe kerül, de lényegesen olcsóbban megoldható, mint a korábbiakban (pl. Web of Science, Science Citation Index). Ezért tökeszegény szakmai szervezetek is nyújthatnak ilyen szolgáltatásokat, a kereskedelmi alapon működő, ilyen technológiákat alkalmazó vállalkozások pedig olcsóbban tudják (és kénytelenek) adni szolgáltatásaikat. A tudományos kommunikáció világában az új lehetőségek új szolgáltatások, eszközök burjánzásához vezetnek és esélyeket adnak az indie folyóiratoknak is.

5 Az Akadémiai Kiadó az OJS kézirat-beküldési funkcióit használja.



Publikációs platformok – az Open Journal Systems

A kiadók, szerkesztőségek számos díj fizetés ellenében használható⁶ és nyílt forráskódú⁷ publikációs platform közül választhatnak.

A Public Knowledge Project (PKP) szabad forráskódú publikációs platform szoftverei a folyóirat-szerkesztőségi működést, publikálást és adminisztrációt automatizáló Open Journal Systems, a konferencia-szervezést és a konferenciakiadvány közreadását szolgáló Open Conference Systems, valamint a könyvkiadást támogató Open Monograph Press.⁸ Az említett publikációs platformok mellett a PKP egy aggregátor-szoftvert is közread: ez az Open Harvester Systems.⁹

A PKP szoftverei a kis létszámú szerkesztőségek részére tökéletesen megfelelő szolgáltatásokat és konfigurálási lehetőségeket nyújtanak. Telepítésük és üzemeltetésük viszonylag egyszerű, de a tapasztalatok azt mutatják, hogy az üzemeltető informatikus mellett mindenképpen szükség van egy informatikai érdeklődéssel rendelkező szerkesztőségi tagra, vagy a szerkesztőséget segítő helyi könyvtárosra, informatikusra, aki az ún. folyóiratmenedzseri feladatokat ellátja. A folyóiratmenedzser jogosult az OJS-ben a weboldal megjelenését testreszabni, a lap munkafolyamatának megfelelő beállításokat (pl. rovatok, melléklet típusok, határidők, lektorálási űrlap, alapértelmezett e-mailek szövege) megtenni, de ez a személy hivatott arra is, hogy segítse a szerkesztőségi tagok és a szerzők munkáját, ha technikai nehézségbe ütköznek a szoftver használata során. Ez pedig garantáltan előfordul, hiszen egy igen sokrétű munkafolyamat támogatására hivatott szoftverről van szó, ami a cikkek benyújtásától a lektorálási, tördelőszerkesztési munkafolyamatokon át az online megjelentetést, de még az utólagos adminisztrációs feladatokat is (mint például a DOI regisztráció) automatizálhatóvá teszi.

A nyílt, adatbázis alapú, szabványokat támogató publikációs platformok nem elsősorban a kiadók munkáját könnyítik meg. Nem mindig könnyű megtanulni ezen szoftverek használatát az e-mail és a kockás füzet után. Nehéz elviselni a szoftverek által megkívánt szabatossgot, elvágólagosságot. Leginkább abban jelentkezik az előnyük, hogy alkalmazásukkal könnyebb az információáramlás az adatbázisok között – a könyvtárosok és persze, végül az olvasók a hasznélvezők.

Atudományos kommunikáció nemzetközi hálózatába való bekerülést nagyban segíti az adatbázis alapú publikációs szoftverek használata. Ezek lehetőséget adnak az automatikus információátadásra, az új technológiájú információ-aggregátorokhoz

6 például: [HighWire Press](#), [Atypon](#), [ubiJournal](#), [Scholastica](#), [ManuscriptManager](#), [JMS](#)

7 például: [OJS](#), [Libero](#), [Janeway](#), [Ambra](#)

8 Az OJS-ről és az MTA KIK-ben való alkalmazásáról Payer Barbara tartott előadást a Networkshopon 2017-ben. Payer Barbara: A megújuló online folyóiratkiadás. Tudásmegosztás, együttműködés, automatizálás. Networkshop 2017.; TMT 64. (2017) 11. 548. Hozzáférés: 2019.07.02. <http://tmt.omikk.bme.hu/tmt/article/view/1330/10363>

9 Open Harvester Systems, Hozzáférés: 2019.07.02. <https://pkp.sfu.ca/ohs/>

való kapcsolódásra¹⁰. Mindeközben természetesen a hagyományos kereskedelmi adatbázisokba is könnyebb bekerülni a megfelelő technológiák alkalmazásával.

Az alábbiakban áttekintjük a PKP platformjainak néhány olyan tulajdonságát, amelyek lehetővé teszik a bibliográfiai információk áramlását. Ezeket mi elsősorban az OJS esetében alkalmazzuk, de adott esetben működnek a másik két szoftverben (OCS, OMP) is.

Bibliográfiai adatok szűretelése

Mind az aggregáció, mind a különböző rendszerek összeköthetőségének fontos eszköze az OAI-PMH¹¹. Az aggregátorok, mint a BASE, Core vagy éppen a magyar Repozitóriumi Közös kereső¹² (OAIkereso) nem csupán repozitóriumokat, de más OJS-t támogató platformokat is aggregálhatnak. Az OAIkereso jelenleg 55 OJS vagy OCS periodika adatait aratja. Folyóirat publikáló platformot üzemeltet Magyarországon többek között az MTA Könyvtár és Információs Központ, a Szegedi Tudományegyetem, a Debreceni Egyetem, a Pécsi Tudományegyetem, a Nemzeti Közzolgálati Egyetem, valamint a Corvinus Egyetem.

MTMT feltöltés, repozitóriumi archiválás

Mivel az OJS-ben tárolt adatok könnyedén legyűjthetőek az OAI-PMH protokoll segítségével, egy erre épülő, ún. OAIuploader, azaz cikkfeltöltő alkalmazás készült az MTA KIK számára. Az eszköz működésének kulcseleme a megfelelően konfigurált OAI kimenet – ill. természetesen a mezők megfelelő adatokkal való kitöltése.

Az alkalmazásba az MTMT „Kiadók” intézményének adminisztrátorai léphetnek be. Akár az adatbázisukban tárolt összes cikk adatait betölthetik az eszköz segítségével az MTMT-be, de szűkíthetnek az adott folyóirat konkrét évfolyamára, füzetszámára, azon belül akár egy konkrét rovatra is. Az általunk elképzelt ideális munkamenet az lenne, ha a megjelenést követően a lehető legrövidebb időn belül elvégeznék ezt az adatfeltöltést, hiszen ez biztosítaná, hogy a szerző munkásságában az ő közreműködése nélkül vagy legalábbis közreműködésének minimalizálásával kerülhessenek be a cikkek. Azaz a fejlesztés során az egy folyóiratszámra való fókuszálás volt a cél mellett, hogy az adminisztrátorok munkájának egyszerűsítését is szem előtt tartottuk. Ennek érdekében az OAIuploader-ben azokat az adatokat, melyek minden folyóiratszámban megegyezők, tároljuk egy ún. folyóirat űrlapon. Így nem szükséges minden egyes feltöltés megkezdésekor megadni a folyóirat ún. OAI base URL-ét és beállítani az adatfeldolgozási sémát, ill. kiválasztani az MTMT-ben tárolt listából a folyóirat címét, nyelvét és a kapcsolódó tudományágat, hiszen ezek minden cikk esetén megegyező adatok lesznek (pl. a Földtani Közlöny folyóirat vélhetően földtudományi cikkeket közöl, így ez a szakterület beállítható, de a szerző

10 Új nyilvános kereső szolgáltatások: [Dimensions](#), [Lens](#), [1Findr](#), [Scilit](#), [Science Open](#). Akadémiai aggregátorok: [CORE](#), [BASE](#). Háttér infrastruktúrák, technológiák, adatbázisok: DOI: [CrossRef](#), [DataCite](#); OpenAccess információ párosítása DOI-val: [oaDOI](#); [Unpaywall](#); szerzőazonosítás: [ORCID](#); technológiák: [SHARE](#). A felsoroltak között vannak olyanok, amelyeket már valamelyik multi felvásárolt, vannak startupok, vannak közösségi - „szövetkezeti” szervezetek.

11 Protocol for Metadata Harvesting. <https://www.openarchives.org/pmh/>

12 <http://oaikereso.sztaki.hu/kereso/index.php>



finomíthatja, ha a cikk bekerült a munkásságába). A működés kulcseleme a séma beállítása, ami az adott folyóiratból kinyert adatok feldolgozásának szabályozására szolgál. Az adatokat a 'source' mező tartalmazza, az egyes adatok felismeréséhez egy ún. sémát kell megadni az alkalmazásnak.

Segítség a séma megadásához

Séma:

Adott OAI repozitóriumhoz tartozó adatok feldolgozására szolgál.
Az adatokat a lekért rekord source mezője tartalmazza.
Itt megadhatja a source mező kulcsszavak általi leírását.

A kulcsszavak az értékek helyét jelölik.

PUB: folyóiratneve
VOL: évfolyam
NUM: füzet
DATE: megjelenés éve
RAN: kezdő és záró oldalszám

Figyeljen a kulcsszavakat közötti karakterek pontos megadására!

Példa:

Az alábbi source mező esetén:

Dissertationes **A**rchaeologicae; Ser. **3**, No. **3** (**2015**); **191-202**

Az elkészített séma:

PUB; Ser. **VOL**, No. **NUM** (**DATE**); **RAN**

1. ábra: A metaadat-megfeleltetés sémájának beállítása az OAIuploader szoftverben

A keresést követően megjelennek az adott számban megjelent cikkek. Sajnos duplumszűrés nincsen a rendszerben, ezért az adminisztrátorok feladata, hogy ellenőrizzék, hogy az MTMT-ben szerepel-e esetleg már a feldolgozandó füzet-számban megjelent cikk. Ha igen, akkor azt nem szabad feltölteniük, ill. ha mégis feltöltik, azt az MTMT-ben duplumösszevonással kezelni kell. Miután kijelölték a feltöltendő cikkeket, a többit már nem viszi tovább a következő munkafolyamatba a rendszer. A következő oldalon megjelennek a folyóirat űrlapján beállított fix adatok, amiken még lehet módosítani, ezek minden cikk adatlapjára be fognak kerülni. Ezután pedig cikkenkénti szerkesztésre van lehetőség egy MTMT-hez hasonló felületen, ahol az egyes cikkek adatai javíthatóak, ha erre szükség van, ill. gazdagíthatóak olyan adatokkal, amiket a folyóirat platform nem tárol (pl. szerző egységesített neve az MTMT-ben, affiliáció). Ezután indítható a feltöltés, ami az MTMT-be és a REAL-ba is feltölti az adatokat, a repozitóriumba a teljes szövegű PDF-fel együtt, valamint az ott létrejött linket beírja az MTMT-be az OA státusszal együtt. A DOI is bekerül az MTMT-be, tehát előbb érdemes azt regisztrálni.

Az alkalmazás bármilyen, megfelelően konfigurált OAI kimenetű forrást támogat, nem csak az OJS platformon megjelenő lapok feltöltését.

További adatcserét támogató funkciók

A cikkeiket DOI azonosítóval ellátó folyóiratok egy ügynökség (leginkább a CrossRef) adatbázisába is feltöltenek az azonosító regisztrációjakor metaadatokat. Ezek az adatbázisok is egyre inkább szolgálják a cikkek láthatóságának növelését. Az OJS DOI pluginja segíti a CrossRef és DataCite azonosítók regisztrálását.

Az OJS támogatja az ORCID API-n keresztül hitelesített azonosító regisztrációt. Az ORCID azonosítók pedig továbbkerülhetnek a CrossRef DOI metaadat-tárába.

Az említetteken túl az OJS bővítmények léteznek a PubMed, a DOAJ, a COUNTER¹³, valamint a LOCKSS¹⁴ adatcsere támogatására. CrossRef-tag folyóiratok számára rendelkezésre állnak kapcsolódási lehetőségek a hivatkozáskövetés (cited by), plágiumellenőrzés és verziókövetés támogatására.

Folyóirat- és konferencia hosting az MTA Könyvtár és Információs Központban

Jelenleg 23 folyóirattal vagyunk kapcsolatban, hat folyóirat jelenik meg az OJS platformunkon. Van olyan folyóirat, amelyik a folyamatszervezési funkciót használja, de más (saját) felületen publikál. Mindezek közül összesen ketten használják a teljes funkcionalitást.

Az MTA KIK a korábban indított OJS folyóirat-platformhoz hasonlóan OCS platformot létesített. Ez a platform ad otthont a Hungarnet Networkshop konferenciasorozat kiadványa számára – de a konferencia bonyolítása jelenleg nem ebben a rendszerben történik. Sajnos, az OCS szoftver új verziójának megjelenése évek óta várat magára – ez nehézségeket okoz a platformunk működtetésében.

Más hasonló szoftverek, szolgáltatások folyóirat-kiadáshoz és konferenciaszervezéshez

Bár a független kiadók körében ez a legelterjedtebb, nem csupán a PKP szoftverei állnak rendelkezésükre. Nem is a szoftver gyártója a lényeg, hanem az, hogy a bibliográfiai adatok adatbázisban – vagy más módon –, gépileg hozzáférhetőek legyenek, a platform támogassa a szabványos protokollokat (OAI-PMH, SWORD), valamint legyenek meg a kapcsolódásai a publikációs ökoszisztéma fontos szereplőinek adatbázisaihoz.

Példaként említhetjük a konferenciaszervezés terén az EasyChair-t (NWS is), ez egy ingyenes, web alapú szolgáltatás, de nem nyílt szoftverrel. Az OpenConf ingyenes közösségi verzióval rendelkezik, letölthető, de felhő alapon is használható.

13 COUNTER – felhasználási statisztikák szabványosítása: <https://www.projectcounter.org/>

14 Loss of Copies Keep Stuff Safe - <https://www.lockss.org/>



A folyóirat kiadói rendszerek közül a DpubS-ot említjük, valamint a Digirati/eLife Libero felhőszolgáltatást (publishing cloud), de számos további rendszer is létezik¹⁵.

Természetesen egyedileg fejlesztett, nem kifejezetten folyóiratok számára készült és általános nyílt forráskódú tartalommenedzsment (CMS) rendszereken alapuló folyóirat-honlapokhoz is fejleszthető OAI-PMH és SWORD kapcsolódási lehetőség – de megfontolandó, hogy ehelyett/inkább az elterjedt és jó képességű célszoftvereket alkalmazzuk.

Felhívjuk a hazai *indie* szakfolyóiratok kiadóit az együttműködésre. Működünk együtt a közös kereső alkalmazásában, az OJS Magyar Felhasználói Csoport működtetésében, egy esetleges jövőbeli idézettséggyűjtés kialakításában!

15 További szoftverek, platformok:
http://edutechwiki.unige.ch/en/Journal_management_software;
https://en.wikipedia.org/wiki/Category:Academic_journal_online_publishing_platforms

Új könyvtári fogalmak gyakorló könyvtárosok szemszögéből

Hubay Miklós
Petőfi Irodalmi Múzeum
hubaym@pim.hu
ORCID: 0000-0002-3241-2988

Gyuricza Andrea
Magyar Tudományos Akadémia Könyvtár és Információs Központ
gyuricza.andrea@konyvtar.mta.hu
ORCID: 0000-0003-3960-5499

As the landscape of libraries changes nowadays, there are more and more new concepts, which are often defined with other, incomprehensible terms in scholarly context. A new trend of library cataloging has appeared, that broke up with the 50-years-old traditions, and besides the simple registering of bibliographic data, it is necessary to put these data in a broader context, in a network. In order to accomplish this task, some of the experimenters use the good old solutions, while others define new theoretical borders. The definitions in the library journals sound very well, while the practice-centered approach is missing.

Now we clear the meaning of some of these library and IT science terms (LRM, discovery, semantic web, MARC21, FRBR, RDA, entity, BIBFRAME, URI, schema.org etc.), so the patron service librarians and also the catalogers can understand the transformation of LIS, can interpret the connections, correlations, and the main goals.

Keywords: library cataloging, entity, MARC21, RDA, BIBFRAME

A könyvtári feldolgozási munkafolyamat lényeges, a hiteles információszolgáltatás szempontjából legfontosabb állomása a bibliográfiai leírások visszakereshetőségét biztosító hozzáférési pontok megállapítása. Ezek az ún. authority adatok az érvényben lévő szabályozás szerint a személy- és testületi nevek, a földrajzi nevek, valamint a címek; ezeken felül pedig különféle tárgyszavakat, tezauruszokat is használunk. Az integrált könyvtári rendszerekben a legkülönbébb adatokat tartalmazó authority rekordok emelhetők be a leíró rekordokba, amelyek összességét authority file-nak hívják. Napjainkban az authority file már nem csupán az egyes rendszerekben, hanem az intézményeken kívül, azoktól függetlenül is létezhet – átjárást teremtve a gyűjtemények között – az ilyen szolgáltatásokat névtérnek nevezzük.

Ezek az állományok a leggyakrabban ugyancsak személynevekből (pl. VIAF, *Virtual Internet Authority File*), ULAN - *Union List of Artist Names* stb.), földrajzi nevekből (pl. GeoNames), testületi nevekből stb. épülnek, de természetesen tárgyi névterek, köznévterek is léteznek (pl. Getty). A személyek, kutatók kétséget kizáró azonosítására mindazonáltal több lehetőség is kínálkozik a VIAF mellett. Ilyen például az ISNI (*International Standard Name Identifier*), vagy az ORCID (*Open Researcher and Contributor ID*) azonosítók. A dokumentumok körében az ISBN-



azonosítás mellett új lehetőség az elektronikus formátumoknál a DOI (*Digital Object Identifier*) azonosító. Az információ- és webtechnológia fejlődésének köszönhetően napjainkban már magukat a „dolgozat” (kézzelfogható és fogalmi egyaránt) is lehetőségünk van a számítógépes feldolgozás számára egyértelműen azonosítani, ezek az URI (*Universal Resource Identifier*) azonosítók.

A könyvtári katalogizálás – amelyet ma forrásleírásként emleget a szakirodalom – elméleti alapjait az 1990-es évek végéig a Párizsi Alapelvek, gyakorlati instrukcióit pedig az európai ISBD (*International Standard Bibliographic Description*), illetve az angol-amerikai AACR2 (*Anglo-American Cataloging Rules*) szabályzatok adták. A XX. század utolsó éveiben azonban, a megsokasodó dokumentumtípusok és újfajta információhordozók megjelenése szükségessé tette, hogy a szakemberek elgondolkodjanak a bibliográfiai kontrollról és jövőjéről, amelynek egy új, általános elveket tartalmazó dokumentum megszületését köszönhetjük (ICP, *International Cataloging Principles*). Megvizsgálták azt is, hogy az információkat visszakereső használó mit vár az egyes rekordoktól, melyek azok a lényeges tematikai egységek, amelyekről tájékozódni szeretne. Ez a használóközpontú, a rekord funkcióit előtérbe helyező elemzés adta az alapját a katalogizálás új, entitásokra épülő elméletének, amelyről 1997-ben olvashattunk elsőként – ez volt az FRBR (*Functional Requirements for Bibliographic Records*), később az FRAD (*Functional Requirements for Authority Data*), majd az FRSAD (*for Subject Authority Data*). Az IFLA ezeket a különálló modelleket 2017-ben egységes szerkezetbe foglalta, megszüntetve a közöttük lévő következetlenségeket (LRM, *Library Reference Model*).

A 2010-es években a funkcionális modellekre alapozva új forrásleírási szabályzatot dolgoztak ki, amely eredetileg az AACR harmadik kiadása lett volna, végül azonban RDA (*Resource Description and Access*) néven látta meg a napvilágot. A gyakorlati megvalósításra azonban csak mostanában kezdenek hangsúlyt fektetni, mivel az informatika jelenlegi eszközei minden eddiginél hatékonyabb implementációt tesznek lehetővé. A közgyűjtemények legelterjedtebb adatcsere-formátumának, a MARC-nak (*Machine-Readable Cataloging*) ugyanis egyre több kritikával kell szembenéznie a világháló és az évről-évre egyre hatékonyabb keresőmotorok megjelenése óta. Ez a tény a már több mint fél évszázada alkalmazott technológiák, szabályok átgondolására, megújítására vagy éppen megkérdőjelezésére indította a szakembereket. Egyesek a MARC lehetőségeit igyekeznek bővíteni, míg mások a világháló előnyeit kihasználó, teljesen új lehetőségeket biztosító adatformátum implementálásán, valamint a keresőoptimalizálás általános elveinek való minél hatékonyabb megfelelésen gondolkodnak.

Az adatok összekapcsolódásán alapuló szemantikusweb-technológia fő építőkövei közé tartoznak a fent említetteken kívül az egyes tudományterületek összefüggéseinek, relációinak formalizált leírását tartalmazó ún. ontológiák, avagy szótárak. A könyvtártudomány területén is számos ilyen ismerünk – pl. a Kongresszusi Könyvtár által fejlesztett BIBFRAME (*Bibliographic Framework*) –, és napjainkban is egyre-másra jelennek meg újak, a már meglévők pedig fejlődnek, tökéletesednek.

A szótárak segítségével, RDF-tripletek (*Resource Description Framework*) formájában kifejezett adathalmazt azután egy nyilvánosan hozzáférhető tárhelyről (triplestore) bárkinek szolgáltathatjuk, akinek szüksége van rá.

Napjainkban igen erőteljes átalakuláson mennek át azok a felületek is, amelyeken a források feldolgozása, valamint szolgáltatása történik. Úgy hisszük, senkinek nem kell bemutatni a magyarországi piacon jelenlévő integrált könyvtári rendszereket (Aleph, Amicus, Corvina, Huntéka, Szikla, Szirén stb.), vagy a díjmentesen (open-source) hozzáférhető megfelelőiket (pl. a KOHA), és az OPAC (*Online Public Access Catalogue*) fogalma sem kíván bővebb definíciót. E felületek azonban az utóbbi években természetes evolúción mentek keresztül, amelynek fő mozgatórugója a világhálóba való szervezettebb beépülés, valamint az egymástól elkülönülő helyeken létező tudás összegyűjtése, integrálása volt. A letölthető, telepíthető, ún. kliensalapú rendszereket az integrált könyvtári platformok, az OPAC-okat pedig a discovery-szolgáltatások – és az ezek működését hatékonyabbá tévő SFX (*Special Effects*) – váltják fel fokozatosan.

A könyvtár ma már nem elégedhet meg a hagyományos, nyomtatott formátumú dokumentumok szolgáltatásával: tartalmait elektronikus formában, vagy éppen online is elérhetővé kell tennie a felhasználók számára. A vonatkozó szakirodalom zavarbaejtően sok kifejezéssel illeti a XX-XXI. századi szolgáltatási formákat implementáló gyűjteményeket, amelyek értelmezése akár szerzőnként, munkaközösségenként is eltérő lehet. Annyit azonban biztosan elmondhatunk, hogy napjainkban a digitális-digitalizált gyűjtemények építésének és kezelésének kiterjedt eszköztára van – repozitóriumok, adatátviteli protokollok, publikációs lehetőségek, illetve az open access-mozgalom –, amely segítséget nyújt a források feldolgozásában, tárolásában, szolgáltatásában, illetve lehetőséget biztosít az elérhető tartalmak egy felületen történő megjelenítésére, az ún. tartalom-aggregációra.

Végezetül mindenképpen érdemes szót ejteni arról a nagy informatikai átalakulási-megújulási folyamatról, amely napjainkban is zajlik az Országos Széchényi Könyvtárban. Az OKR (*Országos Könyvtári Rendszer*) néven ismert projekt számos területen tűzött ki célokat, illetve határozott meg feladatokat, amelyek megvalósítása párhuzamosan, szakmai munkacsoportokon keresztül zajlik. A legfontosabb részterületek az alábbiak:

- új integrált, országos használatra tervezett könyvtári platform bevezetése, mely lehetővé teszi a hazai közgyűjtemények együttműködését (OKP, Országos Könyvtári Platform);
- a munkafolyamatok szabványosítása, a rendelkezésre álló normatív dokumentumok magyar nyelvre fordítása, kiemelt figyelemmel az egységes szakmai terminológiára;
- nagy teljesítményű digitalizáló műhely kialakítása, ahol állományvédelmi és szolgáltatási célú digitalizálás egyaránt történhet;
- az intézmény általános informatikai infrastruktúrájának fejlesztése;
- kiemelt tartalomfejlesztési részterület, amelynek keretén belül olyan



szolgáltatások valósulnak meg, mint az egy-egy szűkebb témát feldolgozó Tudástárak, vagy a könyvtári feldolgozó munkát és az internethasználók általános tájékozódását jelentősen támogató Magyar Nemzeti Névtér.¹

A projektben a fentiek mellett nagy figyelmet fordítanak a magyar vonatkozású webhelyek archiválására és egy internet-archívum kialakítására is.

Az alábbiakban közöljük a fentebb említett témákhoz kapcsolódó legfontosabb fogalmak bővebb, részletesebb magyarázatát.

BIBFRAME: Könyvtártudományi ontológia, avagy szótár, amely a forrásleírások készítéséhez szükséges legfontosabb relációkat, viszonyokat tartalmazza. Készítői szerint ez a szótár lehet a jövőben a MARC helyett a könyvtárak egységes adatsere-formátuma.

COMPASS: A Magyar Tudományos Akadémia könyvtára által gondozott, országos szintű adatbázis, amely a magyar könyvtárakban elektronikus formában elérhető, kurrens és archív tudományos adatbázisok listáját tartalmazza.

Discovery: Az online könyvtári források (a helyi nyomtatott állomány online katalógusa, helyi és kereskedelmi adatbázisok stb.) egyetlen felületen való kereshetővé tétele egyetlen keresőmező alkalmazásával (vö. Google) A legismertebb discovery-szolgáltatások a Primo, a Summon vagy az EBSCO Discovery Service (EDS), a nyílt hozzáférésű megoldások közül pedig a VuFinddal, valamint a Blacklighttal találkozhatunk a legtöbb közgyűjteménynél.

DOI (Digital Object Identifier): Az online dokumentumok körében használt egyedi azonosító rendszer, biztosítja a dokumentumok megtalálását is.

FRBR, FRAD, FRSD: A bibliográfiai, besorolási és tárgyi besorolási rekordok funkcionális követelményeit tárgyaló elméleti modellek. Tartalmazzák a rekordok használói szempontból lényeges egységeinek (entitások) leírását és ezen egységek közötti kapcsolatokat.

ICP (International Cataloguing Principles): A 2016-ban kiadott alapelv-nyilatkozat felváltotta és határozottan kibővítette a „Párizsi alapelvek” alkalmazási területét a csupán szöveges forrásoktól az információforrások valamennyi típusáig.²

ISNI (International Standard Name Identifier): Nemzetközi szabványos névazonosító. Személyek minden kétséget kizáró azonosítására szolgál, rendszerint négyszer négy számjegyből áll.

1 OKR-projekt [elektronikus dok.] <http://www.oszk.hu/okr-projekt> [letöltés: 2019.07.18.]

2 Nyilatkozat a Nemzetközi katalógizálási alapelvekről [elektronikus dok.] URL: http://www.oszk.hu/sites/default/files/ICP_2016_magyarul_blogra_modositott.pdf [letöltés: 2019.07.18.]

Linked data: Az adatok hálózatba szervezésének, összekapcsolásának technológiája. Nem csupán a két kapcsolódó elem, de a köztük lévő viszony leírását is elvégezhetjük a segítségével. Informatikai alapkövei az egységes forrásazonosító (URI), illetve a formalizálást lehetővé tevő forrásleíró keretrendszer (RDF).

LRM (Library Reference Model): Az IFLA által kidolgozott, egységes fogalmi modell, mely összesen 11 entitást határozott meg, amelyekről a forrásleírás során "beszélni" lehet (Res, Mű, Kifejezési forma, Megjelenési forma, Példány, Ágens, Személy, Kollektív ágens, Nomen, Hely, Időtartomány).³ Az LRM megjelenése szükségessé tette az RDA alapos átdolgozását.

MARC21: Az 1998-1999-ben közzétett MARC21 a MARC legfrissebb, ma is folyamatosan aktualizált változata, a 21-es szám a XXI. századra utal. Legutóbbi módosításai figyelembe veszik az új dokumentumtípusok sajátosságait, és már az RDA-ban definiált új adatelemek leírhatóságát, valamint a korszerűbb formátumokra történő konverziót is szolgálják.

Névtér: Besorolási (authority) adatok állománya, az intézményekben végzett forrásleíró munkától többnyire függetlenül épül.

Nyílt forráskód (open source): Olyan szoftver, mely a forráskód elérhetősége révén a tetszőleges célú futtatás, a testreszabás, továbbfejlesztés, tökéletesítés jogát biztosítja a felhasználóknak. A nyílt forráskódú szoftvereken fejlesztéseket végzők egymást segítő közösségekben tevékenykednek.

OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting): A Dublin Core formátumra építő, metaadat-begyűjtő protokoll, melynek segítségével különféle gyűjteményeket lehet egyszerű utasítások segítségével lekérdezni, vagy köztük adatcserét megvalósítani.

OJS (Open Journal Systems): Nyílt forráskódú, ingyenesen letölthető és használható, tudományos folyóiratok online közzétételére kifejlesztett folyóirat-menedzselő és publikáló rendszer.

Open Access: A tudományos közlemények nyílt hozzáférésű publikálása, mely a felhasználó számára biztosítja az olvasás és esetlegesen a további felhasználás legális, finansziális, szervezeti vagy technikai megkötések nélküli lehetőségét.⁴

ORCID (Open Research and Contributor Identifier): Nemzetközileg elfogadott szerzőazonosító kód, mely biztosítja a kutató egyértelmű azonosítását. Segítségével visszakereshető az adott kutató tudományos munkássága. Az ISNI részhalmaza.

3 IFLA könyvtári referenciamodell : A bibliográfiai információk elméleti modellje [elektronikus dok.] https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-hu.pdf [letöltés: 2019.07.18.]

4 Open Access [elektronikus dok.] URL: nkfh.gov.hu/download.php?docID=29473 [letöltés: 2019.07.18.]



Platform: a hagyományos integrált könyvtári rendszerek funkcióit kínáló, böngésző segítségével elérhető felület. Az új környezet egyszerűbbé teszi más, külső adatforrások felhasználását pl. a forrásfeldolgozásban, így megszünteti az információk elszigeteltségét. A könyvtári platformokhoz discovery-szolgáltatások kapcsolódhatnak. Könyvtári platform pl. az Alma, az OLSuite, valamint a közösségi fejlesztésű FOLIO.

RDA (Resource Description and Access): Forrásleírás és hozzáférés. Katalógizálási szabályzat, az AACR harmadik, jelentősen átdolgozott változata. A források leírásának alapjává az entitásokat és azok kapcsolatait tette, a leírás során a feldolgozó könyvtáros az RDA fejezeteit követve gyakorlatilag entitásokat ír le egymás után.

RDF (Resource Description Framework): Forrásleíró keretrendszer, amely lehetővé teszi, hogy a számítógép számára értelmezhető módon (formalizáltan) írjunk le két elemet, valamint a köztük lévő kapcsolatot. Az RDF így tehát három elemből álló "mondatokat", állításokat kezel, amelyeket tripletnek hívunk.

Repozitórium: Egyetemeken és kutatóintézményekben működő dokumentumszerver, amely tudományos anyagok archiválására és világszerte nyílt hozzáférhetővé tételére szolgál.

schema.org: A Google, a Microsoft, a Yahoo, valamint a Yandex által létrehozott kezdeményezés – egy újabb szótár –, amely segít a könyvtári rekordokat a Google-ba juttatni.

SFX: A kontextusérzékeny referenzlinkelés platformfüggetlen eszköze, amely könyvtári forrásokhoz dinamikusan generált linkhalmazokat hoz létre, s teszi mindezt az elektronikus dokumentum metaadataira támaszkodva.⁵

SPARQL: Az RDF-adatbázisok lekérdezőnyelve, adat-átalakításra is alkalmas (pl. egyik szótárról a másikra, pl. BF-schema.org)

Szemantikus web: Más néven web 3.0. A napjainkban ismert világháló egy olyan új rétege, amelyben az egyes elemek, adatok között a számítógépes feldolgozás számára is jelentéssel, értelemmel bíró, azaz informatikai eszközökkel tipizált, minősített kapcsolatok állnak fenn.

Triplestore: Adatbázis-kezelő rendszer, amely a relációs adatbázisokkal ellentétben az adatokat nem adattáblákban, hanem tripletekben tárolja (gráf-adatbázis). Ezen adatbázis lekérdező felületét (ún. SPARQL-endpoint) nyilvánosan hozzáférhetővé lehet tenni.

5 BÁNHEGYI Zsolt: Cummings, Joel-Johnson, Ryan: Az SFX használata és használhatósága = Tudományos és Műszaki Tájékoztatás, 51. évf. (2004) 1. sz., p. 42-44.

URI (Universal Resource Identifier): *Univerzális azonosító. A világon mindennek adható, legyen az kézzelfoghatóan vagy fogalmi szinten létező dolog. Ha a dolgokat és viszonyaikat egyaránt URI-kkal azonosítjuk, a számítógépes feldolgozás számára nem csupán a létezőket, de a köztük lévő kapcsolatot is formalizálni tudjuk.*

VIAF (Virtual Internet Authority File): *Virtuális, internetes besorolási állomány, több könyvtár (leginkább nemzeti könyvtárak) besorolási (authority) állományaiból építették. Alkalmas a személyek egyértelmű azonosítására.*

Webarchiválás: *weboldalak vagy webhelyek valamilyen rendszeres vagy rendszertelen időközönként megismételt mentései. A webarchívumok keresőrobotjai a linkeket követve keresik fel és mentik a weboldalakat, valamint egyéb fájlokat, amelyeket visszakereshetően tárolnak. A legismertebb webarchívum az Internet Archive.⁶*

6 DRÓTOS László-KOKAS Károly - Webarchiválás és a történeti kutatások = Digitális Bölcsészet, 1. évf. (2018) 1. sz., p. 35-53. Elektronikusan hozzáférhető a <https://doi.org/10.31400/dh-hun.2018.1.129> címen [letöltés: 2019.07.18.]

Parazita folyóiratok, a tudományos világ élősködői

Juhász Attila

Debreceni Egyetem, Informatikai Tudományok Doktori Iskola

juhaszattila93@gmail.com

Szabó Dóra

Debreceni Egyetem, Informatikai Kar, Könyvtártudományi Tanszék

szabog6dora@gmail.com

The emergence of open access undoubtedly had a positive impact on the efficiency of scientific communication. Among other things it eliminated the spatial and temporal constraints, improved international cooperation, and caused the shortening of research times. However, parallel to this, predatory publications, that pose a threat to the scientific world due to their profit oriented pseudo-scientific activity, also appeared. Our paper discusses ongoing research that examines the scientific visibility of educators in domestic institutions of higher education and their relation to predatory journals.

Keywords: Open Access, predatory journals, scientific visibility

Bevezetés

Az Open Access (OA), azaz a nyílt hozzáférés megjelenésével a tudományos tartalmak digitális formában, online elérhetővé váltak. Ezen publikációk jogi korlátozás nélkül, a megfelelő hivatkozásokkal szabadon felhasználhatók mindenki számára. Ebből következett, hogy felgyorsult a tudományos kommunikáció sebessége, hiszen a legfrissebb kutatási eredmények bárhol, bárki számára elérhetővé váltak, ezáltal sikerült áthidalni a korábbi tér-és időbeli nehézségeket, ami előnyös lehet a nemzetközi kutatások hatékony összehangolása esetében is. Az elektronikus megjelenésnek köszönhetően nőtt a szerzők olvasottsága, mely a legtöbb esetben megmutatkozott az idézettségükön is. Ugyanakkor a pozitív hozományok mellett a negatívumokról sem feledkezhetünk meg. Megjelentek az úgynevezett ragadozó kiadványok, melyek magukat tudományos kiadványként tüntetik fel, viszont valójában profitorientált, a tudományosságot nélkülöző lapokról van szó. A jelenlétük azért is ártalmas, mert általuk számtalan áltudományos mű válik elérhetővé bárki számára, ezáltal csorbitva a tudományosságba vetett hitet. Alább a jelenleg folyamatban lévő kutatásunkat¹ ismertetjük, amelyben a hazai egyetemek oktatóinak tudományos láthatóságával és a ragadozó kiadványokhoz fűződő viszonyával foglalkozunk.

¹ https://docs.google.com/forms/d/1q9JpNv-Hk2JP-iVh4YBqolcbzdhcracV204dhuPPiBo/edit?usp=forms_home (2019. 06. 30.)

1. A ragadozó folyóiratok

A parazita, vagy más néven ragadozó folyóiratok² (angolul predatory journals) a tudományos világ élősködőiként jelentek meg az elektronikus publikációk térhódításával. A ragadozó kiadványok tudományos folyóiratok álcáját öltik magukra, nívós lapként tüntetik fel magukat, viszont nem válogatják meg a publikálandó cikkek körét, a megfelelő díjak fejében bármit közzétesznek. A szerkesztőbizottság tagjait az interneten toborozzák, a kutatókat pedig kéretlen levelekkel bombázzák, hogy náluk tegyék közzé legújabb eredményeiket. A megtévesztés módszerét alkalmazzák: nívós lapokhoz hasonló nevet és megjelenést választanak. Mivel nem ellenőrzik a cikkeket, nem felelnek meg a tudományos folyóirat-kiadás szakmai követelményeinek. Tevékenységük súlyoskárokat okoz, mivel csorbítja a tudományos világba vetett hitet. Ez azért is jelentős probléma, mert a nyílt hozzáférés korában már bárki hozzáférhet ezen ellenőrizetlen cikkekhez, amelyek, megtéveszthetik azon olvasókat, akik hiteles információhoz szeretnének hozzájutni. Tevékenységük egyik táptalaja a napjainkban jól ismert publish or perish elv,³ miszerint a tudományos előmenetelhez nélkülözhetetlenné vált a megfelelő számú publikáció megléte, azaz létrejött a publikálási kényszer. Az említett elv szerint, aki nem publikál az nem juthat előre a tudományos ranglétrán. A ragadozó folyóiratok természetesen erre sem kínálnak igazi megoldást, hiszen, ha egy kiadóról kiderül, hogy parazita tevékenységet folytat, abban az esetben a szerzők ott megjelentetett cikkeit nem fogadják el a tudományos előrehaladásukhoz, illetve a pályázataikhoz. A ragadozó tevékenység először Jeffrey Beall⁴ amerikai könyvtárosnak szűrt szemet, aki el is készítette 2011-ben az általa parazitának vélt kiadványok listáját (Beall-lista)⁵, illetve folyamatosan frissítette azt. A munkáját 2017-ben kénytelen volt befejezni az őt ért folyamatos támadások miatt, a lista pedig elavulttá vált, hiszen az ilyen jellegű kiadványok száma egyre csak nő. Ezért is fontos a megfelelő körültekintés, mielőtt a szerzők kiadnák kézírataikat a kezükből. Ehhez szükség van a ragadozó lapok ismérveinek alapos ismeretére, valamint érdemes a DOAJ-ban⁶ is megnézni, hogy szerepel-e ott az adott kiadvány, hiszen ott megtalálhatóak az indexelt OA kiadványok.

2. A kutatók és a tudományos láthatóságuk vizsgálata

Mint a fentebb említett publish or perish elvből kiderült, a megfelelő számú publikáció elengedhetetlen ahhoz, hogy a kutató előre juthasson a szakmai ranglétrán. A jelenleg is folyó, terveink szerint reprezentatív mintát elérő elektronikus felmérésünkben (elektronikus Google kérdőív), melyre jelenleg 500 válaszadónk van, arra keressük a választ, hogy mi motiválja a hazai felsőoktatásban az oktatókat.

2 Holl András, Parazita folyóiratok (predatory journals),

URL: https://www.mtmt.hu/system/files/parazita_folyoiratok.pdf (2019. 06. 30.)

3 Erzsebet, Dani: How "Publish or Perish" Can Become "Publish and Perish" in the Age of Objective Assessment of Scientific Quality

URL: [http://www.iiisci.org/journal/CV\\$/sci/pdfs/IP052LL18.pdf](http://www.iiisci.org/journal/CV$/sci/pdfs/IP052LL18.pdf) (2019. 06. 30.)

4 Interjú Jeffrey Bealllel, Könyv, Könyvtár, Könyvtáros, URL: <http://ki2.oszk.hu/3k/2016/09/a-tudomany-eloskodoi-denveri-beszelgetes-jeffrey-beall-lel/> (2019. 07. 21.)

5 Alan Burdick: "Paging Dr. Fraud": The Fake Publishers That Are Ruining Science, The New Yorker 5 BEALL'S LIST OF PREDATORY JOURNALS AND PUBLISHERS
URL: <https://beallslist.weebly.com/> (2019. 03. 11.)

6 Directory of Open Access Journals <https://doaj.org/> (2019. 06. 30.)



Ez a kérdés azért releváns, mert aki motivált az előrejutás felé, az feltehetően szeretné, ha a tudományos munkája szélesebb körben lenne elérhető, hiszen így növelhető az idézettsége, amire az OA folyóiratok tökéletes lehetőséget nyújtanak. Feltehetően a motivált válaszadók a saját intézményük repozitóriumán túl egyéb felületeken is közzéteszik a munkáikat (pld. Google Scholar, Research Gate, esetleg személyes honlap). Hazánkban az elektronikus közzététel egyik legismertebb felülete a Magyar Tudományos Művek Tára (MTMT)⁷, amely nem repozitórium, hanem nemzeti bibliográfiai adatbázis, melynek működése törvény által is szabályozott. A kutatásunk során azt is megvizsgáljuk, hogy a megkérdezettek mekkora része gondolja rendszeresen ezen felületét. Ez azért is fontos, mert itt a szerzőknek lehetőségük van arra, hogy könnyedén összeállíthassanak bibliográfiákat, publikációs listákat, amelyeket felhasználhatnak pályázatok során is, mely szintén összefüggésben állhat a motiváció mértékével. Továbbá vizsgáljuk azt is, hogy a megkérdezettek mekkora része publikált már nyílt hozzáférésű folyóiratban, illetve, ha esetleg még nem, akkor nyitott-e ezen megjelentetési forma irányába, vagy inkább tartózkodnak tőle. A felmérés egyik legfontosabb kérdése pedig azokhoz szól, akik már publikáltak nyílt hozzáférésű kiadványban, esetükben ugyanis azt szeretnénk megtudni, hogy tapasztalták-e a tudományos láthatóságuk növekedését a nyílt hozzáférésnek köszönhetően. A nemzetközi tapasztalat szerint egyértelműen megállapítható, hogy nő azon kutatók idézettsége⁸, akik szabadon elérhetővé teszik eredményeiket. Úgy véljük, hogy ez a hazai válaszadók esetében sem lesz másként.

Továbbá szeretnénk megtudni, hogy a szerzők rendelkeznek-e ORCID⁹ azonosítóval. Ennek az a lényege, hogy a szerzők egyértelműen azonosíthatók legyenek, mivel gyakoriak a névazonosságok, amelyek problémát okozhatnak a teljesítménymérés folyamatában. Az utóbbi időben több kiadó is ehhez a díjmentesen létrehozható azonosítóhoz köti a kéziratok befogadását.

3. A kutatók és a ragadozó folyóiratok kapcsolatának vizsgálata

A tudományos láthatóságra vonatkozó kérdéseket követően a felmérés túlnyomó részében a hazai felsőoktatási dolgozók és a ragadozó folyóiratok közötti kapcsolatot vizsgáljuk.

Először azt szeretnénk megtudni, hogy a megkérdezettek mekkora része találkozott már a tudományos munkája során ilyen jellegű kiadványokkal, vagyis a válaszadók mekkora részét keresték meg valamilyen módon. Úgy véljük, hogy a többség már kapott felkérést ragadozó kiadványoktól valamilyen formában. Ugyanakkor fontos, hogy a szerző el tudja dönteni az adott kiadványról, hogy az valóban parazita tevékenységet folytat-e, melyhez szükség van a megfelelő kritikai érzéken túl arra is, hogy a kutatók ismerjék ezen kiadványok ismérveit. Ennek megismerésére a kitöltés során egy listából kell kiválasztania a lehetséges ismérveket a megkérdezetteknek. Előzetes hipotéziseink szerint a válaszadók nagy többsége képes felismerni az árulkodó jeleket, azaz képes

7 Magyar Tudományos művek Tára, URL: <https://www.mtmt.hu/> (2019. 06. 30.)

8 Gemma Hersh, Andrew Plume: Citation metrics and open access: what do we know?, URL: <https://www.elsevier.com/connect/citation-metrics-and-open-access-what-do-we-know> (2019. 03. 11.)

9 <https://orcid.org/> (2019. 06. 30.)

döntést hozni egy-egy kiadvány jellegéről. A kérdőív további szakaszában Likert-skála segítségével kívánjuk megismerni a válaszadók véleményét.

A ragadozó folyóiratokra vonatkozó kérdések sorában az első arra vonatkozik, hogy a válaszadó mennyire tartja veszélyesnek a kiadványokat a tudományos világra tekintve. Úgy véljük, hogy a megkérdezettek többsége valós veszélyként tekint ezekre a kiadványokra, hiszen súlyosan csorbitják a tudományba vetett hitet, illetve azontúl, hogy gyakran valótlan információkat közölnek, akár érdemtelen előnyhöz is juttathatják azokat, akik ilyen jellegű kiadványokban publikálnak a tisztességes utat választó szerzőkhöz képest. Aki nem szeretne ezen kiadványok csapdájába esni, fontos, hogy alaposan járjon utánuk. Erre vonatkozik a következő kérdés, miszerint alaposan utánajár-e a szerző az egyes kiadványoknak, mielőtt ott megjelentetné cikkeit. Szeretnénk azt is megtudni, hogy a válaszadók hogyan vélekednek, arról, hogy társaik képesek-e eldönteni egy-egy folyóiratról, hogy az áltudományos tevékenységet végez-e. Úgy gondolom, hogy a válaszadók többsége már kapott felkérést ilyen kiadványoktól, illetve mint fentebb az ismérvek kiválasztásánál is részleteztük, feltehetően el tudják dönteni, hogy tisztességes OA folyóirattal van-e dolguk.

A ragadozó folyóiratok veszélyt jelentenek áltudományos tevékenységükkel a tudományos világra, viszont szeretnénk megtudni, hogy a válaszadók tulajdonítanak-e nekik esetleg valamiféle előnyt. Ezen előny lehet akár az is, hogy gyors a publikációs folyamat, hiszen itt nincs valós szakmai bírálat, ebből adódik az is, hogy hamar kerülnek publikálásra a cikkek. Természetesen ez nem tekinthető a szó szoros értelmében vett előnynek, hiszen ez éppen a minőség rovására megy. Ebből a kérdésből kiindulva arra is keressük a választ, hogy a megkérdezettek szerint egyes kutatók a személyes előrehaladásuk érdekében akár önszántukból is publikálnának-e ragadozó folyóiratokban. Ezt mindenképpen elítélendőnek tartjuk, viszont mint lehetőséget nem lehet kizárni, hiszen a valós eredmények hiányában elegendő a megfelelő közzétételi díj megfizetése és a kutató máris több nemzetközi cikkel rendelkezhet, amelyet feltüntethet pályázatok benyújtása során, vagy a tudományos előmenetelét biztosíthatja a segítségükkel. Viszont, mint korábban írtuk, amennyiben egy folyóiratról kiderül, hogy ragadozó tevékenységet folytat, abban az esetben az ott megjelentetett kiadványokat nem lehet elszámolni semmilyen formában sem, illetve szükséges lenne, ezen tartalmak eltávolítása az elektronikus felületekről.¹⁰ Amennyiben egy szerző ilyen jellegű kiadvánnyal találkozik, abban az esetben azt jelentheti az MTMT felé is.¹¹

Továbbá vizsgáljuk, hogy a kutatók kellően tájékozottak-e a ragadozó lapokban történő publikálás veszélyeiről. Ezek nem csupán a tudományos világ hitelességére jelentenek veszélyt, hanem magukra a szerzőkre is, mivel az itt megjelentetett

10 The 9th International Conference on Society and Information Technologies: Proceedings. Konferencia helye, ideje: Orlando (FL), Amerikai Egyesült Államok, 2018.03.13-2018.03.16. Winter Garden (FL): International Institute of Informatics and Systemics (IIIS), 2018. 43-44p. (ISBN:9781941763742) Jeffrey, Beall: [Predatory journals: ban predators from the scientific record](#). *Nature*, 534 (2016), 326p.

11 Parazita (predátor) folyóirat bejelentés URL: <https://www.mtmt.hu/parazita-predator-folyoirat-bejelentes> (2019. 03.11.)



munkák nem közölhetőek újra valódi tudományos lapban, tulajdonképpen a semmire fizetnek publikálási költséget, valamint ezen kiadók megtartják a szerzői jogokat is.¹² A parazita folyóiratokban a publikálókat két csoportra lehet bontani. Az egyik, lényegesen nagyobb részük gyanútlan áldozat, akik nem szándékosan publikálnak a kiadványokban, míg a másik, kisebbik részük tisztában van a kiadó tevékenységével. Mindkét csoport tagjaira érvényesek viszont a fentebb ismertetett hátrányok, így érdemes alaposan utánajárni a kiadványoknak.

Ahhoz, hogy a káros tevékenységet folytató kiadók kiszoruljanak a tudományos világ piacáról, véleményem szerint egységes fellépés kellene. A válaszadók véleményét is szeretnénk megismerni a kérdést illetően, és arra számítunk, hogy a többség szintén egyetért az összehangolt fellépés szükségességét illetően. A kérdőív végén arra vagyunk kíváncsiak, hogy az oktatók mit tartanak a nyílt hozzáférés előnyeinek, illetve hátrányainak. Azt gondolom, hogy a válaszadók az ingyenességet fogják a legnagyobb mértékben kiemelni, míg hátrányként a minőségi fenntartások túlsúlyba kerülésére számítok, ami összefüggésben állhat kutatásom tárgyával, a ragadozó folyóiratok egyre növekvő térnyerésével.

A kutatás tavasz végi lezárása után, a kapott válaszok kielemezését követően fogunk tudni pontos képet adni a tanulmányban felmerülő kérdésekre, melyek azt gondolom, hogy igencsak aktuálisak és a tudományos világ összes résztvevőjét érintik valamilyen módon.

4. Összegzés

Ajelenleg is zajló, reményeink szerint a reprezentatív mintát elérő kutatást két részre oszthatjuk, az első fele a kutatók tudományos láthatóságát vizsgálja, miszerint mennyire motiváltak, illetve amennyiben fontosnak érzik, hogy szélesebb körbe juttathassák eredményeiket, abban az esetben mit tesznek érte. Illetve szeretnénk megtudni, hogy akik már eddig is jelentettek meg nyílt hozzáférésű cikkeket, ők tapasztalták-e a tudományos láthatóságuk növekedését. A felmérés második, nagyobb része a ragadozó folyóiratokkal foglalkozik. Arról szeretnénk átfogó képet kapni, hogy a válaszadók mekkora része találkozott már ilyen jellegű kiadvánnyal, illetve képesek-e eldönteni egy esetleges megkeresés esetén azt, hogy az adott kiadvány valóban OA jellegű, vagy csupán áltudományos tevékenységet folytat. A Likert-skálás mérések lehetőséget adnak arra, hogy alaposabban megismerhessük a kitöltők véleményét a kérdéskörben, illetve átfogó képet adhassunk a hazai helyzetről. A kutatás lezárása 2019 őszén várható, ezt követően kezdődhet el a kapott eredmények kiértékelése.

5. Köszönetnyilvánítás:

A kutatást az „Integrált kutatói utánpótlás-képzési program az informatika és számítástudomány diszciplináris területein” (EFOP-3.6.3-VEKOP-16-2017-00002) című projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

¹² Farkas Boglárka: Open Access források, a publikálás veszélyei, parazita folyóiratok URL: <https://slideplayer.hu/slide/11936495/> (2019. 03. 11.)

Egészségügyi informatikai adatbázisban való szöveges keresés mesterséges intelligenciával támogatott szemantikus keresővel

Kovács Béla Lóránt
Neumann Technology Kft.
bela.lorant.kovacs@negentropics.com

Text search in a health database with a semantic search engine supported by artificial intelligence My purpose is to present a search engine which can process large documents- even books- in the query and search in the medical database based on them. The process works without metadata, so it is enough to upload the search text to the database once, the application executes all the rest of the operations. Searching methods are supported by algorithms based on semantical modeling of natural languages, working in any language. Namely, the algorithm learns languages from the databases. However, the search engine is not only capable of conducting semantic searches but define topics by texts that can be monitored automatically by the program. This process is supported by an artificial intelligence what is suitable for determining semantical groups in one hand, on the other hand, it is learning from the database expansion and user behavior – based on which it can improve the hits, - at the same time, it does not collect any data about the user himself, thus complying with the strictest European data protection rules. I intend to present my description through practical examples demonstrating the innovations that make it unique. At the end, I would like to show in detail what measurable performance the software has and how it differs from other search engines.

Keywords: semantic search, semantic classification, artificial intelligence, search in text documents

A Neumann Technology Kft. a Gazdaságfejlesztési és Innovációs Operatív Program keretén belül, az Európai Regionális Fejlesztési Alapból és hazai központi költségvetési előirányzatból támogatott GINOP-2.1.7-15-2016-00069 azonosító számú projektben egy unikális szoftver prototípusán kezdett dolgozni. A pályázat eredményeként olyan programot fejlesztettünk, amely webes felületen elérhető és általa a felhasználó képes orvosi szövegekben keresni úgy, hogy az eddig rendelkezésre álló keresőknél pontosabb találatot kapjon.¹ A termék a következő elemekből áll:

- Előfeldolgozó
- Index adatbázis
- Kereső adatbázis
- Kereső modul

¹ Méréseinket 2019 márciusában végeztük. Ennek során kifejezetten az orvosi cikkeket tartalmazó adatbázisok és a hozzájuk kapcsolódó keresők vizsgálatára koncentráltunk, az olyan vizsgálatoktól, amely webes keresők pontosságát is mérte volna, mint amilyen Bennett Shenkeré, tartózkodtunk. (v. ö.: [1])

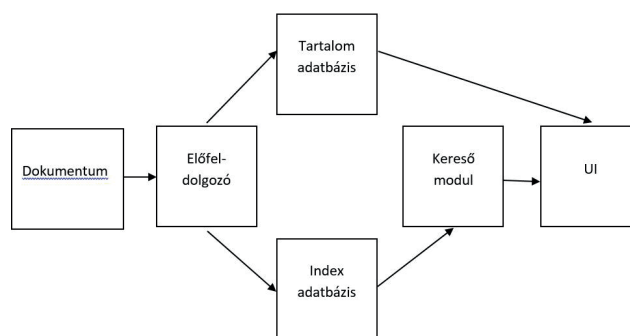


- Végfelhasználók által elérhető kereső funkciókhoz kialakított felület
- Karbantartó munkatársak által használható funkciókhoz kialakított felület
- Vezető munkatársak által használható funkciókhoz kialakított felület

A keresőmotor, amelyet az Aleetheia névre kereszteltünk, két szempontból is figyelemre méltó: egyrészt a természetes nyelvek szemantikájának modellezését és a mesterséges intelligencia alkalmazását a korábbiaktól eltérő módon oldja meg, másrészt a korábbi, hasonló keresőkhöz képest pontosabb és teljesebb találati listát ad. Éppen ezért a továbbiakban a program rövid bemutatása után ez utóbbi tulajdonságára összpontosítok a terméknek – minthogy a felhasználók számára is ez a legfontosabb² – és azt kívánom bemutatni, hogy mit értünk az alatt, hogy pontosabb és teljesebb találati listája, mint a más szoftvereké.

A program működése

A felhasználók a publikus interfész rétegen keresztül kapcsolódhatnak a szolgáltatási réteghez, amelynek a segítségével az alkalmazás funkciókat érhetik el, például a tartalom szerinti keresést. A szolgáltatási rétegben a vezérléseket megfelelő szolgáltatási csatornán fogadjuk, azokat értelmezzük és ütemezetten végrehajtjuk. A szolgáltatási réteg képes fogadni a parancssori és a webes szolgáltatási kéréseket, és az elemző ezeket ütemezetten fogadva és végrehajtva kapja meg. Az alkalmazás magjával készítjük a dokumentum előfeldolgozását. Különböző formátumú szöveges dokumentumok konvertálása, a konvertált tartalmak átadását végzi az elemző részére. Az előfeldolgozó képes a különböző dokumentum formátumok konvertálására. Elvégezzük a tárolást, mely modul feladata kettős: egyrésztől tárolni az előfeldolgozó által átalakítandó bemeneti forrásokat, illetve az elemző által kategorizált kimeneti adatokat, döntésekhez szükséges információkat. Tárolja a tanuló adatbázist, a számított eredményeket, illetve megfelelő formátumra konvertált dokumentumokat. Ezt követően elemzéseket végzünk, az előfeldolgozott forrásokon a kategorizáláshoz szükséges döntéseket hozó algoritmusok végrehajtásával (tanulás, osztályozás, döntés, klaszterezés), illetve a feldolgozott források meghatározott struktúrában való elhelyezésével. Másrészt az említetteken kívül az instrukciók nélküli, tartalom alapján végzett csoportképzéshez szükséges algoritmusokat is tartalmazza és futtatja.



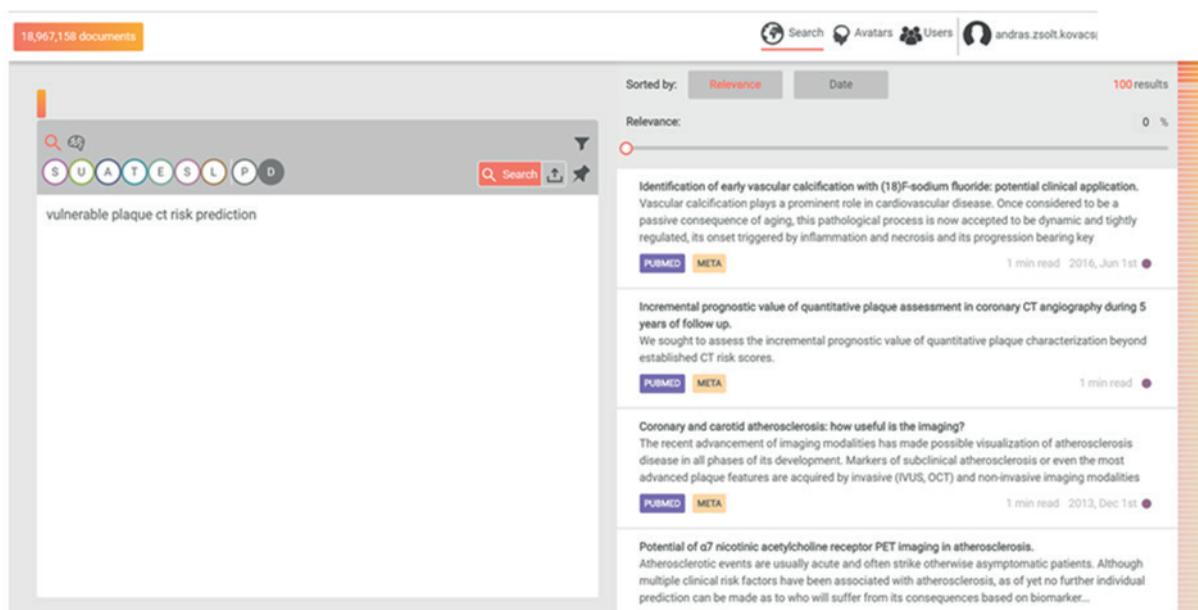
1. ábra: A program felépítése

1. ábra: A program felépítése

² A felhasználók érdeklődéséről: a találatok pontosságáról és teljességéről bővebben lásd.: [2]

A program pontossága

A pontosságot úgy mértük, hogy kutató orvosokat és doktoranduszokat kértünk meg arra, hogy a PubMed adatbázisán egy-egy keresés kapcsán osztályozzák a találatok pontosságát.³ A találatokat egytől tízig pontozták úgy, hogy a tízes találat ért a legtöbbet, az egyes pedig a legkevesebbet, és minden keresésnél a találati lista első húsz elemét vizsgálták meg. A saját találatainkat két keresési algoritmussal adtuk meg: az első kulcsszavak, a második teljes szövegekkel történő keresésre alkalmas. Ezeket a találatokat vetettük össze a PubMed saját keresőjével, amely ugyan alapvetően kulcsszavas, de metaadatokat is figyelembe vesz.⁴ A mi algoritmusaink nem vettek figyelembe metaadatokat és a mérés során nem használtunk keresésre olyan hosszúságú szövegeket, amelyek a PubMed keresőjén nem futnak le. Arra is figyeltünk, hogy ne hibridizáljuk a különböző algoritmusaink által kínált eredményeket, mert ugyan ezáltal még nagyobb pontosságra tehetünk volna szert, ugyanakkor elfedtük volna az egyes algoritmusok valódi képességeit. A továbbiakban éppen ezért a PubMed 19 millió dokumentumot tartalmazó adatbázisán lefuttatott keresések kapcsán három találati listát hasonlítottunk össze: a saját kulcsszavas, a saját teljes szöveges algoritmusaink, illetve a Pubmed keresője által adottakat. A kereső felület a következő módon néz ki:



2. ábra Kereső felület (Jobb oldalon a találati lista, bal oldalon a kereső ablak az algoritmusválasztóval)

A felületen látható az algoritmusválasztó, amely a termék végleges változatában nem lesz a felületre kivezetve, mivel a választást az emberek helyet hibridizációs algoritmusok végzik. Számos további funkcióval rendelkezik a program, amelyeket

3 A méréseket 2019. januárja és márciusa között végeztük négy orvos és két doktorandusz bevonásával.

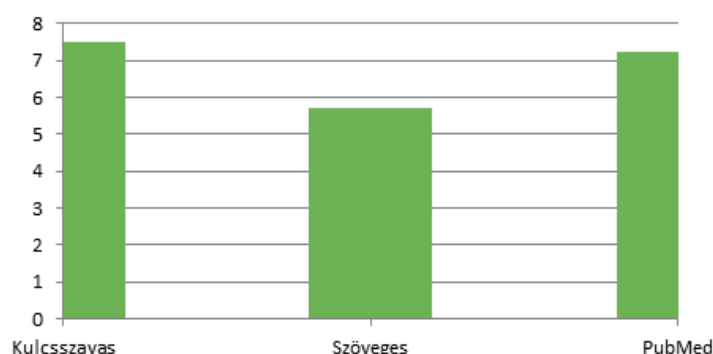
4 A hagyományos kulcsszavas keresők esetében – amilyenek például az egészségügyi adatbázisokban vagy a vállalati adatbázisok keresőiben működnek – kétféle eljárást szoktak kombinálni: az első a keresett kulcsszó gyakoriságából, a második a hozzá kapcsolódó metaadatok (tárgyszavak, tag-ek) előfordulásából szokott kiindulni. (v. ö.: [4].)

azonban hely hiányában nem tudunk bemutatni (témákhoz tartozó szöveghők és szóstatistikák, trendek, szerzőkre vonatkozó statisztikák stb.).

Első mérés

A keresőbe írt kifejezés a következő volt: „*vulnerable plaque ct risk prediction*”.

Az Aleetheia kulcsszavas algoritmus 75%-os pontosságot ért el, a szöveges azonban csak 57%-ot. A hagyományos kereső 72%-os pontosságra volt képes.



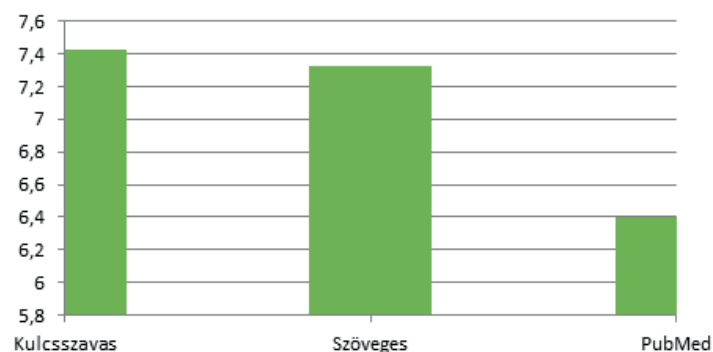
1. Az első keresés pontossága

A találatok részletes elemzéséből kiderül, hogy a kulcsszavas kereső annak ellenére adott pontosabb találatokat, hogy nem használt metaadatokat. Ráadásul nem csak pontosabb, hanem teljesebb is volt a listája, mint a hagyományos keresőnek, hiszen olyan szövegeket is megtalált, amelyeket a másik nem. A szöveges kereső ugyanakkor rosszabb eredményeket adott, köztük teljesen hibásakat is. Ezzel szemben a PubMed keresője ugyan helyes találatokat adott, ám mindössze tizenkét darabot.

Második mérés

A keresőbe írt kifejezés a következő volt: „*cardiovascular risk prediction scores*”.

Az Aleetheia kulcsszavas algoritmus 75%-os eredményt ért el, ám itt már a szöveges kereső is 73%-os pontosságra volt képes. A PubMed itt azonban már csak 64%-os teljesítményt mutatott.



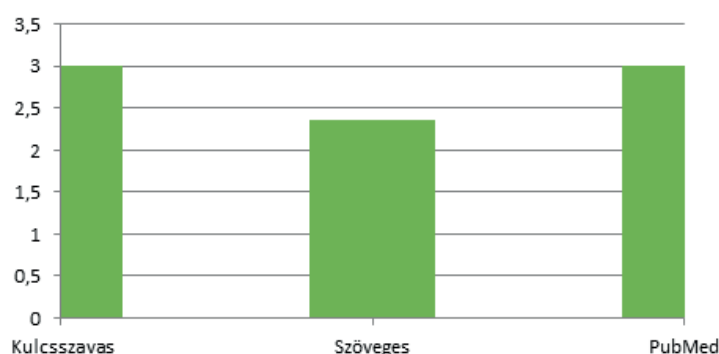
2. A második keresés pontossága

Az eredmények alaposabb elemzéséből az derül ki, hogy az Aleetheia kulcsszavas és a szöveges keresője is jó eredményeket hozott, ám a kétféle algoritmus nem ugyanazokat a jó találatokat jelenítette meg – bár metszete főleg a kiemelkedően jó találatok esetében volt a két listának. Ebből arra a következtetésre jutottunk, hogy érdemes a kétféle keresés erőnyeit egyesíteni. A PubMed keresőjének kudarcát teljessége magyarázza. A keresett kulcsszavak sok dokumentumban felbukkantak és mivel a PubMed keresőjének természetes nyelvi szemantikus képességei nem voltak, így számos hibás találat is bekerült a listába. Itt látszik, hogy a teljesség növekedése hogyan megy a pontosság kárára a hagyományos keresők esetében.

Harmadik mérés

A keresőbe írt kifejezés a következő volt: „*microvesicles microparticles cardiovascular plaque*”.

Ezt a találati eredményt azért mutatom be, mert nagyon jól szemlélteti, hogy mik a különbségek a hagyományos kulcsszavas és a természetes nyelvek szemantikáján alapuló keresők között. Ha csak a listák pontosságát nézzük, akkor szigorú értelemben az Aleetheia kulcsszavas keresője ugyanolyan eredményre volt képes, mint a PubMed-é, szerény 30%-os pontosságra, míg a szöveges kereső ennél is rosszabbra, 24%-ra.



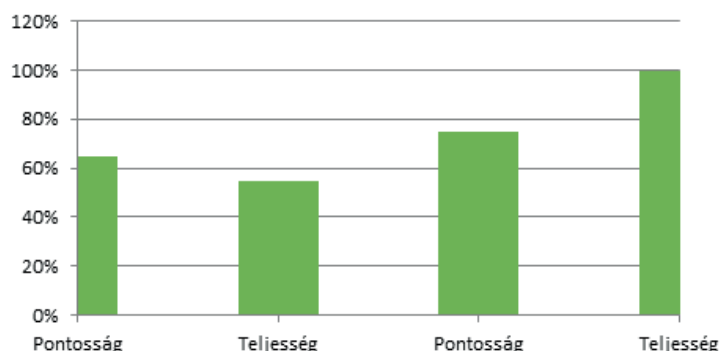
3. A harmadik keresés pontossága

Az eredmények alaposabb tanulsága azonban lényegesen árnyalja a képet. A kulcsszavas keresők ugyanis csak azokat a találatokat jelenítik meg, amelyek minden kulcsszót tartalmazzák vagy pedig minden kulcsszóhoz tartalmazó metaadattal kapcsolatban vannak. Éppen ezért egyetlen, alacsony pontszámmal rendelkező szöveget jelenítenek meg. A szöveges azonban nem ebből indult ki, így számos értékelhetetlen dokumentum mellett több olyat is adott, amelyeket a kutatók sokkal jobbaknak találtak, mint azt, amit a kulcsszavasak adtak.



Tanulságok

Méréseink során több száz keresési listát elemeztünk és mindenhol hasonló eredményre jutottunk – eltekintve természetesen néhány mérési hibától és meglepő szélsőségtől. Az Aleetheia algoritmusai 70-75%-os pontosságra voltak képesek 100%-os teljesség mellett, míg méréseink szerint a hagyományos motorok, köztük a PubMed keresője 60-65%-os pontosságra 55-60%-os teljesség mellett.⁵



4. ábra Kékkel a PubMed, narancssárgával az Aleetheia keresőjének pontossága és teljessége

Az algoritmusok hibridizációjával azonban a pontosság lényegesen növelhető volt 90% közelébe. Mindezt a költséges és időigényes metaadatolás nélkül sikerült elérnünk. Keresőnk különösen jó eredményeket volt képes elérni azokban az esetekben, amelyekben mások nem tudtak. Ilyenek a hosszú szöveges keresések vagy pedig azok a kulcsszavas keresések, ahol a kereső kifejezés csak néhányszor fordult elő, így olyan dokumentumokat is érdemes volt átnézni, amelyek szemantikus kapcsolatban állnak a keresett szóval, de nem tartalmazták azt. Ez a kereséstípus meglepő és szokatlanul pontos eredményeket képes elérni a legnagyobb teljesség mellett.

Összegzés

Az Aleetheia a szemantikus keresés új lehetőségeit teremtette meg. Metaadatok használata nélkül is képes a korábbiaknál pontosabb találatokat adni. A termék hordozza azon előnyöket, melyeket terveztünk – azaz nyelvfüggetlenül működik, mobil eszközön is megjeleníthető, gyors, és pontos keresési találatokat ad. Az orvosi alkalmazás kifejezetten jó teszteredményeket hozott, és tudományos-szakmai területen a PubMed adatbázison orvos-kutatóink továbbra is használhatják a saját kutatási területük újdonságainak megjelenítéséhez.

⁵ Az eredmény fontosságát mi sem bizonyítja jobban, mint az, hogy más keresők, így a Google Scholar vagy a MEDLINE keresője sem képes ilyen eredményre (v. ö.: [5])

Irodalomjegyzék

- [1] Bennett S Shenker, The accuracy of Internet search engines to predict diagnoses from symptoms can be assessed with a validated scoring system. International Journal of Medical Informatics, 83(2), 131-139, February 2014 doi: [10.1016/j.ijmedinf.2013.11.002](https://doi.org/10.1016/j.ijmedinf.2013.11.002)
- [2] Ovidiu Dan, Brian D. Davison, Measuring and Predicting Search Engine Users' Satisfaction, Journal ACM Computing Surveys (CSUR) Volume 49 Issue 1, July 2016 Article No. 18 doi: [10.1145/2893486](https://doi.org/10.1145/2893486).
- [3] Udo Kruschwitz, Charlie Hull, "Searching the Enterprise", Foundations and Trends in Information Retrieval: 2017, Vol. 11: No. 1, pp 1-142. doi: [10.1561/15000000053](https://doi.org/10.1561/15000000053)
- [4] Iqra Safder, Saeed-Ul Hassan, DS4A: Deep Search System for Algorithms from Full-Text Scholarly Big Data 2018 IEEE International Conference on Data Mining Workshops (ICDMW) doi: [10.1109/ICDMW.2018.00186](https://doi.org/10.1109/ICDMW.2018.00186)
- [5] Bramer, W.M., Giustini, D. & Kramer, B.M.R. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev 5, 39 (2016) doi: [10.1186/s13643-016-0215-7](https://doi.org/10.1186/s13643-016-0215-7)

MTA SZTAKI DSD – 25 éve a digitális könyvtárak szolgálatában

Kovács László, Micsik András
MTA SZTAKI Elosztott Rendszerek Osztály

laszlo.kovacs@sztaki.mta.hu, andras.micsik@sztaki.mta.hu

Bevezetés

25 évvel ezelőtt, 1994. január 1-én alapítottuk meg az MTA SZTAKI DSD, Elosztott Rendszerek Osztályt (dsd.sztaki.hu). Az osztály kezdetben, az akkori Magyarországon még újdonságnak számító World Wide Webes technológiák hazai bevezetésében, elterjesztésében volt úttörő. Az ország első webszolgáltatásai (pl. SZTAKI Szótár), kormányzati honlapjai (www.kormany.hu, www.kancellaria.gov.hu), webes műalkotásai (pl. Nightwatch, SZTAKI Gallery) létrehozása mellett gyorsan kialakult az a tematikus profil, mely azóta is töretlenül jellemzi az osztály kutatás-fejlesztési tevékenységét. Ebben a szakmai profilban az (elosztott) digitális könyvtári és archívum rendszerek kutatás-fejlesztése kiemelkedő és meghatározó szerephez jutott és jut a mai napig is, a digitális könyvtárak az MTA SZTAKI DSD osztály alapvető, szakmai identitásképző témaköre.

A 25 év során elért szakmai eredményeinket áttekintve, néhány jelentősebb, a témakörbe eső projektet mutatunk be, konkrét szakmai feladatok köré csoportosítva azokat.

A keresés aspektusai

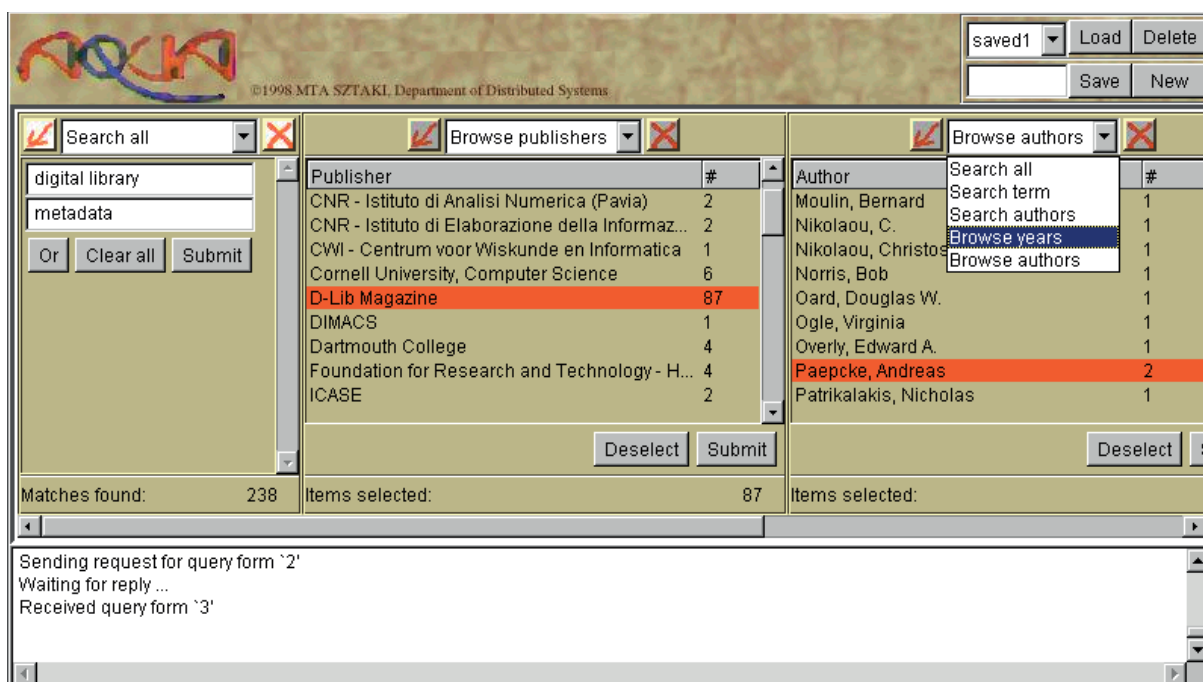
A digitális könyvtárak a World Wide Web elterjedésének egészen korai szakaszában megjelentek (sőt, voltak FTP és Gopher alapú digitális könyvtárak is, de ki emlékszik már rájuk?). Akkoriban azonban még nem voltak univerzális webes keresők (lásd Google), és így a több helyen egyszerre történő, elosztott keresés igénye hamar felmerült. A számítástechnika és számítógéptudomány diszciplínán belül az első digitális könyvtár az USA-ban épült meg és az NCSTRL¹ (Networked Computer Science Technical Report Library) nevet kapta. Az NCSTRL elosztott rendszerként egymáshoz kapcsolt könyvtári csomópontok hálózata volt, melyben egy elosztott keresési algoritmus alapján lehetett megtalálni a keresett digitális objektumokat (kutatási jelentéseket, tudományos cikkeket, publikációkat). A DSD osztály megalakulásával szinte egyidejűleg még 1995-ben felállítottuk az MTA SZTAKI-ban az NCSTRL első európai csomópontját és bekapcsoltuk az NCSTRL hálózatba. A rendszer az elosztott és központi keresési módszereket kombinálta, és sok azóta alapvetőnek tartott szolgáltatást vezetett be, mint például egységes API, perzisztens linkek stb. és a hamarosan megjelenő OAI-PMH protokoll is sokban hasonlít erre a korai API-ra.

A digitális könyvtári kutatási területen Európa jelentősen elmaradt az USA mögött, amit az Unió a DELOS LTR (Long Term Research) ESPRIT projekt indításával igyekezett behozni. A DELOS LTR projekt alapkutatási projektként meghatározó szerepet játszott abban, hogy Európában létrejött a digitális könyvtári szakmai témakörrel foglalkozó kutatók és fejlesztők nemzetközi közössége, és

¹ <http://www.cs.cornell.edu/NCSTRL/>

megteremtődött az európai digitális könyvtári kutatások humán erőforrás alapja, egyben akkor jelentős pénzügyi forrásokat, támogatást is kapott e szakmai terület. Az MTA SZTAKI DSD a kezdet kezdetén történő részvétele a DELOS LTR projektben tette lehetővé azt, hogy elkezdődjön Magyarországon is a digitális archívumok kutatása és fejlesztése, immáron harmonikusan beágyazva az európai digitális könyvtári K+F erőfeszítésekbe. Később az MTA SZTAKI DSD folyamatosan jelen volt az európai digitális könyvtári kutatói közösségben a DELOS NoE, illetve a DELOS NoE2 európai (FP4 és FP5) projektek tagjaként és számos más hazai és nemzetközi K+F projektben kamatoztathatta az európai szinten kifejlesztett, illetve az ott elsajátított technológiákat, műszaki megoldásokat.

Az MTA SZTAKI az ERCIM (European Research Consortium for Informatics and Mathematics) konzorciumba való belépése után jelentős szerepet kapott az ERCIM saját elosztott digitális könyvtári rendszerének létrehozási folyamatában. Elkészült az ETRDL² (ERICIM Technical Reference Digital Library), melyben az MTA SZTAKI hazai csomópontot üzemeltetett és az ETRDL (Európa) - NCSTRL (USA) kapcsolat létrehozásáért volt felelős. Később az ETRDL technikai alapjain készítettük az AQUA keresőnket, amely korai Java applet technológiával az ismétléses finomítás (iterative refinement via facets) keresési módszert támogatta³.



1. ábra: Az AQUA keresőfelület 1998-ból

2 A. Andreoni et al.: [The ERCIM Technical Reference Digital Library](#). D-Lib Magazine December 1999

3 L. Kovács, A. Micsik, B. Pataki: [AQUA: query visualization for the NCSTRL digital library](#). Proceedings of the fourth ACM conference on digital libraries. Berkeley, 1999.



A nagy keresőmotorok hatására a központosított keresőszolgáltatások váltak népszerűvé, mivel gyorsabbak és megbízhatóbbak voltak elosztott társaiknál. Ezek elterjedését roppant mód felgyorsította az OAI-PMH protokoll, melyet először a magyar OSZK-SZTAKI Hektár⁴ projekten belül próbáltunk az OSZK MEK-kel együtt itthon népszerűsíteni. A Hektár projektben kialakított kereső megoldás az általunk kifejlesztett NDA@SZTAKI digitális könyvtári rendszer felületen is megjelent, mely az időközben hasonló elvek mentén megvalósult Nemzeti Digitális Adattár alternatív keresőfelülete volt. Ez a szolgáltatásunk jelenleg az OAI kereső⁵ nevű reinkarnációjában él tovább, és a nyílt teljes szövegű repozitóriumok és folyóiratok közös, országos keresőfelületét nyújtja.

Továbbra is foglalkoztatott minket azonban az elosztott digitális könyvtárak és archívumhálózatok problémaköre, és a kis- és közösségi rádiózás szükségleteinek megfelelő peer-to-peer elosztott rádióarchívum-hálózatot fejlesztettünk ki a StreamOnTheFly EU projektben⁶. Itt a csomópontok meghatározhatták nyelvi, regionális, vagy bármely más alapon, hogy mely más csomópontokkal cserélnek metaadatokat.

A StreamOnTheFly európai hálózat több mint 10 éven keresztül üzemelt megbízhatóan és biztosította az európai közösségi rádiócsatornák archiválási igényeit, illetve a csatornák közötti műsorcserét, mintegy virtuális, ingyenes piacot hozva létre a multimédia (leginkább audió) tartalmak kicserélésére és újrafelhasználására. A kis- és közösségi rádiócsatornák ugyanis mindig is erőforráshiánnyal küzdöttek, ezért a StreamOnTheFly hálózat jelentősen hozzájárult e csatornák mindennapi takarékos működéséhez, fennmaradásához.

A StreamOnTheFly hálózat felbomlása után több mint 10 évvel digitális "maradványként" a radio.sztaki.hu⁷ oldalon ma is behallgathatunk magyar városi rádiók korábban archivált műsoraiba.

A nem szöveges média formátumok (audió, kép, videó stb.) terjedésével az ezekben való keresés lett az aktuális probléma. Az MTA SZTAKI CrossMedia projektünkben a képi információk és metaadatok (szemantikus) keresésének kombinálási lehetőségeivel kísérleteztünk⁸.

Végző soron a plágiumkeresés is egy ilyen újfajta keresési technika, amely sokféle dokumentumformátumból kivonja a szöveget, és észleli az egyező szövegrészeket a különböző dokumentumokban. Az MTA SZTAKI KOPI⁹ plágiumkeresőjét 2004-ben

4 <http://hektar.sztaki.hu/index.html>

5 <http://oai.kereso.sztaki.hu/>

6 Kézdi Tamás, Kovács, László, Micsik András, Pataki Máté:
[Elosztott digitális hangtárak a közösségi rádiózásért](#). Networkshop 2003, Pécs

7 <http://radio.sztaki.hu/>

8 Gergő Márton, Havasi László, Mátételki Péter, Micsik András, Kovács László, Szirányi Tamás:
[Képi és szemantikus keresőalgoritmusok kutatását támogató közösségi platform](#)
Networkshop 2012, Veszprém

9 <https://kopi.sztaki.hu/>

hoztuk létre, és azóta üzemeltetjük. Ez volt az első plágiumkereső Magyarországon, amely jól kezelte a magyar nyelvű szövegeket, és később 2011-ben, a világon elsőként, valósított meg jó minőségű fordítási plágiumkeresést¹⁰ is, amely például detektálni tudta, ha valaki a dolgozatába az angol nyelvű Wikipedia-ból magyarra fordított szövegrészeket illesztett be.

Metaadatok, dokumentumok, kapcsolt adatok

Az 1998-ban megjelent Dublin Core (DC) metaadatleíró rendszer forradalmasította a metaadatok világát. A Dublin Core megnyitotta az utat az eddig egymástól elkülönülve fejlődő leíró rendszerek (mint például a MARC), átjárhatóvá, egymásra leképezhetővé tételére, meghatározva a leglényegesebb, esszenciális metaadatok legszűkebb körét. Kezdetben úgy képzelték, hogy az egyes szakágak (pl. kereskedelem, könyvtárak, közigazgatás stb.) mind kialakítják a DC, majd később qDC alapú, de specializált változataikat, az ún. alkalmazási profilokat (application profile). Ezek szabványosítására, áttekintésére és kényelmes, grafikus szerkesztésére szolgált az azonos nevű EU projekt keretében létrehozott CORES rendszer, melynek fejlesztésében segédkeztünk, és melyet sokáig üzemeltettünk¹¹.

Az alkalmazási profilok szakmai körökben történő lassú terjedésével egyidejűleg megjelent egy alapvetőbb - persze bonyolultabb - koncepció, a Szemantikus Web, amellyel egyszerű metaadatrekordokat, tezauruszokat és logikai modelleket (ontológiákat) egyaránt le lehetett írni. A cél tehát az lett, hogy minden területen létrejöjjön a tudásreprezentációhoz szükséges ontológiák megfelelő halmaza. A Szemantikus Web később egyszerűbb és hatékonyabb formában kapcsolt adatok (linked data (LD), linked open data (LOD)) néven indult rohamos fejlődésnek.

A kapcsolt adatokat a már korábban említett CrossMedia projektben használtuk szemantikus keresésre, vagyis konkrét szóelőfordulások helyett egy tezaurusz részgráfja alapján kerestünk (pl. szinonimák, hiponimák bevonásával) megfelelő képeket.

2011-ben elindult a lod.sztaki.hu szolgáltatás az Intézetben, amely az OAI alapon összegyűjtött hazai kulturális adatokból előállított kapcsolt adathalmaz szolgáltatás, és több mint 11 millió egyszerű tényből (RDF triple) áll össze.

Végül 2016-ban, a COURAGE EU projekt indulásával sikerült egy a kezdetektől és alapjaiban is kapcsolt adatos (RDF alapú) adattár-megoldást létrehoznunk.

Teljes rendszerek fejlesztése és üzemeltetése

A fejlesztések mellett mindig is szerettük, ha a szoftvereink működnek és széles körben használják is azokat. Ezért fokozott gondot fordítottunk az üzemeltetésre és a folyamatos fenntartásra, támogatásra. Néhány komplexebb rendszert kiemelnénk a létrehozott szolgáltatás-portfóliónkból.

10 Pataki Máté: [Plágiumkeresés különböző nyelvek között](#). Workshop 2011, Kaposvár

11 Fülöp Csaba, Kovács László, Micsik, András:

[Metaadatsémák nyilvántartása szemantikus web alapon](#). Workshop 2004, Győr.



A már említett StreamOnTheFly európai rádióarchívum rendszer több csomópontból álló, elosztott rendszer volt, ahol az egyes csomópontok dönthettek arról, hogy mely más csomópontokkal lépnek partnerségre. Egy csomóponton belül több "rádióadó" archiválhatta műsorait, a szerkesztői jogok szabályozásával akár sorozatonként más-más ember végezhetette kényelmesen a hanganyagok, metaadatok feltöltését és/vagy szerkesztését. A felhasználók pedig saját, akár több órás egyéniesített (perszonalizált) rádióműsort állíthattak össze az archívumból és stream formájában élvezhették azt.

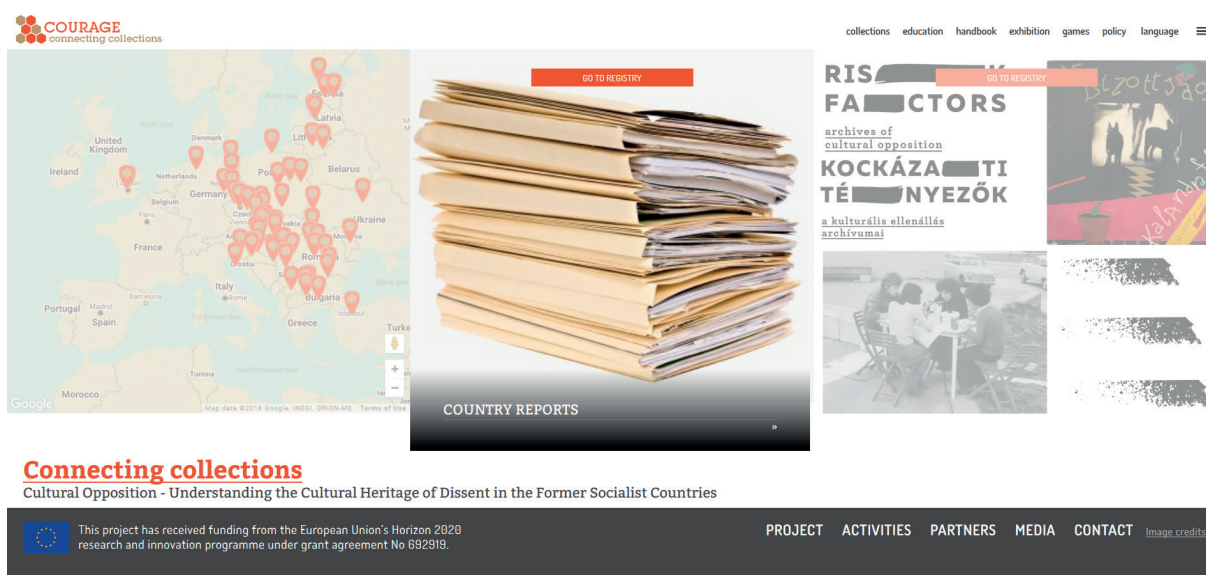
A KOPI plágiumkereső is összetett osztott rendszer, mivel a keresés összetett belső folyamatának lépéseit különböző célszerverek végzik. A SZTAKI Szótár is ide kívánczok, bár az csak szavak archívuma, de napi 80-100 ezer látogatót szolgál ki 8-10 szerver precíz együttműködésével. A szótár tartalmak tárolását pedig egy gráfadatbázis (neo4j) szolgálja ki. A SZTAKI Szótár kapcsán érdemes megjegyeznünk, hogy itt gondosan ügyeltünk a szótári kereső URL-ek hosszútávú megőrzésére. Ezért, ha valaki 1995-ben a szótár indulásakor berakott egy linket (keresést) egy szóra a honlapjára, ha erre ma 2019-ben ráklikkelnek, akkor a szótár ma is megadja a megfelelő fordítást.

The screenshot displays the MTMT2 editor interface. At the top, there are navigation tabs for 'MTMT2 szerkesztő', 'Rendszerezés', 'Fórum', 'Üzenetek', 'Cédlapok', and 'micsika'. Below these, a sidebar on the left contains a 'Közlemény' section with sub-items like 'Közlemény', 'Idézőkapcsolat', 'Szerző', 'Admin', and 'Intézmény'. The main area shows a list of publications with columns for 'Közlemény', 'Szerző', 'Admin', and 'Intézmény'. The list includes entries such as 'Kovács, Ádám Tamás ; Micsik, András' and 'Micsik, A ; Felker, T'. Each entry has a 'Kijelölés' column with checkboxes and a 'Szűkítés' button. The bottom of the interface shows a 'Listák' section with a 'Riportok és sablonok' button.

2. ábra: Az MTMT2 szerkesztői felülete

Az MTASZTAKI Elosztott Rendszerek osztályán fejlesztettük ki a Magyar Tudományos Művek Tára 2018. novembere óta éles üzemelésű 2-es (MTMT2) szoftververzióját¹². Ez volt pályafutásunk során eddig a legbonyolultabb és legnagyobb saját fejlesztésű szoftver-rendszer. Az MTMT2 a külvilággal egy REST API-n keresztül kommunikál, ezt használja a teljesen új nyilvános felület, amelyen bejelentkezés nélkül akár mobilon is lehet böngészni a szerzők, csoportok munkásságát, vagy az egyes témakörökben megjelent cikkeket. A szerkesztői felület ún. egyablakos Javascript alkalmazás, amely egy teljes professzionális munkakörnyezetet (workspace) ad a közlemények felviteli, kiegészítési stb. teljes körű adatkurátori feladatainak elvégzésére. Az MTMT2 rendszert több mint 60.000 felhasználó, az ország teljes kutatói szférája használja rendszeresen.

Végezetül a COURAGE¹³ EU projekt keretében az általunk létrehozott digitális archívum rendszerét említjük meg, amely a technikai fejlettségével emelkedik ki. Az adatokat RDF triple store tárolja a külön e célra készült COURAGE ontológia, mint adatséma alapján. A felhasználói felület nagy részét is kapcsolt adat konfiguráció, illetve SPARQL lekérdezések vezérlik. Az adatfelvitel során egyből létrejönnek a kétirányú adatkapcsolatok, relációk, melyek mentén a létrehozott szemantikus tudásgráf sokoldalúan böngészhető. A rendszergazda pedig menet közben a rendszer leállása nélkül fel tud venni új adatmezőket, vagy meg tudja változtatni a meglévőket (séma módosítás), azok megjelenési módját is beleértve.



3. ábra: A COURAGE projekt honlapja

¹² Micsik András, Pataki Balázs, Kovács László et al.: [A Magyar Tudományos Művek Tára 2.0 verziójának fejlesztése](#). Workshop 2016

¹³ Micsik András: [Besúgók és provokátorok - történelmünk kutatása kapcsolt adatokkal](#). Workshop 2017



Utószó

Az MTA SZTAKI DSD, a kutatóintézet Elosztott Rendszerek Osztálya 25 éven keresztül töretlenül dolgozott a digitális könyvtárak és archívumok kutatás-fejlesztése területén és létrehozott egy egyedülálló tudás és technológiai megoldás portfóliót, mely jelenleg a teljes magyar felhasználói közösség szolgálatára áll. A jövőben szeretnénk tovább foglalkozni a tudásreprezentáció és tudásfúzió felmerülő elméleti és gyakorlati problémáival. A portfólióban nem csupán korábbi megoldásaink, illetve az elsajátított technológiák, know-how-k stb. reprezentálnak értéket, hanem az az időközben az osztályon kialakított képességünk, mely bonyolult műszaki-tudományos problémák kezelését holisztikusan szemlélve egyidejűleg képes akár felfedező, alkalmazott kutatási és fejlesztési tevékenységeket kombináltan végezni. Az osztály munkatársai e képesség birtokában bátran vállalkoznak bármely új kihívás esetén egy lehetséges, a gyakorlatban működő, korszerű és tudományosan is értékelhető megoldást megtalálni.

A múzeumi adatrepresentáció és közzététel formái a Közgyűteményi Digitalizálási Stratégia keretében

Kómár Éva
Magyar Nemzeti Múzeum
komar.eva@mnm.hu

The museum data representation and publication included in the Public Collection Digitalisation Strategy The directions of the Hungarian cultural heritage digitalization are set by the Public Collection Digitalisation Strategy (KDS) until 2025, in which the most important goal is to provide digital content based on the common principles of the public collections for the broader society. One of the basic requirements of providing services, unifying heterogeneous resources, is that a data mapping should be available between data models used in public collection branches, and the forms and tools of data representations should be in synergy. Although different disciplines, like museology, usually use their own datamodels and publication schemas. This presentation examines, what steps should museums take, to meet the requirements in connection with providing content, defined in KDS. How can they get from current practises (just making publications focusing on descriptive data) to creative applications provided with extra informations, which satisfy user needs. Furthermore it will be presented, how can museum collection management systems connect to common platforms based on cooperation, like the Hungarian National Namespace Project or the search service of the National Repository Project.

Bevezetés

A hazai kulturális örökség digitalizálásának irányait 2025-ig a Közgyűteményi Digitalizálási Stratégia (KDS)¹ jelöli ki, melynek egyik legfontosabb célja, hogy a közgyűtemények egységes elvek alapján működő digitális tartalomszolgáltatást valósítsanak meg a társadalom minél szélesebb rétegei számára. A heterogén forrásokat egységesítő szolgáltatások biztosításának egyik alapfeltétele, hogy a közgyűteményi ágakban használt adatmodellek közelítsenek egymáshoz, az adatrepresentációk formái és eszközei szinergiában legyenek. Ugyanakkor a különböző diszciplínák, így a múzeumi terület is, jellemzően saját adatmodellekben és publikációs sémákban gondolkodnak.

Jelen írás azt vizsgálja, hogy a múzeumoknak milyen lépéseket kell tennie ahhoz, hogy a KDS-ben megfogalmazott tartalomszolgáltatási követelményeknek megfeleljenek. Hogyan juthatnak el a jelenlegi gyakorlattól – a csupán a nyilvántartási adatokra koncentrált publikálástól – a széles felhasználói igényeket kiszolgáló, plusz információkkal ellátott kreatív alkalmazásokig. Továbbá bemutatja, hogy a múzeumi gyűjteménykezelő rendszerek hogyan kapcsolódhatnak olyan együttműködésre épülő közös platformokhoz, mint a Magyar Nemzeti Névtér Projekt vagy a Nemzeti Adattár Projekt keresőszolgáltatása.

¹ Közgyűteményi Digitalizálási Stratégia (2017-2025). Hozzáférés: 2019.06.30.
https://www.kormany.hu/download/9/ac/11000/K%C3%B6zgy%C5%B1tem%C3%A9ny%C3%A1s%C3%A1si%20Strat%C3%A9gia_2017-2025.pdf

1. Digitalizálás a múzeumokban

Ha visszatekintünk az elmúlt tíz évre, láthatjuk, hogy a múzeumi szféra a közgyűjteményi digitalizálás területén egyre nagyobb ütemben igyekszik felzárkózni a könyvtári és a levéltári szakágakhoz. A múzeumok jelentős része elkezdett digitális nyilvántartást vezetni, megjelentek az intézményi webes katalógusok, virtuális kiállítások születtek, látványos kreatív megoldások a kiállítási terekben, és több múzeum sikeresen kapcsolódott valamelyik Europeana által indított projekthez. A hazai múzeumi ágazat digitalizálási gyakorlatára ugyanakkor jellemző, hogy a mai napig sem a metaadatokra, sem a digitális objektumokra nincsenek egységesen alkalmazott szabványok. A felsorolt eredmények többségükben szigetszerű vállalkozások, amelyekből nincs átjárás más muzeális intézmények, vagy más közgyűjteményi ágak szolgáltatásaihoz.

1.1 A digitalizálás célja

A digitalizálás célja a múzeumokban elsősorban a digitális nyilvántartás vezetése, az archiválási szándék és a kutatói kör kiszolgálása. Ahogy az 1. ábrán is látható a digitális tartalmak hasznosulását tekintve egyértelműen az állományvédelmi rangsor érvényesül a közzétételi igényekkel szemben.



1. ábra A digitális tartalmak hasznosulási modellje (forrás KDS, 2017)

Egy, az OMMIK² által 2017-ben végzett országos felmérés is bizonyítja, hogy a múzeumi adatbázisokban a képpel rendelkező rekordok száma rendkívül alacsony a szöveges rekordok arányához képest (17%), továbbá a publikus rekordok aránya az összes digitalizált rekordok állományához képest szintén elenyésző (11%).

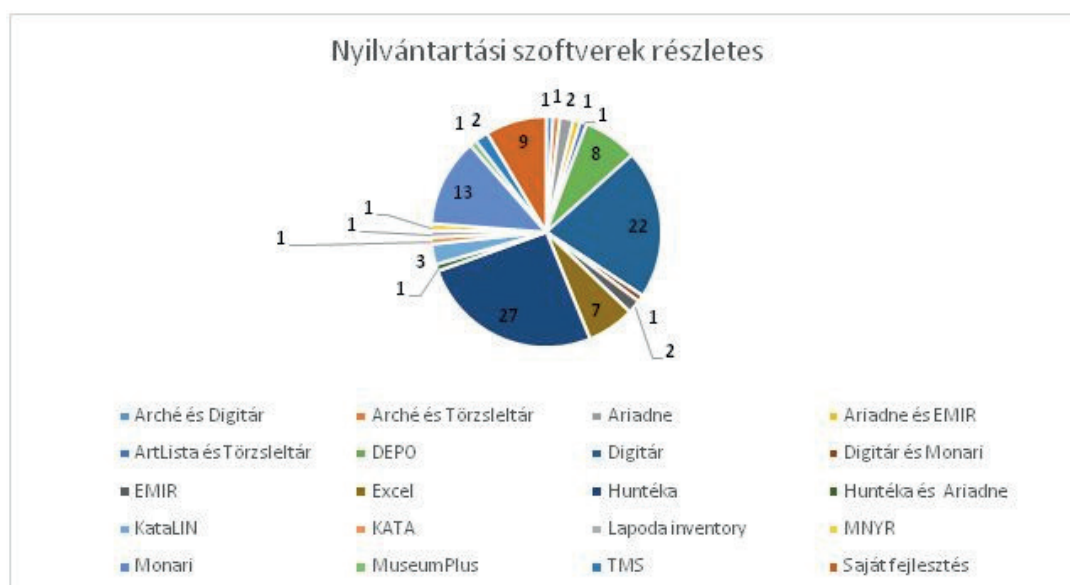
2 Országos Muzeológiai Módszertani és Információs Központ, a Magyar Nemzeti Múzeum főosztálya



2. ábra Szöveges és képes rekordok aránya a digitális nyilvántartásokban, 2017-es felmérés

1.2 Múzeumi adatbázisok

Bár egyre több múzeum használ már valamilyen elektronikus nyilvántartást, a metaadatok nagy része még valamilyen táblázatkezelő szoftverrel képződik. A gyűjteménykezelő rendszerek heterogén képet mutatnak, a „dobozos” termékek mellett sok a helyi fejlesztés is, ugyanakkor ezekben az adatbázisokban biztosított a strukturált adattárolás és a szabványos kimeneti formátum.



3. ábra Múzeumi nyilvántartási szoftverek, 2017-es felmérés

A nemzetközi gyakorlatban létezik speciálisan a múzeumok számára készült harvesztálási-adatcsere formátum (LIDO³), ám közös projektek hiányában hazai területen mégsem terjedt el a használata.

3 Lightweight Information Describing Objects. Hozzáférés: 2019. 06. 30.
<http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>



1.3. Digitális objektumok

A gyűjteményi digitalizálás első szakaszában az intézmények többsége a legértékesebb és a leginkább keresett műtárgyairól készített másolatot. A jó minőségű mesterpéldányok alkalmasak a nyomdai célú továbbforgalmazásra. A későbbi szakaszokban, a tömeges digitalizálással előállított példányok többsége azonban csupán a műtárgyak beazonosítására szolgál, a digitális objektumok közül kevés a jó minőségű, élményszerű felhasználást biztosító médiatartalom. A bevételkieséstől tartva a múzeumok nem szívesen publikálják online térben a műtárgyak digitális másolatait, főleg nem nagy felbontásban. Leginkább a tartalmakhoz való hozzáférést tekintve jelentős a múzeumi terület lemaradása a többi ágazathoz képest: 2016-ban az összes digitalizált múzeumi állományból csak 9% volt hozzáférhető a felhasználók számára.⁴

1.4 Leíró metaadatok

A hazai múzeumi területen jelenleg nincs hivatalosan elfogadott metaadatszabvány. A gyűjteménykezelő rendszerekben található leíró metaadatokat a ma már igencsak elavultnak tekinthető 20/2002. (X.4.) NKÖM rendelet⁵ alapján képzik a muzeológusok. A rendeletnek megfelelően rögzített adatok csak a nyilvántartáshoz szükséges információkat tartalmazzák. A publikus metaadatok köre ennél még szűkebb, sokszor hiányoznak a részletes leírások, így a rekordok nem elég informatívak.

A fentiekből következik, hogy az online térben elérhető múzeumi tartalmak az átlag felhasználó igényeit nem elégíti ki. Hiába tesszük közzé egy fibula vagy egy mángorló rekordját, olyan adatmezőkkel, amelyek a beazonosítást szolgálják (megnevezés, anyag, technika, méret), ennyi információból még nem derül ki, hogy mi is az a tárgy és mire használták.



4. ábra Egy fibula és egy mángorló rekordja a MuseuMap portálon
<http://resolver.museumap.hu/object/5859436>
<http://resolver.museumap.hu/object/5984036>

4 Kögyűjteményi Digitalizálási Stratégia (2017-2025). Hozzáférés: 2019.06.30.
https://www.kormany.hu/download/9/ac/11000/K%C3%B6zgy%C5%B1tem%C3%A9ny%C3%A1s%C3%A1si%20Strat%C3%A9gia_2017-2025.pdf

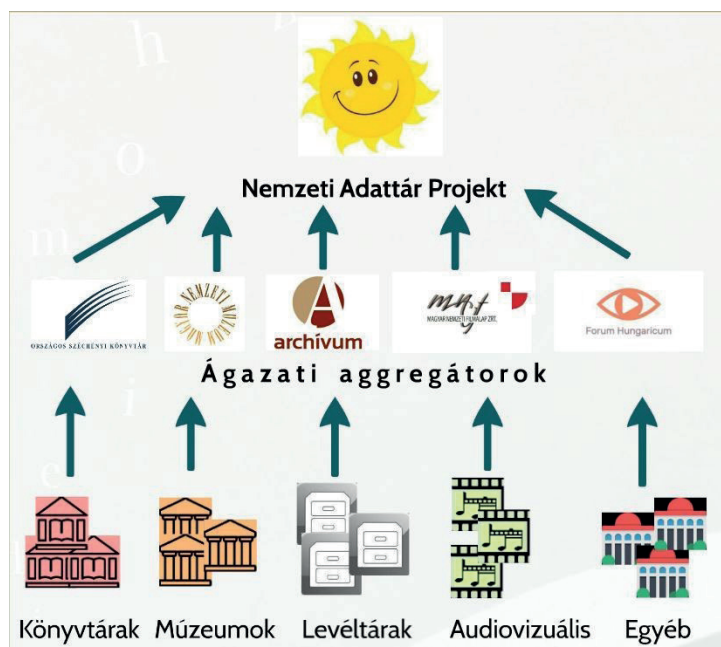
5 20/2002. (X. 4.) NKÖM rendelet a muzeális intézmények nyilvántartási szabályzatáról.
 Hozzáférés: 2019.06.30. <https://net.jogtar.hu/jogszabaly?docid=a0200020.nkm>

2. Új típusú elvárások

Az előzőekben leírtak azért is jelentenek problémát, mert a KDS a közzétételt, a közzétételi célú digitalizálást hangsúlyozza, ennek érdekében jó minőségű, az átlag felhasználó számára is informatív digitális tartalmak létrehozását várja el. Szolgáltatásorientált szemléletet kíván megvalósítani, melynek során a közgyűteményi digitális tartalmak elsősorban az oktatás/köznevelés, de a továbbiakban a turizmus, a rekreáció területén is hasznosulnak, végső soron hozzájárulva az állampolgárok jobb életminőségéhez.

Ahhoz, hogy a múzeumok meg tudjanak felelni a KDS-ben megjelölt közzétételi kívánalmaknak, előfeltétel, hogy rendelkezzenek megfelelő mennyiségű és minőségű digitálisan elérhető tartalommal, és ezekhez hozzáférést is biztosítsanak különböző platformokon. A tartalomelőállítást tekintve alapvető szemléletváltásra van szükség, azaz az intézményi digitális gyűjteményfejlesztési és szolgáltatásfejlesztési stratégia és gyakorlat megújítására, valamint egységes módszertan használatára. Ez utóbbihoz ad majd segítséget a KDS keretében elkészülő Fehér Könyv⁶, amely a digitalizálás teljes munkafolyamatához nyújt módszertani útmutatót a közgyűjtemények számára.

A KDS szolgáltatási modellje háromszintű. Az első a primér szint, az intézmények szintje, ahol a tartalomelőállítás történik, a második, az ágazati aggregátorok szintje, ahol begyűjtik az intézményektől a tartalmakat, összefésülik, adatgazdagítják, és így továbbítják a Nemzeti Adattár Projekt (NAP) összközgyűjteményi keresőjébe, amely a jövőben készül el.



5. ábra A Nemzeti Adattár Projekt háromszintű modellje

6 Az 1175/2018. (III.28.) Korm.hat.3.pontjában a Nemzeti Adattár Projekt megvalósítása érdekében kidolgozott módszertani segédlet



A múzeumi terület központi intézménye a Magyar Nemzeti Múzeum, amely már 2015-ben elindította ágazati aggregációs szolgáltatását. A szolgáltatás működési modellje az Europeana, mint az európai szintű aggregáció működési modelljének hazai adaptációja. A tartalomszolgáltatók háromféleképpen juttathatják el a rekordjaikat a MuseuMap számára: intézményi adatbázisokból, OnlineData webes alapú beviteli rendszerből vagy országos múzeumokon keresztül. Az aggregáció harvesztálási sémája a múzeumi területen nemzetközileg is elterjedt LIDO formátum, de a MuseuMap bármilyen rendszerből képes szabványos formában átadott rekordokat fogadni. A múzeumi aggregációs szolgáltatás publikus felülete a MuseuMap portál (www.museumap.hu), ahol jelenleg 30 múzeum több mint 300 ezer rekordja között böngészhet a látogató.

A KDS keretében 2019-ben indított hároméves Mintaprojektnek legfőbb célkitűzése, hogy a múzeumoknál jó minőségű, a felhasználói igényeket széles körűen kielégítő digitális tartalmak készüljenek, amelyek az aggregátori szolgáltatásokban továbbhasznosulhatnak.

2.1 Minőségi digitális objektumok

A magas színvonalú közzétételi szolgáltatásokhoz korszerű technológiával előállított mesterfájlok szükségesek, amelyekből további felhasználásra alkalmas, élményszerű szolgáltatási példányokat lehet generálni. A digitális objektumok közzétételi elvárásaira vonatkozóan a Fehér Könyv az Europeana Publikációs Keretrendszer Digitális tartalmakra vonatkozó útmutatóját⁷ ajánlja, amely minden médiatípusra négy közzétételi szintet határoz meg, és ezekhez paramétereket rendel. Az első szint, amikor a tartalmainkkal csak a keresőmotorokat szeretnénk kiszolgálni, a második szint, amikor már alkalmasak a tartalmak tematikus gyűjtemények vagy virtuális kiállítások létrehozására, a harmadik és a negyedik szint pedig a kreatív, nem kereskedelmi (pl. oktatási célú) és kereskedelmi célú (pl. kreatív ipar) újrafelhasználási igényeit írja le. Az utóbbi két esetben a jogi szabályozásban vannak különbségek.

Képfájlok esetében:

- 1., 0,1 megapixel, ~ 400 pixel széles előnézeti kép – (keresőmotorok)
- 2., 0,5 megapixel kb. ~ 800 pixel széles képfájl (tematikus gyűjtemények, virtuális kiállítások)
- 3., Minimum 1200+ pixelszéles ajánlott (nem kereskedelmi célú újrafelhasználás; pl. oktatás)
- 4., Minimum 1200+ pixel széles ajánlott (kereskedelmi célú újrafelhasználás, kreatív ipar)

2.2. Minőségi leíró adatok

A digitális objektumokhoz fűzött, szöveges információk egyik részét a műtárgy metaadatai jelentik. Az informatív tárgyleíráshoz elengedhetetlen, hogy a nyilvántartási rendelet mezőiből minél többet, minél részletesebben adjanak

⁷ Europeana Publishing Framework. Hozzáférés: 2019.06.30.
<https://pro.europeana.eu/post/publishing-framework>

meg a muzeológusok. Másrészt az újrafelhasználás szempontjából fontos, hogy megszülessenek azok a felhasználói igényekhez alakított narratív leírást tartalmazó részek, amelyek további kontextusba helyezik a műtárgyakat és megkönnyítik a befogadást. Már intézményi szinten elindulhatna saját vagy külső névterek használata az adatbázisban rögzített entitások (személy, hely, fogalom stb.) beazonosítására. A strukturált formában tárolt, karbantartott adatok eredményesebben hasznosulnak további, közös szolgáltatásokban is.

2.3 Adatgazdagítás

A digitális rekordok plusz információval való ellátását a KDS az aggregátorok szintjére helyezi át. A múzeumi ágazat aggregátor intézménye a Magyar Nemzeti Múzeum⁸, amelynek feladata, hogy az intézmények rekordjait begyűjtse, egységesítse, további információkkal gazdagítsa, és így továbbítsa a tartalmakat a NAP keresőrendszerébe.

Az adatgazdagítást tekintve a múzeumi terület is a klasszikus értelemben vett négy dimenziójú adatgazdagítást veszi alapul: a személyre, helyre, időre, fogalomra történő szemantikai dúsítást.

Az adatgazdagításhoz leggyakrabban használt források:

Személy

VIAF: The Virtual International Authority File	https://viaf.org/
ULAN: Union List of Artist Names	http://www.getty.edu/research/tools/vocabularies/ulan/
DBpedia	https://wiki.dbpedia.org/
wikidata	https://www.wikidata.org/
Petőfi Irodalmi Múzeum személynévtére	https://opac-nevter.pim.hu/
FOAF: Friend of friend ontology	http://www.foaf-project.org/

Idő

Time Ontology in OWL	https://www.w3.org/TR/owl-time/
PeriodO	http://perio.do/en/
DBpedia	https://wiki.dbpedia.org/
wikidata	https://www.wikidata.org/

8 A Magyar Nemzeti Múzeum 2015-től működteti ágazati aggregációs szolgáltatását MuseumMap néven. www.museummap.hu



Hely

GeoNames	https://www.geonames.org/
TGN: The Getty Thesaurus of Geographic Names,	http://www.getty.edu/research/tools/vocabularies/tgn/index.html
Pleiades	https://pleiades.stoa.org/
Fogalom	
AAT: Art&Architecture Thesaurus	http://www.getty.edu/research/tools/vocabularies/aat/
DBpedia	https://wiki.dbpedia.org/
wikidata	https://www.wikidata.org/
GEMET: General Multilingual Environmental Thesaurus	https://www.eionet.europa.eu/gemet/en/themes/
Iconclass	http://www.iconclass.org/

2.4 Kapcsolódás a Magyar Nemzeti Névtér szolgáltatáshoz

Mivel a múzeumi nyilvántartási gyakorlatban nem hangsúlyos a besorolási állományok gondozása, például megengedett a személynevek és testületi nevek (pl. fotóműhely, szerkesztőség) egy szótárban való kezelése, a Nemzeti Névtér használata komoly előrelépést jelentene a múzeumok számára, hogy a gyűjteményeikhez kapcsolódóan pontosabb és bővebb információszolgáltatást valósítsanak meg.

A Nemzeti Névtér hasznosításának lehetőségei:

- A névtér névjegyadatainak használata a helyi rendszerekben szereplő nevek egyértelmű azonosításához
- Hivatkozás a névtérre: URI-k beemelése a helyi rendszerek besorolási állományaiba
- A Nemzeti Névtérben található adatok, névrekordok honosítása a helyi rendszerek besorolási állományaiba

A múzeumi szféra nemcsak felhasználóként, hanem adatszolgáltatóként is közreműködhet a Nemzeti Névtér-szolgáltatásban. A helynevek pontos megjelölése kiemelt fontosságú több múzeumi szakanyag esetén, ezért a régészeti, a néprajzi vagy a természettudományi gyűjtemények lelőhelyre/gyűjtőhelyre vonatkozó információi bővebb adatkörrel rendelkeznek, mint a Földrajzi Névtér névjegyadatai.

Lehetőségek a múzeumi adatok szolgáltatására a Földrajzi Névtérbe:

- A régészeti lelőhely-nyilvántartás integrálása a Nemzeti Névtérbe
- További lehetőség a Magyar Nemzeti Múzeum Régészeti adatbázisának⁹ (Ariadne) összekötése a Nemzeti Névtérrel.
<https://archeodatabase.hnm.hu/hu>
- A Magyar Természettudományi Múzeumban épített gyűjtőhely-listák integrálása a Nemzeti Névtérbe

9 Régészeti adatbázis. Hozzáférés: 2019.06.30. <https://archeodatabase.hnm.hu/hu>

Amennyiben a szakági aggregátorok egységesen alkalmazzák a Nemzeti Névtér-azonosítókat, lehetővé válik, hogy a heterogén állományok ellenére kapcsolatok épüljenek fel a különböző közgyűteményekben található entitások között, az így létrejövő szemantikus háló segítségével pedig további ráépülő, interdiszciplináris szolgáltatások valósulhatnak meg.

2.5 Oktatási célú adatgazdagítás

A KDS szempontjából kiemelt fontosságú a közgyűteményekben keletkező digitális tartalmak újrahasznosítása az oktatás területén. A köznevelésben résztvevők számára a kereshetőséget és a hozzáférést a leíró adatok bővítése biztosítaná a köznevelés tartalmi szabályozóinak fogalomkészlete alapján. A kategóriaképzés elve az altémák szerinti csoportosítás, a nevelés szakaszaihoz és a tartalmi szabályozókhoz (Nemzeti alaptanterv, kerettanterv) való hozzárendelés. A fogalomkészlet megadását a múzeumi terület is az oktatási szereplőktől várja.

Új technológia alkalmazása

A digitális átállás egyik következménye, hogy a múzeumi területen is egyre nagyobb számú adat keletkezik, aminek a kezeléséhez nem áll rendelkezésre elég szakember. Az újonnan keletkező tartalmak mellett folyamatos feladat a már meglévő állományok gondozása is. A tartalmak gyorsabb és hatékonyabb feldolgozására a mesterséges intelligencia bevonása jelent megoldást. A múzeumi adatbázisokban található rekordok digitális objektumait (médiatartalom) és a hozzájuk tartozó szöveges információkat elemezhetjük tanuló algoritmusok segítségével, külső források (szótárak, teauruszok, névterek) figyelembevételével. A tartalomalapú elemzésbe beletartozhat: az arcfelismerés, színelemzés, formák, helyszínek, korok vagy stílusirányzatok felismertetése. Az új technológia előnye, hogy a sokkal több időt igénylő manuális munkavégzést kiváltva gyorsabb és eredményesebb lehet a tartalmak feldolgozása, valamint a gépi intelligencia addig nem ismert kontextusokat is feltárhat a műtárgyak között.

A hazai múzeum szakma tehát jelentős változások előtt áll, amennyiben a KDS-ben megfogalmazott szolgáltatási igényeket figyelembe kívánja venni. A digitalizálási gyakorlat szemléletének megújítása és a korszerű technológia integrálása egyaránt segíthet, hogy a múzeumi ágazat hatékonyabban járuljon hozzá a kultúra átörökítéséhez és a digitális kulturális örökség társadalmi hasznosulásához.

A nyílt hozzáférés kérdése a hivatalos statisztikában Some Aspects of Open Access and Official Statistics

Lencsés Ákos

ORCID: 0000-0003-4461-1105

MTA Könyvtár és Információs Központ ELSZ Titkárság (grid.496758.1)

ELTE BTK Könyvtár- és Információtudományi Intézet

Könyvtártudományi Doktori Program (grid.5591.8)

The past decades have revealed the importance and want of Open Access in the academic sphere. The recent steps, e.g. Plan S show that research funds are all aware of Open Access and are ready to step forward to a fully OA world. The study focuses on how OA appears in the landscape of official statistics – while academic sphere is not the main focus in this field, still official statistics is relevant for the academics in many ways.

The study analyses the publications – books, yearbooks, journals – of European official statistics regarding their availability. With a few exceptions, national statistical offices are all committed to OA. There are future challenges for official statistics as metadata and other tools that help to integrate official yearbooks and other publications into the research workflow are poorly represented.

Keywords: statistical yearbook, open access, official statistics, statistical dissemination

Bevezetés

A tudományos világban az elmúlt évtizedekben egyre nagyobb hangsúly kapott a nyílt hozzáférés kérdése. A legújabb fejlemények, pl. a Plan S létrejötte azt mutatják, hogy a kutatásfinanszírozó testületek a jövőt a teljes körű nyílt hozzáférés jegyében képzelik el. Érdekes megvizsgálni, hogy más – a kutatás szempontjából releváns, de nem elsősorban kutatási anyagokat előállító – területeken hogyan jelenik meg a nyílt hozzáférés kérdése.

Az előadás a hivatalos statisztika évkönyveit és folyóiratait vizsgálja, elsősorban a nyílt hozzáférés szempontjából. A 2017 első félévében végrehajtott felmérés segít megállapítani, hogy a hivatalos statisztikai intézmények mennyire elkötelezettek a nyílt hozzáférés irányába. A vizsgálat óta természetesen több esetben történt változás, így például a *Statisztikai Szemle* is nyílt hozzáférésű folyóirattá vált. Jelen írásban a 2017 első félévének megfelelő állapotokat vizsgáljuk.

Statisztikai évkönyvek Európában

Valamennyi nemzeti statisztikai hivatal számára a statisztikai évkönyv a leginkább reprezentatív kiadvány. Az évkönyvek vizsgálatából több következtetés is levonható a hivatalok működésére vonatkozóan.¹ Az alábbi felmérésben az európai országok gyakorlatát mutatjuk be a statisztikai évkönyv szabad hozzáférhetőségére

¹ Lencsés Ákos. „A Magyar statisztikai évkönyv szerkezetének változásai 1872 és 2016 között.” *Statisztikai Szemle* 95, 11–12. sz. (2017): 1144–1158. DOI: [10.20311/stat2017.11-12.hu1144](https://doi.org/10.20311/stat2017.11-12.hu1144)

vonatkozóan. Az adatgyűjtés során nem vettük figyelembe a statisztikai zsebkönyveket, kizárólag az egyes országok átfogó statisztikai évkönyvére koncentráltunk. Így nem okoz torzítást a vizsgálatban, ha az ország statisztikai zsebkönyve vagy más rövid, statisztikai jellegű ismertetést tartalmazó kiadványa nyílt hozzáféréssel elérhető, az átfogó statisztikai évkönyv azonban csak embargóval tekinthető meg.

Több országot kizártunk a vizsgálatból. Ennek oka méretük vagy az, hogy a közelmúlt politikai változásai miatt nem értelmezhető még relevánsan a hivatali intézményrendszer minden irányvonala. Ennek megfelelően a torzítások kiküszöbölése érdekében nem tartalmazza a táblázat Andorra, Bosznia-Hercegovina, Ciprus, Koszovó, Liechtenstein, Luxemburg, Macedónia, Málta, Moldova, Monaco, Montenegró, San Marino, Szerbia, Ukrajna és Vatikán adatait. A felmérésben így Albánia, Ausztria, Belgium, Bulgária, Csehország, Dánia, Észtország, Fehéroroszország, Finnország, Franciaország, Görögország, Hollandia, Horvátország, Írország, Izland, Lengyelország, Lettország, Litvánia, Magyarország, Nagy-Britannia, Németország, Norvégia, Olaszország, Oroszország, Portugália, Románia, Spanyolország, Svájc, Svédország, Szlovákia és Szlovénia kiadványai szerepelnek. Az adatok forrása minden esetben az ország nemzeti statisztikai hivatalának honlapja volt. Az adatgyűjtés során nagy segítséget jelentett a Központi Statisztikai Hivatal Könyvtár által üzemeltetett Digistat adatbázis.²

A felmérésben az alábbi kérdéseket vizsgáltuk:

- megjelenik-e nyomtatásban az ország statisztikai évkönyve;
- megjelenik-e online az ország statisztikai évkönyve;
- az online változat nyílt hozzáférésű-e;
- elérhető-e a kiadványok archívuma;
- mennyi a nyomtatott kiadványok listaára.

Az adatok alapján kijelenthető, hogy a statisztikai hivatalok túlnyomó része elkötelezett a nyílt hozzáférés irányában. Ugyanakkor az intézmények nagy része fontosnak tartja a nyílt hozzáférésű online kiadványok mellett a nyomtatott évkönyvek megjelentetését is. A nyomtatott kiadványok ára nagy változatosságot mutat, és nem minden esetben mondható el, hogy a nyugat-európai országok statisztikai évkönyvei drágábbak lennének a kelet-európai hivatalok kiadványainál.

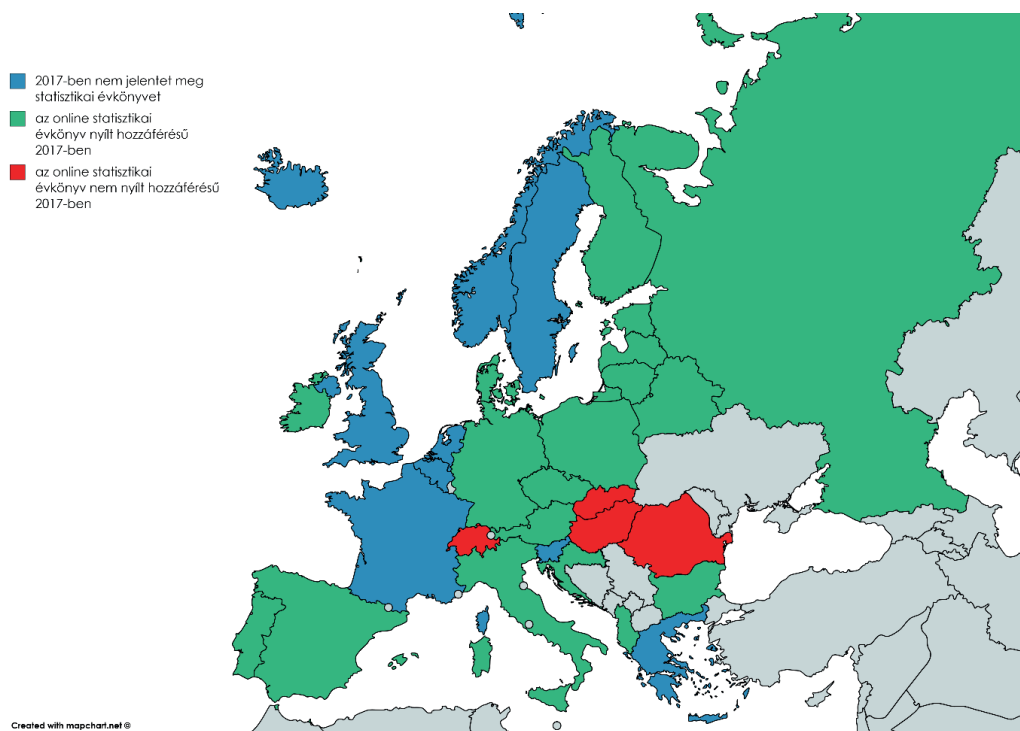
A vizsgálatban szereplő 31 ország közül 21 megjelenteti nyomtatásban is az átfogó statisztikai évkönyvet. A többi 10 ország egy része átfogó statisztikai évkönyvet nem jelentet meg, de kisebb terjedelmű, gyakran ismeretterjesztő kiadványt nyomtatásban is közread. A nyomtatott évkönyveket megszüntető hivatalok általában az évkönyvek szerkesztését is megszüntették. Egyetlen kivétel Írország, ahol továbbra is előállítják a kötet formájára tördelt PDF-változatot a statisztikai évkönyvből, azonban nyomtatásban már nem jelentetik meg. A nyomtatott változatot (illetve általában a statisztikai évkönyvet) elsőként Franciaországban és Nagy-Britanniában szüntették meg, 2005-ben. Ennek oka valószínűleg

² Digistat. http://konyvtarksh.hu/index.php?s=db_digistat [2019. június 20.]



nemcsak az informatikai fejlődés, hanem a két ország intézménystruktúrájának sajátossága. Mindkét országban több hivatal foglalkozik a statisztika előállításával és publikálásával, és ezek az intézmények gyakran önállóan jelentették, illetve jelentetik meg kiadványaikat.³ Az észak-európai országok elsőként döntöttek úgy, hogy az adatok közreadását egyáltalán nem kötik kiadványokhoz. A 2010-es években megszűntek a norvég és a svéd statisztikai évkönyvek. Ez a változás azonban már nem az intézményi környezetnek köszönhető, hanem tudatos döntés a kiadványstruktúra leépítése mellett az online, adatbázisokban történő megjelenítés javára. Hasonló fejlődést járt be a belga és a holland statisztikai hivatal kiadványpolitikája is.

Írország statisztikai hivatala egyetlen intézményként kizárólag online teszi közzé az évkönyvet. Svájc pedig egyetlen országgént kizárólag nyomtatásban forgalmazza a statisztikai évkönyvet – nem teszi lehetővé a PDF-ben történő letöltést vagy megvásárlást sem. Mindössze négy olyan hivatalt találtunk, amely a létező online változatot nem teszi szabadon hozzáférhetővé. Három esetben (Magyarország, Szlovákia, Románia) nyolc-tizenkét hónapos embargóval lehet szabadon elérni a kiadványokat, de természetesen megvásárolható a hozzáférés az embargó időszakában is. Svájc az egyetlen ország, amely se szabadon hozzáférhető, se fizetős módon nem biztosít hozzáférést a statisztikai évkönyv online változatához. A vizsgált országok közül 19 ország teszi a megjelenés idejében nyíltan hozzáférhetővé a statisztikai évkönyvet, tehát az európai statisztikai hivatalok túlnyomó része elkötelezett a nyílt hozzáférés iránt.



1. ábra A statisztikai évkönyvek és a nyílt hozzáférés helyzete Európában

3 Rózsa Dávid, Kalmár Csilla, Lencsés Ákos. „A statisztika és a digitális átállás.” = *Tudományos és Műszaki Tájékoztatás* 60, 2. sz. (2013): 47–65

Az archív statisztikai évkönyveket mindössze négy hivatal nem teszi hozzáférhetővé (Albánia, Ausztria, Franciaország, Izland). Ez az arány elsőre jónak nevezhető. Azonban mindössze 6 olyan hivatalt sikerült azonosítani, amely hangsúlyt fektet arra, hogy a statisztikai évkönyvek digitális korszakot megelőző példányai is online hozzáférhetőek legyenek (Norvégia, Svájc, Lettország, Görögország, Németország, Szlovénia). Természetesen a hat ország mellett egyes további országokban is elérhető az archív tartalom, de nem a hivatal saját felületén. A statisztikai hivatalok túlnyomó többsége nem áldoz erőforrást arra, hogy a digitális világ előtti kiadványok online elérhetőek legyenek. 17 esetben csak az 1990-es vagy 2000-es évektől hozzáférhetőek az eleve digitálisan létrejött kiadványok. 4 további esetben is csak az elmúlt 30 esztendő dokumentumai érhetőek el.

Érdekes a hivatalok egy részének nyomtatott kiadványokhoz való ragaszkodása annak fényében, hogy egyes szerzők már a kétezres évek elején a nyomtatott statisztikai kötetek eltűnését jósolták.⁴ Megmaradásuk oka talán nem is praktikusságukban, hanem reprezentatív szerepükben rejlik. A német statisztikai hivatal egyik felmérésében a statisztikai adatokat rendszeresen felhasználók 85%-a ismerte az átfogó statisztikai évkönyvet, de csak 8%-uk az ágazati kiadványokat.⁵

Vizsgáltuk a nyomtatott statisztikai évkönyvek árát is. Az összehasonlíthatóság érdekében a Magyar Nemzeti Bank 2017. évi középárfolyamával számolva próbáltunk durva becslésre alkalmas árakat előállítani. 18 évkönyv esetén sikerült a hivatal által meghatározott fogyasztói árat azonosítani. Az évkönyvek átváltás utáni becsült ára 19 és 120 euró között változik. Legolcsóbb a cseh és az észt (19 és 20 euró), legdrágább a svájci (120 euró) statisztikai évkönyv. Az árakat vizsgálva nem érződik a nyugati és északi országok gazdasági dominanciája: a lengyel és a német egyaránt 71 euróba kerül, az olcsóbb kötetek között megtaláljuk a spanyol vagy a portugál statisztikai évkönyvet is. A magyar statisztikai évkönyv a vizsgált kötetek közül a harmadik legolcsóbb kiadvány, csak az észt és a cseh évkönyv olcsóbb nála.

Külön kategóriát képvisel az egyes európai statisztikai évkönyvek közül az Eurostat statisztikai évkönyve (*Europe in Figures*). Ennek utolsó nyomtatásban megjelent tárgyéve 2012.⁶ Azóta a kiadvány online módon érhető el, azonban nem egyszerre megjelenő, változásmentes kiadványról van szó. Az adatok folyamatos változása és frissítése inkább adatbázisra vagy a Wikipédiára emlékeztető felületet eredményez. Könyvészeti szempontból a megoldás nehezen értelmezhető monográfiaként vagy évkönyvként – bár az oldal nevében még őrzi a *yearbook* kifejezést.

Afentvizsgálatalapján kijelenthető, hogy nincsen egységes gyakorlat az egyes országok statisztikai hivatalai részéről az átfogó statisztikai évkönyvek megjelentetésére. A

4 Cannon, San, Marc Rodriguez. „Survey of On-line Data Dissemination Practices for Government and International Statistics.” *The Statistics Newsletter* 49. sz. (2010): 5–8.

5 Pfeiffer, Annette. „Das Image des Statistischen Bundesamtes.” *Wirtschaft und Statistik* 12. sz. (2001): 981–987.

6 Rózsa Dávid, Kalmár Csilla, Lencsés Ákos. „A statisztika és a digitális átállás.” – *Tudományos és Műszaki Tájékoztatás* 60, 2. sz. (2013): 47–65



megoldások sokfélék a kiadványalapú adatközlés elhagyásától, a kizárólag online formátumban való megjelentetésen át, a hagyományos nyomtatott kiadványokon keresztül a fizetős és nyílt hozzáférésű online kiadványokig. Európában érezhető az észak- és nyugat-európai statisztikai hivatalok elkötelezettsége a nyílt hozzáférés felé, ami gyakran párosul azzal, hogy elhagyják a kiadványalapú adatközlést.

A hivatalos statisztika folyóiratai

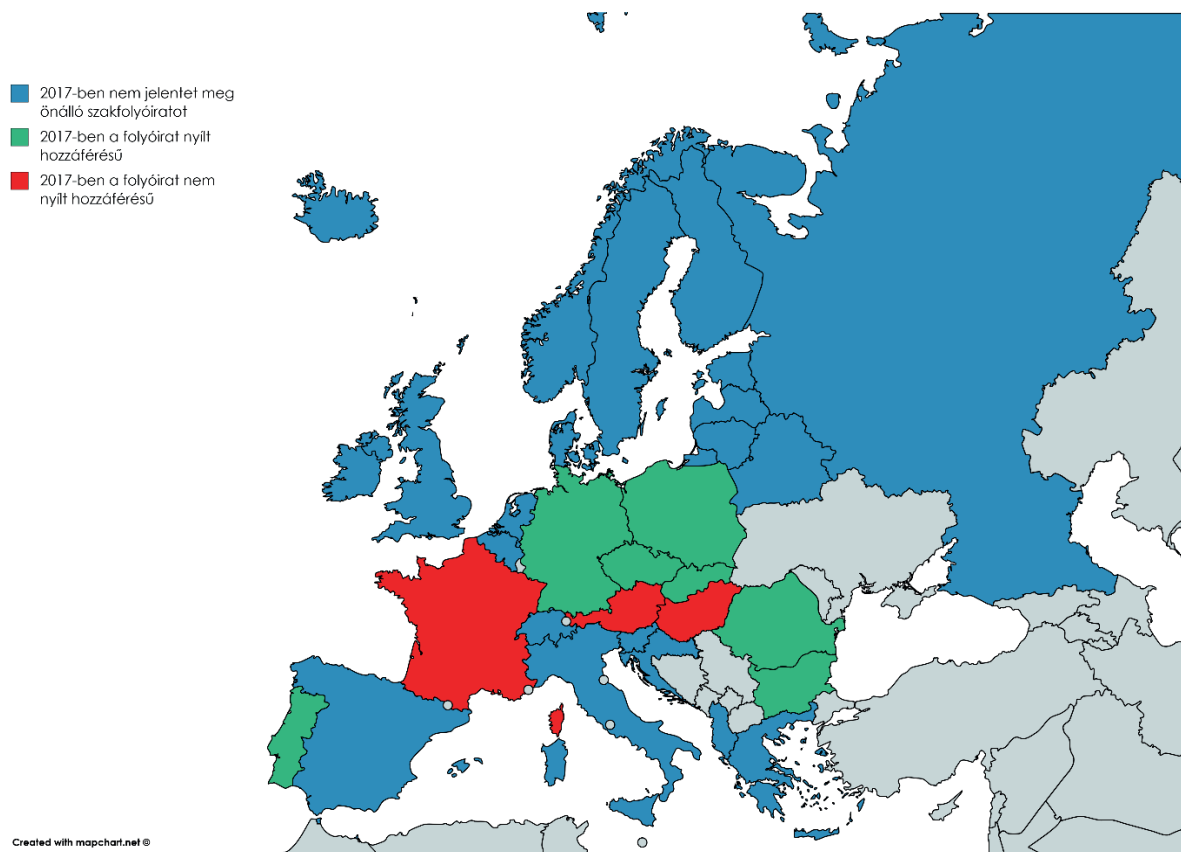
A vizsgálat folytatásaként a statisztikai évkönyvek mellett az egyes hivatalok által megjelentetett folyóiratok nyílt hozzáférését mértük fel. A vizsgálatnak a statisztikai hivatalok nyílt hozzáférés felé való elkötelezettségének megállapítása volt a célja, nem pedig általánosságban a statisztikai folyóiratok nyílt hozzáféréseinek felmérése. Így a tudományos kiadók vagy nem hivatalos szakmai szervezetek által kiadott folyóiratok nem kerültek az összegzésbe. A vizsgálatba ezért azokat a folyóiratokat vontuk be, amelyeket a statisztikai hivatalok önállóan jelentetnek meg. Nem kerültek be az elemzésbe a szakkiadók vagy szakmai szervezetek által megjelentetett kiadványok, így például az osztrák statisztikai társaság által kiadott *Austrian Journal of Statistics* vagy a Wiley kiadó által megjelentetett brit Royal Statistical Society folyóiratcsaládja. Ezeket a folyóiratokat abban az esetben sem vontuk be a vizsgálati körbe, ha nyílt hozzáféréssel elérhetők, így a svéd statisztikai hivatalnak a De Gruyter kiadó gondozásában megjelenő *Journal of Official Statistics* folyóirata sem szerepel a kutatásban. Bekerültek a vizsgálatba azonban a statisztikai hivatal és állami egyetemek, kutatóintézetek közös kiadásában megjelenő folyóiratok, például a cseh statisztikai hivatal és a cseh közgazdaság-tudományi egyetem által együtt jegyzett *Statistika* folyóirat.

A felmérésbe végül tíz ország tizennégy folyóiratát vontuk be: Ausztria (*Statistische Nachrichten*), Bulgária (*Statistika*), Csehország (*Demografie; Statistika*), Franciaország (*Population; Économie et Statistique*), Lengyelország (*Wiadomości Statystyczne*), Magyarország (*Statisztikai Szemle; Demográfia; Területi Statisztika*), Németország (*Wirtschaft und Statistik*), Portugália (*Revstat*), Románia (*Romanian Statistical Review*), Szlovákia (*Slovenská Štatistika a Demografia*). A kutatásban a 2017. első negyedév szerinti állapotnak megfelelően vizsgáltuk a folyóiratok nyelvét, periodicitását, nyílt hozzáférését és az archív számok elérhetőségét.

A hivatalok által megjelentetett folyóiratok jelentős része nemzeti nyelven jelenik meg, nem elhanyagolható azonban az angolul megjelenő címek száma sem. Általában azok a folyóiratok tartoznak ez utóbbiak közé, amelyeket egy-egy statisztikai hivatal valamelyik felsőoktatási intézménnyel vagy kutatóintézettel közösen jelentet meg. Így a cseh statisztikai hivatal és a cseh közgazdaság-tudományi egyetem *Statistika* folyóirata kizárólag angolul jelenik meg, a francia *Population* pedig egyedi módon francia és angol kiadásban is elérhető.

Az európai hivatalos statisztika intézményei által kiadott folyóiratok túlnyomó része nyílt hozzáféréssel megtalálható az interneten. A vizsgált címek közül egyedül a *Statisztikai Szemle*, a *Population* és a *Statistische Nachrichten* nem érhető el nyílt hozzáféréssel. Ez utóbbi különösen meglepő, mivel az osztrák statisztikai évkönyv

embargó nélkül elérhető a hivatal honlapján. Bár a vizsgálat szempontjai között nem szerepelt, de valamennyi folyóirat megjelenik nyomtatásban is. Franciaországot leszámítva minden szakfolyóiratot megjelentető európai statisztikai hivatal a statisztikai évkönyvet is megjelenteti nyomtatásban. Ugyanakkor az évkönyvet nyomtatásban nem megjelentető hivatalok általában online megjelenő folyóiratot sem üzemeltetnek.



2. ábra A hivatalos statisztika folyóiratai és a nyílt hozzáférés helyzete Európában

Az archiv számok az évkönyvekhez hasonlóan általában az online változat megjelenésétől érhetők el. A korábbi nyomtatott évfolyamok digitalizálása és közzététele általában egy-egy kerek évfordulóhoz köthetően jelenik meg, mint a *Statisztikai Szemle* kilencvenedik évfolyama vagy az *Économie et Statistique* elődfolyóiratának számító *La Conjoncture* első megjelenésének hetvenedik évfordulója.

A KSH gondozásában megjelenő gazdag folyóirat-portfólió nem jellemző a külföldi statisztikai hivatalokra. Ugyanakkor más országokban gyakrabban fordul elő, hogy a hivatalos statisztikai szolgálat szervezetei mellett egyesületek, tudományos társaságok kiadványai gazdagítják a statisztikai szakirodalmat, és így egymás mellett és egymást kiegészítve jelennek meg a statisztikai hivatal és az attól független statisztikai szervezetek folyóiratai.



A folyóiratok periodicitása meglehetősen nagy változatosságot mutat, de a lehető leggyorsabb esetben jelennek meg havonta. A *Statisztikai Szemle* mellett kizárólag a *Statistische Nachrichten* és a *Wiadomości Statystyczne* jelentkezik minden hónapban. A KSH folyóiratai között is kizárólag a *Statisztikai Szemle* rendelkezik havi periodicitással, a *Területi Statisztika* és a *Demográfia* kéthavonta, illetve negyedévente jelenik meg. Általánosságban a kéthónapos mellett inkább a negyedéves megjelenés jellemző a statisztikai hivatalok folyóirataira. A nemzeti nyelven megjelenő periodikák szinte mindegyike megjelentet angol nyelvű különszámot is éves vagy kétéves gyakorisággal.

A hivatalos statisztika folyóiratai általában nem rendelkeznek impakt faktorról – és ez alól nem jelentenek kivételt az angol nyelvű lapok sem. Az impakt faktor értékeit nyilvántartó és szolgáltató Clarivate Analytics adatbázisai szerint (Journal Citation Report, Web of Science) 2017-ben összesen 123 statisztikai folyóirat rendelkezett impakt faktorról. Ezek a kiadványok kivétel nélkül tudományos társaságok és profi szakkiadók termékei. Mindössze két folyóirat rendelkezik impakt faktorról a hivatalos szervezetek lapjai közül. Ezeket azonban nagy, nemzetközi kiadók jelentetik meg a hivatallal együttműködve: a kanadai statisztikai hivatal *Survey Methodology* és a svéd hivatal *Journal of Official Statistics* folyóirata. Ez a jelenség is rávilágít arra, hogy a hivatalos statisztika szervezetei kiadványaikat nem kizárólag a tudományos közösség számára jelentetik meg – legalábbis nem szorúlnak rá, hogy részt vegyenek olyan versenyhelyzetekben, amelyeket a kiadói folyóiratokra kényszerítenek a tudománymetriai mutatószámok.

Összegzés

A felmérés adatai alapján megállapítható, hogy a hivatalos statisztika intézményei nagyrészt elkötelezettek a nyílt hozzáférés irányába. Ugyanakkor érdekes, hogy a statisztikai évkönyvek és a szakfolyóiratok más mintázatot mutatnak a nyílt hozzáférés szempontjából: találni olyan intézményt, ahol az évkönyv szabadon elérhető, a szakfolyóirat azonban nem.

A vizsgálat alapján érdemes volna a hivatalos intézményrendszerben jobban tudatosítani a nyílt hozzáférés kérdését, hogy az intézmények egységes, a széles nyilvánosság és a szakmai közösségek számára is egyértelmű irányvonalak mentén képviseljék a nyílt hozzáférés gyakorlatát – és természetesen csatlakozzanak a mozgalomhoz azok az intézmények is, amelyek egyelőre csak fizetős módon biztosítják egyes online kiadványok hozzáférését.

Irodalomjegyzék

1. Cannon, San, Marc Rodriguez. [„Survey of On-line Data Dissemination Practices for Government and International Statistics.”](#) *The Statistics Newsletter* 49. sz. (2010): 5–8.
2. Lencsés Ákos. „A Magyar statisztikai évkönyv szerkezetének változásai 1872 és 2016 között.” *Statisztikai Szemle* 95, 11–12. sz. (2017): 1144–1158. DOI: [10.20311/stat2017.11-12.hu1144](https://doi.org/10.20311/stat2017.11-12.hu1144)

3. Pfeiffer, Annette „Das Image des Statistischen Bundesamtes.” *Wirtschaft und Statistik* 12. sz. (2001): 981–987.
4. Rózsa Dávid, Kalmár Csilla, Lencsés Ákos. „A statisztika és a digitális átállás.” *Tudományos és Műszaki Tájékoztatás* 60, 2. sz. (2013): 47–65.

Entitások és kapcsolódások a könyvtári rendszerekben és azokon túl

Lendvay Miklós

Országos Széchényi Könyvtár, Információtechnológiai Igazgatóság

lendvay.miklos@oszk.hu

ORCID: 0000-0001-8065-5282

The limitations of time and physical space required in the workflows of cooperation and provided services seems to vanish as a result of the data and information revolution brought about by IT technology. The entity-based approach in the library sector means that we split all major components of the interlinkages into small units: the data elements become interlinked atomic descriptions; the units of connection may be institutions or individual scientists, the flexibility is limitless; the software components are small modules, microservices. The National Library Platform will be a distributed cooperation tool. This open source, cloud-based platform will go live in 2020, and it will be available for any type of library. Currently we have completed three modules of the platform: the National Namespace, a flexible cataloging tool, and an integrated text search – all these software-components break down the data into entities.

Keywords: entity, new generation library platform, namespace

Az információs technológia és a kiszolgáló IT megoldások fejlődésével a világról szóló legszélesebb körű, legkülönbözőbb tematikájú és legváltozatosabb metaadat-struktúrájú információk összekötése vált lehetővé felhőalapú megoldásokkal, szemantikus hálóban. A világháló képes áttörni a témakörök, az intézmények és a szakirányok, a nyelvek és a kódrendszerek korlátait. A hagyományos szerkesztésű, ellenőrzött és engedélyezett publikációval ellentétben a világháló egyáltalán nem a szakirányi meghatározottságból, hanem az információ sokkal kisebb egységeiből építkezik. Nem a nagy ívet bontja le apróbb egységekre, hanem az apró mozaikokból készíti összetett építményeit.

Az utóbbi évtizedekben hagyományos könyvtári világot alapvetően átalakító információs technológiák jöttek létre. Az egyre fejlődő, mesterséges intelligenciával bővülő új technológia lehetővé teszi, hogy a könyvtári állományok rendszerezése során egy magasabb szintű hierarchia jelenjen meg a gyűjtemény elemeinek leírásában. A mű, annak változatai, példányai és az ezekhez kapcsolódó releváns feldolgozások és információk egyértelműen azonosítottan és értelemszerűen kapcsolhatók össze – létrehozva ezzel az értelemszerű, azaz szemantikus hálót a megbízható, ellenőrzött adatok tartományában is – kiküszöbölve a szabad világháló esetlegességének és ellenőrizetlenségének hátrányait.

A könyvtárak akkor képesek az informatika által adott lehetőséghez kapcsolódni, amikor ezen alapelvek alapján létrejön egy széleskörűen használt együttműködési megosztott platform és a könyvtárak képesekké válnak a leírások, a metaadatok struktúráját a fentiek szerint használni. A meglévő adatok konszolidálása, az új adatok

célzott és megosztott katalogizálása az együttműködés és munkamegosztás révén a célja a könyvtárak által létrehozandó Országos Könyvtári Platformnak. Ebben az írásban részeredményekről számolok be a Nemzeti Névtér, az integrált szövegkereső és a flexibilisen paraméterevezhető katalogizáló felületünk kapcsán. Ezek a megvalósult fejlesztések az entitások kapcsolódását részben automatikusan, részben könyvtárosi munkával lehetővé tevő modulok, amelyek messze túlmutatnak a könyvtári világ határain, és más intézmények és tudományágak bekapcsolódását is lehetővé teszik – integrálják más intézményeket és az intézményeken kívüli szereplőket egyaránt.

Az entitásokat sokan és sokféleképpen értelmezik, habitus, nézőpont és egyes nyelvek / népek gondolkodásmódja szerint. Ebben az írásban az entitást önálló, zárt egységként definiálok, amely képes önállóan működni, és az őt körülvevő entitásokkal összekapcsolódni. Mindehhez az összes kapcsolódás, esemény, az entitás teljes környezete hozzátartozik, és meghatározó abban a kérdésben, hogy mennyire van az entitás értelmezhető, értelmes, hasznos kapcsolatban az őt körülvevő világgal, a vele kapcsolódó többi egységgel, entitással. Az entitás fogalma egyszerre foglal magába olyan különböző tárgyakat, mint dolgok, tulajdonságok, relációk, viszonyok, kapcsolódások és események. Az entitás ennél lecsupaszítottabbban, szűkebben és egyszerűbben nem értelmezhető, létét a definíciója értelmében maga a kommunikáció, a szimbiózis, az együttműködés adja.

A könyvtári rendszereket szemlélve entitásnak nevezhetjük a könyvtári publikációs egységeket, az információ tagolt elemeit, a könyvtári folyamatok résztvevőit, intézményeket, szereplőket, könyvtárosokat és olvasókat egyaránt. Önállóan, elszigetelten értelmét veszti mindegyik, nem képesek kiteljesedni. Ha a kapcsolódások kötöttek, meghatározottak, akkor távolabbi vagy újszerű elemek nehezen vagy egyáltalán nem képesek értelmes módon kapcsolódni velük, zárt és kommunikációra nem képes rendszerek alakulhatnak ki. Miképpen tágítható a kapcsolódások köre, miképpen tehető az individuális tartalom univerzális jelentőségűvé és széles körben használhatóvá és újrahasznosíthatóvá?

A hagyományos könyvtár entitásai elsősorban fizikai hordozóikon léteznek. Ezek sokféleképpen szemlélhetők, a fizikai határok megrajzolásával számos entitást tekinthetünk egységnek: lehet entitás egy gyűjtemény, hagyaték, de akár a könyv gerince, egy folyóiratban szereplő idézet is. A szellemi jelentés oldaláról nézve egy újabb absztrakciós szintre kell emelkednünk: a hordozójától független jelentés, gondolat, mű válik a szellemi szempontból megragadható, leírható entitássá.

A digitális feldolgozás és tárolás egy újabb dimenziót nyit az elérés szempontjából: ebben az esetben az ember érzékszerveivel a jeleket nem az adott fizikai hordozón, hanem egy elektronikusan kódolt és dekódolt jel megjelenítő felületen (monitoron, fejhallgatóban stb.) megjelenő lenyomatát érzékeli, és alakítja át jelentéssé – az ember szellemi lényként hozzáteszi a saját értelmezését, gazdagítja, torzítja, feldolgozza azt. A különbség a hordozó szempontjából elementáris: nem kell



fizikailag megközelítenünk a tárgyat, a sok ember általi elérhetőség kedvéért már nem kell sokszoroznunk az objektumot. Az elérés többszörözéséhez az elérési utak sokszorozása, a távoli elérés megvalósítása, az egy leképezett példány sok helyről való elérhetővé tétele a feladat. A hozzáférés forradalma a megközelítés gyökeres átalakulásából adódik. Akkor forradalmi lépés ez, mint annakidején a tömeges nyomtatás és sokszorozás kialakulása volt.

(Olyannyira más filozófia ez, hogy az internet szünet nélkül való elérhetősége már az adatcserét, és ezáltal a könyvtárban annyira megszokott adatcsere formátumokat, a géppel olvasható katalógizálási formátumban (MARC - Machine Readable Code), BIBFRAME-ben stb. is hosszabb távon feleslegessé teszi, hiszen ezek helyére a megosztás lép.)

Az egész könyvtári adathalmaz nemcsak a weben való megjelenésével, publikálásával lép a nyilvánosság elé. Maga a katalógus is a háló szerves része, webből van. Szükségszerűen egyre kisebbek a szervesen összekapcsolódó egységek – a fizikai hordozó szerepe háttérbe szorul, a digitális feldolgozás mindenből képes „egynemű” hordozót teremteni, a látványból képet, a térbeli élményből 3D-leképezést, a hangzathangfelvételt, az előadásból videofelvételt, képes az eleven tartalmakat leképezni, és a 0 és 1 kombinációiból álló kódsorok képesek ezt tárolni és szükség esetén újra jelként visszaadni. Ez a flexibilis transzformációs lehetőség és ezek átviteli sebessége felszámolja a tér és az idő korlátait, bármi bármikor elérhető, átalakítható, feldolgozható és újra visszaalakítható. Átlépjük azokat a fizikai korlátokat, amelyeket a szellemi lény, az ember a fejlődése során tapasztalatszerzésre használhatott, és belépünk egy virtuális valóságba, a „minden lehetséges” és a „minden súlytalan” birodalmába. A lélek gyakran lassabb annál, minthogy a gyors, és tértől és időtől független váltásokat képes legyen egészséges módon beutazni. Eljutottunk addig a határig, ahol a szellemi jelenlét és az anyagtalanság, szubsztanciamentes információ találkozik – és egyre jelentőségtejtésebbé válik a tény, hogy a szellemi lény tanúságtétele szükséges ahhoz, hogy a szellemi tartalmak valódi jelentéssel bírjanak. Úgy is feltehetnénk az ismeretelmélet kérdését: van-e zene abban a teremben, amelyben felhangzik egy oratórium, de amelyben nincs ember, aki azt átélje? Létezik-e a könyvben leírt tartalom a papírlapon, vagy csak az olvasó szellemi újratemtése által teremődik újra a jelentés? De ebben az írásban ez túl messzire vezetne, így megmaradunk a szellemi tartalmak fizikai jeleinek (ha mégoly illékony elektronikus) területén.

A mai könyvtárosi kihívások egyik legjelentősebbje az adatszövegteljesítés, amivel minden tudományterületen szemben állunk. Ma már mindenki és minden formában publikálhat és publikál is, fotókat, szövegeket, gondolatokat; könnyű akár könyveket megjelentetni, de még könnyebb előadásokat rögzíteni képen és hangban, és elérhetővé tenni a világhálón. Korábban a művek megszületése előtt végig kellett gondolni az ívet, amelyet azok bejárni kívántak, elejét, közepét, végét, és utána volt érdemes papírra, kottára vetni – ezt követelte a hordozó, a papírra vetett írás. Az informatika ebben is kitágítja a lehetőségeket, és ezzel részben be is szűkíti az emberi koncentráció jelentőségét: ma a szövegszerkesztővel el lehet

kezdeni egy fogalmazványt gondolat- és mondatfoslányokkal, majd azokat átalakítva, kibővítve, átformálva, összekötve, átrendezve egységes írást lehet alkotni. Teljes kísérletező festői életműveket lehet képszerkesztő programmal percek alatt szimulálni. A korábbi minőségellenőrzést végző, adott esetben inspiráló szellemi műhelyek, szerkesztőségek munkáját ma talán már fel is váltotta az internetes látogatók tetszésindexe, televíziós publikációk esetén a nézettség és eladhatóság a mérvadó. Az előállítási költségek minimálisra csökkentek, a technikai eszközök tömeges gyártása és széles körben való elérhetősége az árak radikális csökkenéséhez vezettek a nagy tömegek számára. Egy igényes drónos koncertfelvétel professzionális minőségű élő közvetítése könnyen megfizethetővé válik akár „amatőrök”-nek is. A könyvtárak számára ez azt jelenti, hogy a potenciális gyűjtőkör szinte végtelenné tágu, a szelekciót és minőségellenőrzést az esetek nagy százalékában nem végzi senki sem intézményesítve, sem civil módon szerveződve.

Más megközelítésben: áradnak ránk kontrollálatlan, strukturálatlan adatok. Az adatok jelentős része gépek által generált, sok kiadványt robotok állítanak össze. A mesterséges intelligencia képes az adatáradatot struktúrába szervezni, kapcsolódásokat felderíteni – és ebben a képességében emberi segítséggel tanítható. A WIKI birodalom (WikiPedia, WikiData, WikiMedia, WikiBase) eszközt ad azon emberek kezébe, akik szeretnék a rendelkezésükre álló tudást, információkat másokkal megosztani, közösségi szerkesztéssel és minőségellenőrzéssel. A mai információs robbanásban egy hatalmas strukturált web-enciklopédia és média-tár alakult ki, amelyet sokan, sok forrásból, sok nyelven építenek, egy hatalmas, erre alkalmas szoftver platformon. Lehetséges-e erre úgy tekintenünk, mint a világkönyvtár egy jelentős, talán legnagyobb méretű, struktúrájában finoman tagolt kötetére?

A könyvtárak intézményi működési módon, nagy és képzett személyezettel, állami finanszírozással járulnak hozzá, hogy a fizikailag és digitálisan publikált könyvek, folyóiratok, fotók, plakátok, webes kulturális tartalmak stb. az olvasók számára elérhetővé váljanak. A könyvtári egységek leírása lineáris, követi a jól definiált katalóguscédulák logikáját. Zárt egység, jól definiált kapcsolódási pontokkal – de rugalmatlan az újra, ahhoz, hogy zártságából kiemelhető legyen, keresési adatbázisokat kell fölébe építeni.

Két alapvetően eltérő irányból tárhatjuk fel az értelmes kapcsolódásokat az információöszönben fellelhető entitások között. Egyrészt a meglévő, definiált leírásokból kialakíthatunk kapcsolódásokat, amelyek segítenek az egymáshoz kapcsolódó tartalmakat csoportosítani, összekötni. A művelet során a gép az ember általi leírásokból összefüggéseket generál. Másrészt kialakíthatunk egy struktúrát, amelyben az entitásokat a legautentikusabb forrásból vesszük, és a platform emberi résztvevői által hozzáépítjük az információt, gazdagítjuk az ismeretek tárházát. Eszközt adunk a képzett emberi intelligencia számára, hogy strukturált módon gazdagítsa a meglévő halmazt.



Nincs akkora személyzete a legnagyobb könyvtárnak se, hogy képes legyen megbirkózni a rázúduló feldolgozandó információhalommal. Nincs meg a szerteágazó szaktudása a könyvtárosoknak, hogy minden területen igazán érdemi módon, kellő mélységben elvégezzék a szakmai feltárást. Ezért szükséges a szereplők körét kontrollált és szervezett módon kibővíteni: bevonni a kiadókat, a szerzőket, a szakterületek tudósait stb., hogy az általuk legjobban ismert területeken maguk legyenek képesek az adatokat feldolgozni, a munkatársaikkal megosztani, a széles közönség számára publikálni. Új generációs feldolgozó eszközre és szerkesztőségi munkafolyamatokra van szükség a könyvtári munka során.

A könyvtárak által feldolgozott publikációk összefüggéseinek feltárását a szellemi egységgel, a művel célszerű kezdeni. Csokorba kell fognunk mindazokat a publikációkat (fordítás, illusztrált kiadás, operaváltozat, megfilmesítés stb.), amely teljesen azonos az eredetivel, illetve, ha nem is azonos, de még mindig ugyanaz a mű. Felismerhető kapcsolatot kell teremtenünk azokkal a kiadványokkal, amelyek a műhöz egyértelműen kapcsolódnak, de már egy új művet jelentenek: a műről szóló értekezés, ismertető, paródia, recenzió, kritika, kommentár stb. A könyvtári világban az entitásalapú megközelítés megvalósítója a BIBFRAME adatcsereformátum, amelynek leglényegesebb eleme, hogy a művet tekinti a feldolgozás központi egységének. Az entitásalapú katalogizálási szabvány, az RDA (Resource Description and Access) utat jelent ahhoz, hogy a katalogizálás során ne betűsorokat, hanem entitasokat jelző egyértelmű azonosítókat adjunk meg az egységek azonosításánál. Az entítások összekötésének alapvető adatmodellje a bibliográfiai tételek funkcionális követelményeire vonatkozó FRBR - Functional Requirements for Bibliographic Records (amely nemcsak a könyvtári, de a múzeumi és levéltári világban is kialakította a szakterületnek megfelelő adatmodelljét).

A könyvtári platformot megvalósító szoftver esetében is az entitás-alapú megközelítést célszerű követnünk, mert ezáltal képes a teljes rendszer kielégíteni a legindividuaisabb követelményeket úgy, hogy egy univerzális képességű, felhőben üzemeltetett platformon a munkafolyamatok a legmesszebbmenőkig testreszabottan legyenek kialakíthatók. Ebben az entitás a legatomibb egységre szedett funkció, a microservice, tulajdonképpen egy kicsi zárt entitás, amely képes sokféleképpen és sok más entitással kapcsolódni. A fejlesztés során olyan alapelvek figyelembevételével dolgoznak az informatikusok, amelyek alapján az egységek saját magukon belül teljesen szabadok (programozási nyelv, adatmodell, felépítés tekintetében), és a külső kapcsolódások a többi microservice-el szabadon alakíthatók.

Az eddigiekben felvázolt entítások, az adat-entítások, a résztvevő-entítások (könyvtárak, munkatársak stb.) valamint a szoftver-entítások sokszínű kapcsolódása az alapja annak a nyílt forráskódú könyvtári platform fejlesztésnek, amelyet sok ország sokféle könyvtáros közössége együttes erővel tervez és specifikál, és sok informatikus fejlesztő valósít meg. A FOLIO mozaikszó jelentése: a könyvtárak jövője nyitott – the Future of Libraries is Open. Az Országos Széchényi Könyvtár által 2017-2019-ben lebonyolított nyílt nemzetközi tárgyalásos tenderen több induló közül

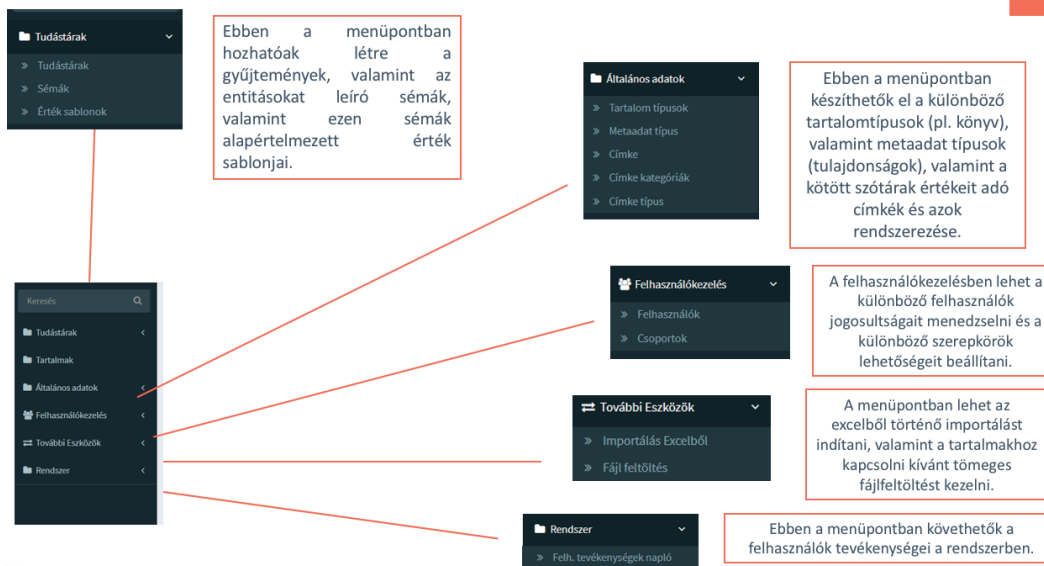
ez a megoldás nyert, és 2019 áprilisában a nyertes pályázó aláírta a Kormányzati Informatikai Fejlesztési Ügynökséggel (KIFÜ) azt a szerződést, amely alapján 2020 év végéig a KIFÜ felhő alapú, georedundáns infrastruktúráján a nyertes szállító által az Országos Könyvtári Platform installálásra kerül. Ez a platform a tender követelményekben meghatározott funkcionalitásokat valósítja meg. 2020 végén az OSZK és a Könyvtártudományi Szakkönyvtár összes meglévő szolgáltatását (az országos szolgáltatásokat, közös katalógus, könyvtárközi kölcsönzés stb. is beleértve) ebben fogja megvalósítani, 2021-től pedig nyitva áll a platform, hogy bármely magyar könyvtár beköltözzön, akár teljes funkcionalitását megvalósítva, akár valamely szolgáltatását illetően. Ahhoz, hogy ne csak a nemzeti könyvtár funkcióit, hanem bármilyen egyetemi, akadémiai, egyházi, közkönyvtár stb. funkcióit is maradéktalanul megvalósíthassa a platform, már a tenderkiírás elkészítésébe hat nagy, különböző funkciójú magyar könyvtár szakemberei kapcsolódtak be – és ezek és más könyvtárak is részt vesznek a követelmények pontosításában, és a rendszer kidolgozásában, végső soron megvalósításában, tesztelésében.

Az elmúlt két évben három modul készült el és került bevezetésre, mintegy a fent vázolt entitásalapú platform előkészítéseként.

A **katalogizáló modul**ban a metaadatok elkészítésének alapja a nagyon szabadon definiálható struktúra, a séma. Meghatározzuk a leírni kívánt tartalomtípust, és ehhez tulajdonságokat rendelünk, értéksablonokkal. Elkészítjük a kötött szótárak értékeit adó címkéket és azok rendszerezését. A leírás és a leírt digitális objektum egységet képez.

Ameddig a saját rendszerünkben dolgozunk, teljesen szabadok vagyunk a tulajdonságok meghatározásában, azok hierarchiájának kialakításában. Ha a külvilághoz szeretnénk kapcsolódni, akkor nemcsak a szoftverünk kell csatlakozóképes legyen, hanem az adataink is: azaz a külvilág számára értelmezhető (ideális esetben standard) formátumba kell őket hoznunk, de már a tulajdonságok által felvehető értékek meghatározásakor érdemes olyan értékkészletet választani, amely a külvilág számára is közvetlenül értelmezhető, de legalábbis külső azonosítóknak megfeleltethető.

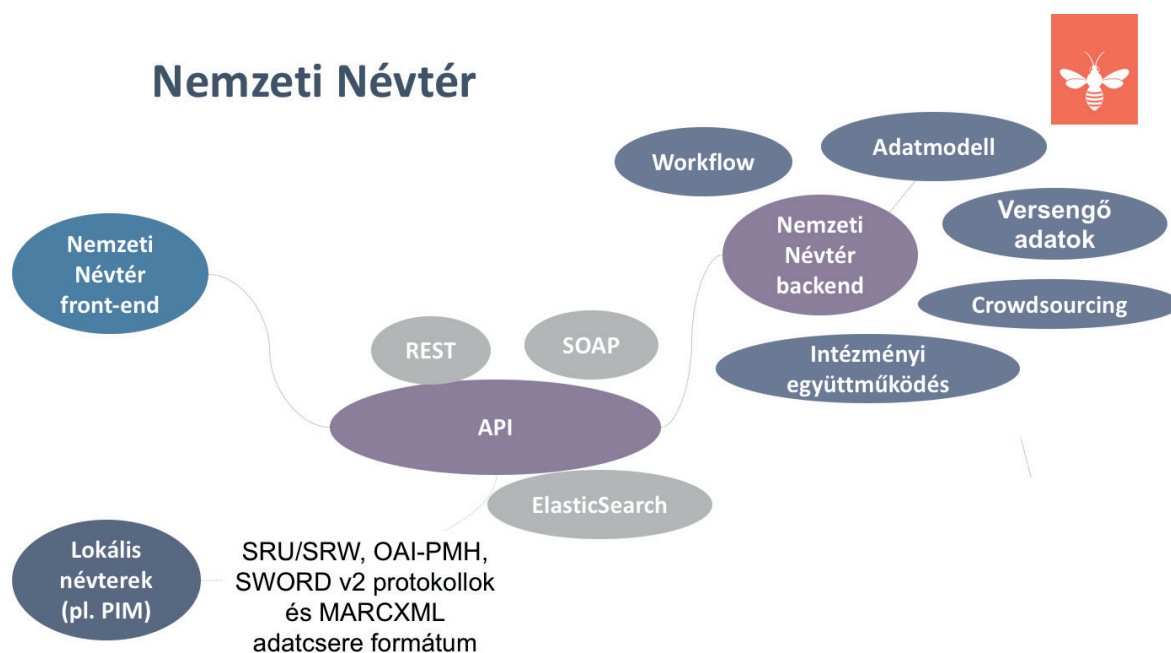
A katalogizáló funkciók áttekintése



14 | www.folio.org

folio+

Emiatt elengedhetetlen jelentőségű a **Nemzeti Névtér** létrehozása, amely célszerűen túlmutat a könyvtári világ határain. A Névtér olyan konzorciumban működik, ahol az adatok gyarapításában más típusú kulturális intézmények és más nem kulturális területen tevékenykedő intézmények egyaránt részt vesznek. Tetszőleges típusú névterek hozhatóak létre a platformon, eddig a személy-, testületi- és geo-névtér valósult meg. Az entitások azonosításának alapja a nyelvi modell – az adott nyelvben szereplő összes jelentésnek saját egyedi azonosítója van. A szoftver megvalósítja egy modern platform összes lényeges jellemvonását: képes az adatok közötti hierarchikus viszonyokat paraméterezhető módon leírni, flexibilisen kialakítható munkafolyamatok definiálhatóak, létezik az adatjavítást és gazdagítást megvalósító szerkesztőségi együttműködés, nemcsak az intézményi együttműködés, de intézményen kívül tevékenykedő hozzáértők bevonása is lehetséges a flexibilis jogosultságkezelés alapján (citizen science, crowdsourcing), a rendszer képes a kitüntetett névalakok és a versengő adatok kezelésére, valamint képes lokális névterekkel összekapcsolódni, több módon: felterjesztés – szinkronizálás – származtatás.



Az adatoknak időbeli érvényessége van. A földrajzi névtérben a hely névhordozói adatlapján idővonalon jelöljük a különböző névalakok és típusok kronologikus változásait. Külön egységekben szerepelnek az adott földrajzi hely feletteseinek és tipizált részei. A földrajzi geometria, a terület formája is időpont-függően megjelenített a térképen. A személynévtér szintén megjeleníti a névalakok kronológiáját az idővonalon. Ha szerzőről van szó, egy kiegészítő dobozban jelennek meg az OSZK katalógusában szereplő művei.

A Névter, a maga azonosítóival és jól definiált, gazdag egységeivel az entitás-koncepció legtisztább megvalósulása.

Földrajzi névtér – Névhordozói adatlap

[illegible]

A legkülönbözőbb forrásból származó szövegek esetén van szükségünk arra, hogy értelemszerű összefüggéseket tárjunk fel a szöveg tartalmi, lényegi jelentéséből levezetve. Ezt szolgálja az **integrált szövegkereső modul**. Ez feldolgozza a digitalizált anyagok optikai szövegfelismeréssel (OCR – Optical Character Recognition) előkészített szövegeit, egy automatikus helyesírás ellenőrzés (HunSpell) és beépített algoritmusok segítségével javítva az értelmezhetőséget. Ezután következik a szótövezés, nyelvi feldolgozás, szövegbányászat, azaz a szöveg feldolgozása. Majd a keresetőség érdekében a Solr search-engine segítségével az adatok indexelése történik. Az entitások felismerése után egy NoSQL gráf adatbázis (Neo4J) tárolja az entitásokat és a közöttük levő kapcsolatokat. A szinonimák kezelésére, szemantikus összekapcsolására ontológia-és tezaurusz-menedzsmentet használunk. A felhasználó számára fazettás, kulcsszó alapú és teljes szövegben való keresés egyaránt elérhető ezáltal. A gép tanítható ember által, és ez fontos eleme a hatékonyabb feldolgozásnak.

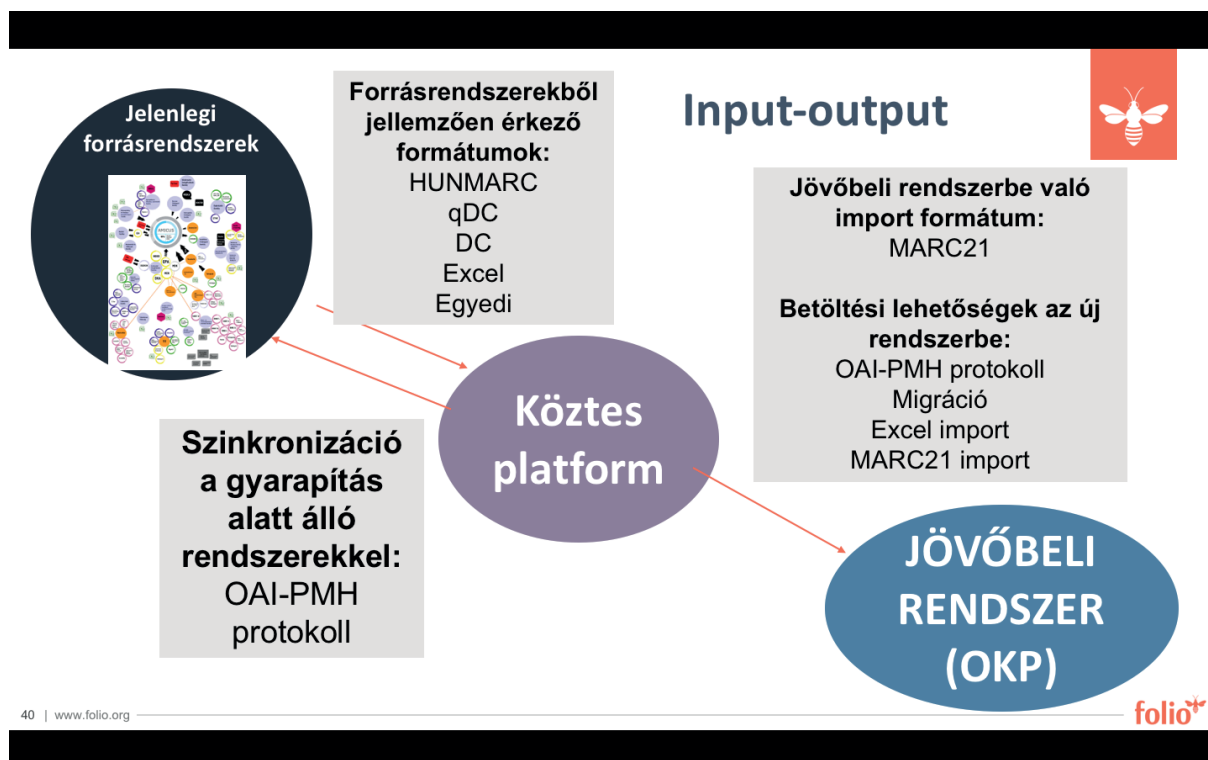
Integrált szövegkereső megjelenítés



A keresésre adott találati listában megjelenítésre kerülnek a fontosabb metaadatok, valamint a keresett kifejezés bővebb kontextusának megjelenése a szövegben. Minden találat esetében megvannak a lehetőségek az adott mű bővített megjelenítéséhez, betekintéséhez, full-text szöveges olvasásához.

Valamint a találati lista mellett jelenik meg a különböző szűrőket megvalósító sáv is.

Készülve az új megvalósuló Országos Könyvtári Platformra, szükségszerű a meglévő adatok minél alaposabb konszolidálása. Ez a munka másfél éve folyik, melynek során egy köztes platformra való betöltés érdekében gyűjtjük az adatokat, szűrjük a duplumokat, és az adatokat a MARC21 adatcsereformátum aktuálisan érvényes szabványos formájára konvertáljuk, figyelembe véve az RDA katalogizálási szabvány újdonságait is.



Mindhárom fejlesztés központi eleme az adatok elemekre, entitásokra tagolása. Ez az újfajta megközelítés teszi lehetővé, hogy nem kell a munkát többször elvégezni, támaszkodni lehet a megbízható és ismert forrásból származó feldolgozott adatokra. A kapcsolódásokon keresztül olyan lehetőségek nyílnak meg a feldolgozók és a végfelhasználók (olvasók) előtt egyaránt, amely az adatok soha nem látott komplexitását képes feltárni, szemantikus, értelmes módon. A bevont adatgazdagítók köre jelentősen tágul, lehetővé téve a széleskörű és szakértő módon történő feldolgozást.

A célunk tehát egy olyan központi szolgáltatás, amely paraméterezhető módon kínál individuális megoldást minden bekapcsolódó résztvevőnek. Az adatfeldolgozás és a szoftverfejlesztés egyaránt nyitott, várjuk ennek a lelkesítően újszerű, új generációs platformnak a megalkotásába, fejlesztésébe a bekapcsolódni vágyókat! Biztosak vagyunk benne, hogy a nemzetközi együttműködésben egy hosszabb távon fenntartható, nyelveken és kultúrákon átívelő, együttműködő, megosztott eszközt leszünk képesek létrehozni, amely messze túlmutat a piacon jelenleg megtalálható monolitikus megoldásokon. Az apró entitások teszik lehetővé, hogy ez a komplexitásában soha még nem látott, összetett platform olyan módon fejlődhessen a jövőben, hogy egyes elemei úgy újulhassanak meg, hogy a teljes egész szervesen gyarapodva, folyamatosan működni legyen képes.

Büszkék vagyunk az eddig elért eredményeinkre, és örömmel tekintünk az előttünk álló kihívások elébe. Örülünk, amennyiben minél több könyvtár és fejlesztő bekapcsolódik ebbe a munkába!

Open Access pályázati rendszer technikai megvalósítása és a szerzők támogatása a Szegedi Tudományegyetemen

Muzs Krisztina

SZTE Klebelsberg Könyvtár
krisztina.muzs@ek.szte.hu

Molnár Tamás

SZTE Klebelsberg Könyvtár
tamas.molnar@ek.szte.hu

Hoczopán Szabolcs

SZTE Klebelsberg Könyvtár
szabolcs.hoczopan@ek.szte.hu

Abstract

A Szegedi Tudományegyetem Open Access stratégiájának meghatározó része a támogatás nyújtása a szerzők nyílt hozzáférésű publikációinak közzétételéhez, a megfelelő folyóirat kiválasztásához illetve a szerzői támogatás biztosítása. Egy Szegedi Tudományegyetem méretű intézményben a saját büdzséből fizetett, vagy konzorcionális forrásból fedezett publikációs költségek, illetve az eleve nyílt hozzáférésűnek szánt cikkeken felül, a rendelkezésünkre álló, egyre növekvő számú, opcionális (Read and Publish) Open Access közzételési lehetőség adminisztrációja komoly kihívás a munkafolyamatba bevont kollégák számára. Nemcsak a ténylegesen elfogadott közlemények igénylik kollégáink figyelmét, de a készülőkben és az utánkövetés alatt lévők is, szigorú összhangban az intézményi Open Access támogatási szabállyal.

A szerteágazó feladatok ellátásához szükségünk volt egy online adminisztrációs rendszerre, mely számos automatizmussal könnyíti meg kollégáink munkáját, és minden számunkra fontos funkciót egy felületen tesz elérhetővé a támogatási döntéshozatal előkészítésétől kezdve a pénzügyi adminisztráción keresztül a publikációk utánkövetéséig. A rendszer szükségszerűen összehangolva működik az egyes kiadók saját online Open Access adminisztrációs felületével, máskülönben nagyon hamar elvesznénk az adatok tengerében. Az adminisztrációs rendszer része a Szegedi Tudományegyetem publikálás támogató felületének, a Szerzői Eszköztárnak.

The main part of the University of Szeged Open Access strategy is the provision of support for the payment of APCs, the selection of the appropriate journal and the author support. In an institution like the University of Szeged, where publishing costs are paid from several budgets or covered by many consortium contracts, and in addition the Open Access born articles, the administration of the increasing number of optional Open Access possibilities is a serious challenge for the colleagues involved in the workflow. Of course not only the accepted publications requiring the attention of our colleagues, but also the under preparation articles and the follow-ups, in strict accordance with the institutional Open Access policy.

In order to perform diverse tasks, we needed an administrative system that facilitates the work of our colleagues with a number of automations and provide every important function for us, from the preparation of decision making through financial administration to the follow-up. The system will necessarily work with each publisher's own Open Access administration interface, otherwise we would be lost in the sea of data very soon. The administration system is part of the so-called Author's Toolbox, which provide more and more author support services for the University of Szeged.

Keywords: open access, Article Processing Charge, Read and Publish, institutional author support

1. A szerzők támogatása

Az SZTE Klebelsberg Könyvtár 2015-ben kezdte meg a Szerzői Eszköztár¹ projekt keretében az egyetem szerzőinek publikációs támogatását. A Szerzői Eszköztár valójában egy folyamatosan bővülő "szolgáltatás csokor" melyben az egyetem szerzőit kívánjuk segíteni a publikációs folyamat során felmerülő problémák megoldásában. Természetesen senki helyett nem írjuk meg a cikkét, de segítséget nyújtunk a folyóirat kiválasztásban, publikációs workshopok szervezésében, az előzetes lektorálásban, plágiumellenőrzésben, a publikációs költség fedezésében, a repozitálásban, az MTMT feltöltésben és a szerző jogi kérdésekben².

Az utóbbi időben a szerzőket támogató szolgáltatásaink mellett elindítottuk a szerkesztőségeket támogató szolgáltatásainkat, melynek keretében Open Journal System alapú folyóirat platformot, és technikai támogatást biztosítunk az egyetemi folyóiratok számára, hogy szabályos Open Access folyóiratként jelenhessenek meg. A platformon futó folyóiratok esetében a megszülető cikkeket digitális objektum azonosítóval, és ezzel plágiumvédelemmel látjuk el. A szerkesztőségek kérésére a folyóiratplatformunkat összekötöttük a CrossCheck plágiumellenőrző rendszerrel is, így akár minden a szerzők által beküldött cikk, automatikusan plágiumellenőrzésen is áteshet. Az esetlegesen felmerülő szerzői jogi ügyletek rendezésére az egyetem a Copyright Clearance Centerrel áll kapcsolatban, hogy az esetlegesen az egyetem szerzői által benyújtott cikkekben szereplő külső forrásból felhasznált ábrákra, szövegrészekre stb. gördülékenyen tudjunk újrafelhasználási engedélyeket szerezni³.

Magának a publikációs támogatási programnak a következő alapvető céljai vannak: segíteni a szerzőket a nyílt publikációk megjelentetésében, az ügyintézés gyorsítása, az SZTE láthatóságának növelése, a nemzetközi rangsorokban való

1 Keveházi Katalin. Nyílt hozzáférés a gyakorlatban. *Tudományos és Műszaki Tájékoztatás* 63. 5. sz. (2016.), 193-197.

2 Dr. Meskő Eszter. Kutatást támogató könyvtár, a tartalomszolgáltatástól a tartalom-gazdaságig. *Workshop 2016 konferencia, Tartalomszolgáltatások*
<https://kifu.videotarium.hu/hu/recordings/12958/kutatast-tamogato-konyvtar-a-tartalomszolgáltatastol-a-tartalom-gazdasagig>

3 <http://szerzoknek.ek.szte.hu>



előrelépés elősegítése, az MTMT feltöltöttségének javítása és általánosságban az Open Access ügyének támogatása.

Az utóbbi években azt tapasztaltuk, hogy az Open Access publikációk körüli adminisztráció legalább annyira munkaigényes, mint az előfizetések intézése. Az egyetlen szervként, ami intézményi szinten vállalja fel a nyílt hozzáférés adminisztrációját, a kiadók számára a Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára vált az elsődleges kapcsolattartóvá, így azok az esetek többségében nem csak azokkal a beküldött cikkekkel keresnek meg minket, melyek szigorúan véve a támogatási körünkbe tartoznak, hanem minden egyetemi publikációval kapcsolatban. Tulajdonképpen annak köszönhetően, hogy a tagsági szerződéseket könyvtárként kötöttük meg a kiadókkal, rálátást nyertünk majdnem minden, az egyetemen megszülető Open Access, vagy potenciálisan azzá alakítható publikációra.

2. Open Access támogatási feltételek

A fentiekből következően az elsődleges feladatunk, hogy még a kiadói rendszerben megfelelő irányba indítsuk el a cikkek ügyintézését: a központilag támogatott publikációkat a könyvtár kiadói adminisztrációs rendszerben kialakított munkaterülete felé, a többit pedig a szerzők felé irányítva. Ahhoz, hogy pontosan tudjuk, melyik feltöltéssel mi a teendők, szükségünk volt egy átfogó egyetemi Open Access adminisztrációs rendszerre, amely már előzetesen minden adatot tartalmaz a gyors döntéshozatalhoz.

Az egyetemi Open Access támogatást az intézmény olyan feltételekhez kötötte, melyek mind a szerzőknek, mind az egyetemnek hasznosak. A feltételek ugyan folyamatosan változnak, követve az új lehetőségeket, igényeket és kihívásokat, de alapvetően az alábbi pontokkal kezdtük meg a munkát: a folyóiratnak tiszta Open Accessnek és Q1-Q2 besorolásúnak kell lennie. A levelező szerzőnek SZTE-s affiliációjának kell lennie, a cikket SZTE-es email címmel kell feltölteni, az SZTE-s társszerzők arányának magasabbnak kell lennie 50%-nál, a szerzők 2014 óta megjelent publikációinak szerepelnie kell az MTMT-ben és a repozitóriumban. A szabályok egyértelmű célokat szolgálnak: a célfolyóiratok első két kvartilise elősegíti a szerzőknek az idézetek megfelelő számát, ami az egyetem érdeke is, az SZTE-s email cím egyértelműsíti az affiliációt. Az MTMT és a repozitórium minél alaposabb feltöltöttsége pedig a szerző és az intézmény közös érdeke.

3. A Read and Publish szerződések beillesztése

A feltételeket folyamatosan alakítják azok a Read and Publish⁴ publikációs szerződések, amelyeket az egyetem köt az EISZ konzorciumon keresztül. Ezen szerződések, összehasonlítva a korábbi, egyetemünk által kötött szerződésekkel a Gold Open Access kiadókkal, nagyon kedvezményes és nagyszámú publikációs lehetőséget biztosítanak, a tiszta Open Access-en felül a hibrid típusú folyóiratokban is⁵.

4 Az olvasáshoz és a nyílt publikáláshoz való jogot egyben kezelő szerződés a kiadókkal
Dér Ádám. Könyvtári és konzorciumi kezdeményezések az Open Access átállás elősegítése érdekében. *ELTE Egyetemi Könyvtár és Levéltár, Hagyományok és kihívások VI.*
<http://videotorium.hu/hu/recordings/26100>

Már a kezdetekkor alapvető döntést kellett hoznunk. A R&P szerződéseknek nem feltétlenül kellett volna betagozódniuk a már megkezdett OA programunkba, működhetek volna függetlenül, felhasználási feltételek nélkül. Azonban azt tapasztaltuk, hogy az OA adminisztrációs felületek (dashboard) kezelésére ugyanúgy könyvtári munkaerőt kell dedikálnunk, ráadásul rengeteg extra feladatot jelent a szerzők folyamatos tájékoztatása. Így az tűnt a logikus lépésnek, hogy a meglévő struktúrába illesztünk be minden elérhetővé vált új forrást.

2018-ban ezek a típusú szerződések még kuriózumnak számítottak, így a fenti feltételrendszerben egyfajta kivételként szerepeltettük őket, hiszen nem tiszta Open Access kiadókról beszélhettünk, és a rendelkezésre bocsájtott publikációs kereteket igyekezni kellett kitölteni. Időnként a folyóirat színvonalából is engedtünk, hogy ne menjen veszendőbe a rendelkezésre álló OA kvóta.

A 2019-es évre annyi Read and Publish típusú szerződést kötött a Szegedi Tudományegyetem az EISZ konzorciumon keresztül, hogy számszerűleg sokkal több Open Access publikálási lehetőséghez jutottunk hozzá, mint amennyit a hagyományos Gold OA vonalon fedezni tudtunk volna. Így az vált kihívássá, hogy a rendelkezésre álló kereteket hasznosan töltsük ki, valamint ne menjen veszendőbe egyetlen APC sem⁶. Az első nagyobb kiadói szerződés, mely azonnal átütő sikerrel indult az SZTE-n, a Springer Nature-rel kötött volt⁷.

Nyilvánvalóvá vált, hogy a korábbi gyakorlat, amikor kivételként, eseti elbírálással kezeltük a Read and Publish típusú publikálási igényeket, nem tartható tovább, hiszen az arányokat tekintve komoly eltolódás történt. Az eredetileg tiszta nyílt hozzáférésű folyóiratokra kidolgozott szabályokat nem akartuk megváltoztatni. Ezeket a roppant drága APC-eket⁸ a tiszta OA folyóiratokban továbbra is csak komoly előzetes szűrést követően, az egyetem érdekeit szem előtt tartva akartuk átvállalni a szerzőktől. A költségkimélőbb R&P kvótákat viszont maradéktalanul fel szeretnénk volna használni, minél több szerzőt bevonva ezzel a programba. Emiatt megfogalmaztunk egy második szabályrendszert, ahol engedményeket tettünk az előzőekhez képest. A folyóiratok besorolását tekintve a harmadik kvartilisbe tartozó címeket is elfogadtuk, az SZTE-s szerzők arányát pedig mindössze 30%-ban határoztuk meg. A választóvonal a két szabályrendszer között a megkötött R&P szerződések mentén húzódik, szerzőinket gyakran tájékoztatni kell, hiszen könnyedén elveszhetnek az információk tengerében.

Az új "APC állományok" használatba vételekor nagyon hamar kiderült, hogy ezúttal az Open Access szerzői igényeket nem elég passzívan várni. Minden tájékoztatósi erőfeszítésünk ellenére sok szerző nem értesült a lehetőségekről, vagy nem tudja hogyan kell élni velük. Sok esetben, egy cikk elfogadásáról szóló kiadói értesítést

6 <http://eisz.mtak.hu/index.php/hu/open-access.html>

7 <http://eisz.mtak.hu/index.php/hu/open-access/286-open-access-megallapodas-a-springer-nature-kiadoval.html>

8 Article Processing Charge. A nyílt hozzáférésű folyóiratok esetében a cikk megjelentetéséért felszámított eljárási díj.




kézhez véve azt tapasztaljuk – a később bemutatandó Open Access pályázati rendszerünkben –, hogy a szerző nem pályázott előzetesen. Feltehetőleg nem értesült az OA lehetőségről, vagy félreértette a kiadói tájékoztatást. Ilyenkor egyenként megkeressük őket, ismertetve a publikálási lehetőségeket és azok feltételeit.

4. Open Access adminisztrációs rendszer

A fentiekből kiolvasható, hogy az Open Access munkacsoportunknak nem csak a pályázatot igénylő, egyértelműen más forrást használó, hanem a feltehetőleg hozzánk tartozó igényeket is kezelnie kell. Egy ilyen teljes intézményt átfogó adminisztráció elképzelhetetlen lenne, ha nem állna rendelkezésünkre egy olyan saját fejlesztésű rendszer, amely összeköti az összes résztvevőt, lehetővé téve a pályázatok adminisztrálását, a megszületett cikkek utánpótlását, valamint segítve a pénzügyi és APC adminisztrációt, abból fakadóan, hogy minden adatot egy felületen tárol. A szerző ebben a rendszerben szerkeszti és nyújtja be pályázatát, illetve az esetleges hiánypótlást, majd ezt követően a könyvtáros kollégák elvégzik a döntéshozó ellenőrzést, hogy az egyetem felső vezetése meg tudja hozni a döntést. Így amikor egy kiadótól értesítés érkezik egy SZTE affiliációjú OA vagy potenciálisan Open Access cikk elfogadásáról, a munkacsoport tagjai már pontosan tudják, hogy adott cikket elutasítani, jóváhagyni kell, esetleg a szerzőt kell megkeresni tájékoztatás céljából. A pályázatokon minden olyan adatot bekérünk a szerzőktől, ami a pályázati feltételek ellenőrzését lehetővé teszi, úgymint a folyóirat adatai, az APC várható összege, a szerzők MTMT, Scopus azonosítója illetve ORCID-ja.

Szegedi Tudományegyetem

 Klebsberg Kuno Könyvtára

Tisztelt Teszt Elek

Beadott pályázatának adatai

Azonosító	4458
Kérelmező neve	Teszt Elek
Kérelmező e-mail címe	szabolcs.hoczopan@ek.szte.hu
Kar	BBMK - Bartók Béla Művészeti Kar
Tanszék	Rézfüvös Tanszék
OA folyóirat címe	Open Access Journal
ISSN	0001-5350
Kiadó neve	OA Publishing
Publikáció címe	Open Publikáció
APC összege (Publikációs költség, a kiadó által használt devizában.)	2048 EUR
APC befizetés (várható) határideje	2019 / 11 / 28
Leadás időpontja	2019 / 11 / 01
Megjegyzés a pályázat elbírálójának	

Szerzők

Név	Affiliáció	Rendelkezik MTMT szerzői azonosítóval?	MTMT szerzői azonosító	Scopus Author ID	ORCID	Szerző típusa
Próba Szilárd	Összehasonlító Jogi Intézet	Igen	22221111		https://orcid.org/0000-3333-2222-3333	Első
Társszerző Pál	Élelmiszer-mérnöki Intézet	Igen	33331111		https://orcid.org/0000-3333-2222-1111	Nincs megadva
Teszt Elek	Rézfüvös Tanszék	Igen	1112222		https://orcid.org/0000-1111-2222-3333	Levelező
Név	Affiliáció	Rendelkezik MTMT szerzői azonosítóval?	MTMT szerzői azonosító	Scopus Author ID	ORCID	Szerző típusa

1. ábra Szerzői adatlap

A beérkező adatok alapján az OA munkacsoport adminisztrátora elkezdi a döntéselőkészítési adatolást. Megjelöli, hogy adott cikk melyik szerződés keretében, melyik szabályzat alapján lesz nyílt hozzáférésű. Leellenőrzi a megadott azonosítók alapján, hogy a szerzők cikkei fel vannak-e töltve az MTMT-be és a repozitóriumba. A folyóirat esetében a Scopus és a Web of Science adatok alapján felviszik az adatlapra a kvartilis besorolást, ami a legfontosabb bírálati szempontunk.

SZTE Open Access

OA folyóirat címe
Open Access Journal

ISSN
0001-5350

Kiadó neve
OA Publishing

Membership
Igen

SJR
Q2

WOS
Q2

Publikáció címe
Open Publikáció

DOI

APC összege (Publikációs költség, a kiadó által használt devizában.)
2048

APC Devizanem
EUR

APC befizetés (várható) határideje
2019-11-28

Oldal tetejére

Megjegyzés a pályázat elbírálójának

2.ábra Folyóirat adatlap

Amennyiben valamely feltételnek a szerző nem felel meg, pályázatát a könyvtári adminisztrátor elutasíthatja, ha csak hiánypótlásra van szükség, akkor visszaküldheti a szerzőnek szerkesztésre, pontosan megjelölve, milyen adatokat kérünk még feltölteni.

Előfordulhatnak kivételes esetek, például egy tudományterületen kevés nyílt hozzáférésű folyóirat érhető el és azok kvartilis besorolása is az általunk húzott határon billeg. Ebben az esetben az egyetemi felsővezetés automatikusan elutasítaná a pályázatot, azonban az űrlapba be lett építve egy szabadszöveges mező, ahol az ilyen jellegű kivételes pályázatokat meg tudjuk indokolni, adatokkal alátámasztani.

5. Döntéstámogatás

Ha lefutott a pályázat és a cikk megjelent, az OA munkacsoport munkája nem ér véget: a cikkek utóéletét is folyamatosan figyeljük, idézettségi szempontból. Erre a célra a megjelent cikkek DOI-jait⁹ gyűjtjük az adatbázisban, azokból építünk egy egyedi

9 Digital Object Identifier - digitális objektum azonosító



entitást az Elsevier Scival tudományometriai rendszerében. Ennek köszönhetően folyamatos rálátásunk van az általunk finanszírozott cikkek idézettségére, valamint hatására.

Bár a módszer működik, a DOI-k kézi gyűjtögetésénél professzionálisabb megoldásra volt szükségünk. Az utánkövetés egyszerűsítésére 2019 júniusától regisztráltuk magunkat a CrossRef FundRef adatbázisában. Az egyetem szerzőit arra kérjük, hogy a FundRef azonosítónkat rögzítsék a cikk feltöltésekor. Az így rögzített azonosító alapján a cikkeink már bármely bibliográfiai adatbázisból lekérdezhetők és elemezhetők lesznek.

Az elemzés és statisztikakészítés másik vonulata a pénzügyi utánkövetés. Mivel folyamatosan el kell számolnunk az egyetem felé az elköltött APC összegével, ezért természetesen az adatbázisba felvett APC összegek alapján folyamatos kimutatásokat készítünk. Az APC-k folyamatos monitorozásának a könyvelési szükségszerűségeen túl, döntéselőkészítő jelentősége is van. Amennyiben egy adott kiadónál a számlák összege, gyakorisága kiemelkedő, lehetőség szerint Open Access tagsági (membership) szerződést kötünk velük¹⁰. Ezzel a lépéssel a későbbi kifizetésekből kedvezményeket tudunk elérni, illetve az APC kifizetés idejét jelentősen redukálni tudjuk, ami a publikációs versenyben nem utolsó szempont.

Külön vonalat képvisel a pénzügyi adminisztrációnkban, hogy azon APC árakat is gyűjtjük, melyek Read and Publish szerződés keretében születő cikkekhez tartoznak. Nyilvánvaló, hogy itt klasszikus értelemben nem tudunk elszámolni a publikációs költségekkel, de egy-egy szerződés későbbi megítélésének szempontjából ez ugyanúgy fontos adat lehet, mint például a "Read" szerződési rész letöltési adatainak gyűjtése és elemzése.

6. Összefoglalás

A nyílt hozzáférésű publikációkkal kapcsolatosan nyilvánvalóan elképzelhető lenne másfajta ügyintézési metódus is, saját megoldásunkban nem titkoltan a központosított dokumentum beszerzési rendszerünket vettük alapul, ahol a központosított munkafolyamatban az SZTE Klebelsberg Könyvtár hozzáadott értéke az adminisztráció gyorsításában és az egyetemi szinten is jelentős mértékű megtakarításokban mutatkozik meg. A központosításnak hála, hasonló eredményeket érhetünk el publikációs vonalon, bár a centralizáció 2019 tavaszán még nem teljesen fejeződött be. Az SZTE beszerzési szabályzatának módosítása 2019 nyarára azon kevés APC adminisztrálását is az Egyetemi Könyvtár hatáskörébe utalta, melyek eddig el tudták kerülni a figyelmünket. Így már valóban minden feltétel megteremtődött a hatékony és központosított intézményi Open Access politikához.

¹⁰ <http://szerzoknek.ek.szte.hu/tagsagi-szerzodesek/>

Haladó funkciók és innovatív fejlesztések az EPrints és Omeka szoftverek körében

Nagy Dóra
SZTE Klebelsberg Könyvtár
dora.nagy@ek.szte.hu

Nagy Gyula
SZTE Klebelsberg Könyvtár
gyula.nagy@ek.szte.hu
ORCID: 0000-0002-8391-2851

Advanced features and innovative developments of softwares including EPrints and Omeka

The University of Szeged Klebelsberg Library has been using EPrints for handling institutional repositories since 2010. As we have progressed together with the program, we have been continually trying to exploit most of its components. In the presentation, we focus on introducing features that may be less widely used within the framework of Hungarian practice. Furthermore, some of our own innovations will also be revealed, for example, the possibilities of setting-up subject headings to the EPrints, managing file formats beyond PDF, the experience of using the statistics module, and the flexibility of designing metadata input forms.

In addition to text documents, several image collections have been recently digitized in our library. Access to these images is not provided through the institutional repositories, we use another open source software called Omeka. This system is equipped with a wide range of customizations through various extensions. It is able to manage user and access types, perform geolocations and offer various crowdsourcing techniques (collaboration, commenting, and sharing). After a brief introduction to the Omeka, some practical solutions will be demonstrated: Which types of extensions are we using? What are our future plans concerning this service?

Keywords: repozitórium, képtárház, EPrints, Omeka, tárgyszórendszer, crowdsourcing

Bevezetés

A könyvtárak mindig is élen jártak a nyílt forráskódú szoftverek használatában, így nem meglepő, hogy az elmúlt évtizedben a magyar repozitóriumi hálózat is ilyen alapokon épült ki. Tanulmányunk első felében ennek a hálózatnak az egyik alapvető építőkövét jelentő szabad szoftver, az EPrints repozitóriumkezelő rendszer kevésbé ismert, haladó lehetőségeit-funkcióit szeretnénk bemutatni (1.1.-1.6. fejezet). A második részben pedig egy a hazai gyakorlatban még kevésbé elterjedt, elsősorban képek és más médiaállományok kezelésére kifejlesztett ingyenes, Omeka nevű szoftver néhány bővítményének bemutatása olvasható (2.1.-2.2. fejezet). A programok általános szintű ismertetésén túl elsősorban azokra a magasabb szintű műveleteket megvalósító kiterjesztett szolgáltatásokra koncentrálunk írásunkban, amelyek használata még kevésbé gyökeresedett meg a hazai gyakorlatban. Erre próbálunk utalni írásunk címében a "haladó funkciók" szófordulattal. Azokat a



plusz szolgáltatást nyújtó lehetőségeket nevezzük haladó funkcióknak, amelyek nem érhetőek el a szoftver alap telepítésekor, vagy esetleg valamilyen rejtett funkcióként jelennek csak meg. A bemutatott esetekben ezen funkciók teljes körű kihasználásához bővítmények telepítése, konfigurálása, esetleg önálló fejlesztések adaptálása is szükséges.

1.1. EPrints – Statisztika bővítmény

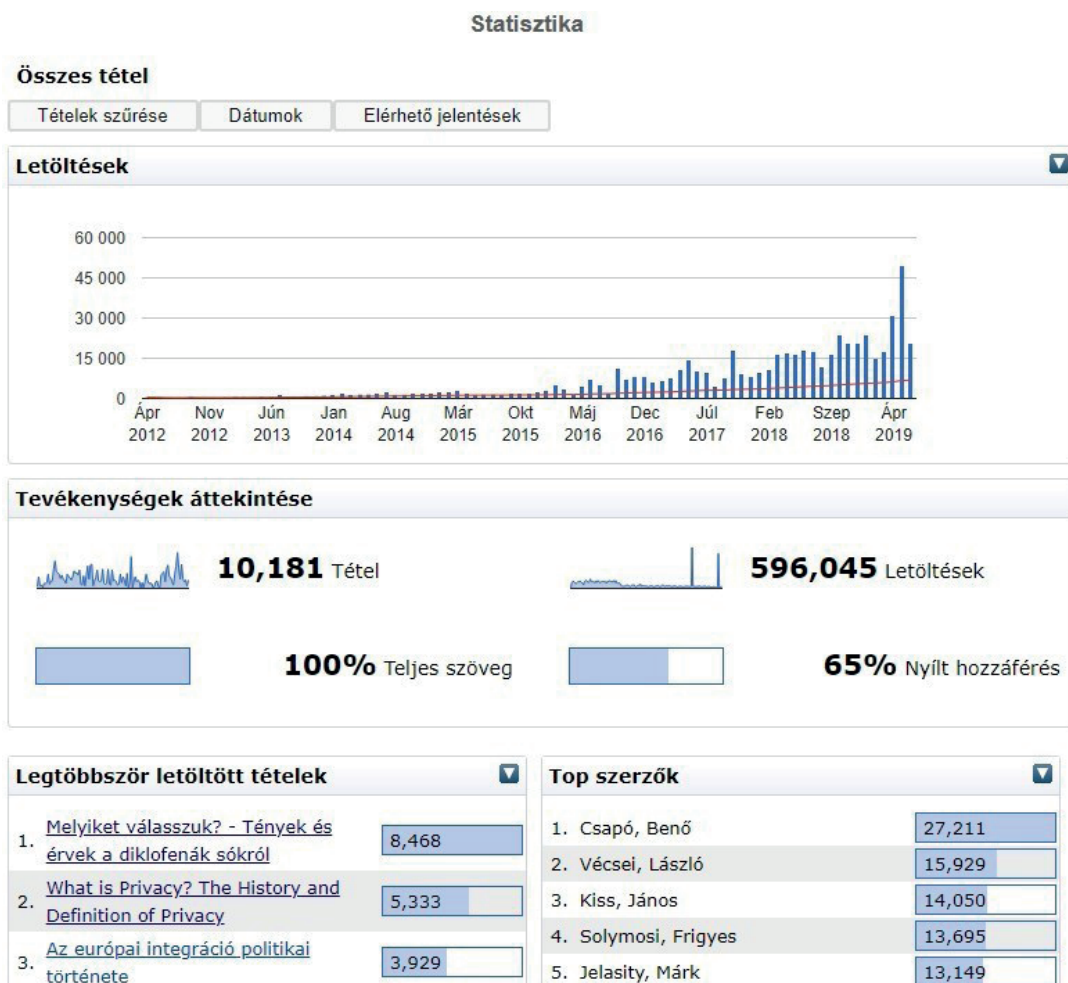
Ezek körébe tartozik az EPrints statisztika bővítménye, amely 35 hazai EPrints repozitóriumot tekintve csak 14 esetben van telepítve, melyből 12 SZTE megvalósítású. Saját EPrints repozitóriumainkon kívül a BGE Publikációtárnál és Dolgozattárnál, valamint a Corvinus Disszertációk és Kutatások repozitóriumoknál érhető el ez a modul.

Maga az *IRStats2* nevű statisztika bővítmény az EPrints Bazaar bővítménykezelő felületéről telepíthető. Rengeteg féle statisztikai adatot kinyerhetünk segítségével. Többek között kilistázza a legtöbbször letöltött szerzőket és műveket, a repozitóriumban lévő tételek számát, az összes letöltésszámot, illetve a bővítmény telepítésétől kezdve grafikonon is ábrázolja a letöltési adatokat, piros vonallal megjelenítve az összesített átlagot. A bővítmény segítségével különféle jelentések elkészítése, illetve a tételek többféle szempont szerinti szűrése is lehetséges. Kimutatható például a letöltések országonkénti forrása (térképen ábrázolva), évenkénti összehasonlítás végezhető a repozitórium gyarapodásáról, illetve a dokumentumok típus szerint is megjeleníthetők grafikonon. Az eredmények XML, JSON és CSV formátumban exportálhatók. Az SZTE Publicatio Repozitóriumnál az egyes tételek összefoglaló oldalán szintén megjelenik az adott rekordhoz tartozó letöltési statisztika.

Az összes SZTE-s repozitórium nyitóoldalán elérhető a statisztika főoldalára mutató link, amely szabadon böngészhető. Felhasználóink pozitív visszajelzései alapján helyes döntésnek bizonyult a statisztikai adatok nyilvánossá tétele, hiszen így pontos képet kaphatnak a repozitóriumban lévő anyagok használatáról, ami tovább növelheti a népszerűséget és az érdekeltséget.

1.2. EPrints – Kötegelt szerkesztési lehetőségek adminisztrátori felületen

Munkánk során rendszeresen használjuk az adminisztrátori felületen elérhető kötegelt szerkesztési lehetőséget. Több száz rekord betöltésekor a *Batch Editing Tool* segítségével történik az élő archivumba mozgatás, emellett hatékonyan használható különféle metaadatok tömeges javítására. Ilyenkor az adminisztrátori felületen a 'Tételek keresése' menüben lekeressük a módosítani kívánt rekordokat, majd a találati lista bal felső sarkában a 'Batch Edit' lehetőségre kattintva jutunk arra a felületre, ahol a legördülő menüből kiválasztható az az EPrints mező, amelyet korrigálni szeretnénk, majd a 'Hozzáadás' gombra kattintva írhatjuk be az új értéket. Öt féle művelet hajtható végre, ami a mező típusától függően eltérő. Lehet törölni a mező teljes tartalmát vagy a tartalom egy részét, hozzáírni új adatot a meglévők elé vagy mögé, illetve karaktersorozatot cserélni.



1. ábra – Az IRStats2 bővítmény nyilvános főoldala

Batch Editing Tool

Állapot valamelyike megtalálható "Felhasználói munkaterület" ÉS Cím "teszt". Az eredmény -Dátum, Szerző, Címszerint rendezve.
Applying batch alterations to 8 items (only first 8 shown):

- Teszt Elek [Teszt](#) (2018) [Jegyzet, tankönyv]
- Teszt Elek [Teszt](#) (2018) [Online oktatási csomag (e-learning lecke/téma)]
- Teszt Elek [Teszt](#) (2018) [Online oktatási csomag (e-learning lecke/téma)]
- Teszt Elek [teszt](#) (2018) [Jegyzet, tankönyv]
- Teszt Elek [teszt cím batch](#) [Online oktatási csomag (e-learning lecke/téma)]
- Teszt Elek [teszt](#) [Jegyzet, tankönyv]
- Teszt Elek [teszt cím](#) [Jegyzet, tankönyv]
- Teszt Elek [teszt 2](#) [Online oktatási csomag (e-learning lecke/téma)]

Modify Records **Remove all Records**

Kiegészítő névelem Családi név Keresztnév Email

Append new value Teszt Elek

Tanszék, intézet

Prepend new value Angol Tanszék

Dátum

Replace value with Év: 2019 Hónap: Április Nap: 25

Dátum

2. ábra – A Batch Editing Tool használat közben: új szerző felvétele, tanszék és dátum mező kitöltése



1.3. EPrints – Összetett űrlap több szereplő kiszolgálására

Az EPrints képes arra, hogy olyan beviteli űrlapot definiáljunk, amelynél az egyes mezőkhöz és állomásokhoz csak bizonyos jogosultsággal rendelkezők férhetnek hozzá. Az SZTE Publicatio Repozitóriumnál és az SZTE Elektronikus Tananyag Archívumnál beépítettük ezt a lehetőséget, ugyanis ezek esetében külső felhasználók is töltenek fel anyagokat: előfordulnak olyan adatmezők, melyek csak a szerkesztők számára relevánsak, ezekkel nem akartuk feleslegesen terhelni a feltöltőket, emiatt számukra egy egyszerűsített űrlap jelenik csak meg a legfontosabb mezőkkel.

Emellett, a munkafolyamat konfigurációja során lehetőség van különböző elágazások létrehozására is. Nem csak felhasználó, hanem EPrints típus szerint is változtatni lehet a különböző metaadatmezőből álló űrlapokat: például videók esetében releváns a 'Lejátszási idő', míg könyv típusnál az 'ISBN'. A fent említett két repozitórium esetében a jogi nyilatkozatot is hasonló elágazással oldottuk meg. Ez a legelső állomás, melyre a felhasználó érkezik, ha új rekordot szeretne létrehozni, és csak akkor tud továbblépni, ha elfogadta a nyilatkozatot.

1.4. EPrints – eduID

Az SZTE Open Online Oktatás (SZTE O3) stratégia és az EFOP-3.4.3-16-2016-00014 projekt megvalósításának részeként az SZTE Elektronikus Tananyag Archívum feladata az egyetemen keletkezett tananyagok hosszú távú megőrzésének biztosítása, szakszerű feldolgozása és szolgáltatása. A tananyagok megfelelő elérésének biztosítására szükségessé vált egy új, eddigi repozitóriumainknál még nem használt hozzáférési típus bevezetése. Az eduID használatára több hazai adatbázisnál láthatunk példát (Szaktárs, Typotex), amely megoldás kivitelezhető intézményi repozitóriumoknál is. Jelenleg csak saját IDP-hez (Identify Data Provider) van bekötve az autentikáció, így egyelőre csak SZTE-s eduID-vel működik a rendszer.

1.5. EPrints – PDF-en túli fájlformátumok

A repozitóriumokban általában PDF formátumban tároljuk, szolgáltatjuk a digitalizált vagy eleve elektronikusan keletkezett anyagokat, azonban az EPrints valójában sokféle fájlformátumot képes kezelni: MP3, ZIP, PPT stb. Ha az adott fájlformátumot a böngésző támogatja, akkor külső program igénybevétele nélkül, a böngészőben is megtekinthetők a tartalmak. Ilyen formátumok a PDF, a HTML, a különböző kép és videó fájlok (JPG, MP4 stb.).

Az SZTE Elektronikus Tananyag Archívumban elkezdtünk HTML oldalakat is archiválni, ugyanis rengeteg ilyen típusú tananyagot találtunk, melyeket nem lehetett PDF formátumba konvertálni az interaktív megoldásaik miatt.

HTML oldalak archiválása a következőként történik: a lementett oldalt egy tömörített mappában feltöltjük EPrints-be, majd rendszeren belül kicsomagoljuk 'Single' módban, ezután lehet kiválasztani a fő fájlt (ami általában index.html). Nyilvános felületen a rekordban csak a fő fájl jelenik meg, a többi háttérben marad. Ezzel a megoldással felhasználóbarát módon lehet böngészni a weboldalt EPrints-en belül, nem kell a csomagot saját gépre letölteni.

Tétel nézet: Fizikai optika

Ez a tétel URL címmel van az Repozitóriumban:
<http://eta.bibl.u-szeged.hu/1693/>

Tétel visszavonása Mozgatás vissza a Feladatok közé Tétel megsemmisítése

Részletek Akciók Történet Kiadások

jogi nyilatkozat Adatok szerkesztése

Jogi nyilatkozat: Elfogadom a feltételeket és kijelentem, hogy a feltöltött anyag a saját szerzői művem.

típus Adatok szerkesztése

Oktatási anyag típusa: Jegyzet, tankönyv

feltöltés Adatok szerkesztése

Típus: Weboldal. **Nyelv:** magyar. **Hozzáférés szintje:** Nyilvános. **Fő fájl:** fizika_optika/index.html.

- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_7.19.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_84.1.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_0.2.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_6.24.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_48.8.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_5.22.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_22.2.gif
- fizika_optika/titan.physx.u-szeged.hu/tamop411c/public_html/Fizikai_optika/eXe_LaTeX_math_36.6.gif

3. ábra – Egy HTML alapú tananyag tételnézetben

1.6. EPrints – Szakterületi besorolás

Mivel a tananyagok az egyetem minden karáról érkezhettek, a tartalmi feltárást elősegítendő bevezettünk egy szakterületi besorolást. Ehhez a H2020 projekthez készített hétszintes, angol és magyar¹ nyelven is rendelkezésre álló tárgyszórendszert választottuk, melyet a megújult MTMT2 is beépített. A taxonómia két ágra bomlik: „Tudomány” (N=~3400); „Gazdálkodás és innováció” (N=~300). Az EPrints illesztés során a témák mindegyike új ID-t kapott, illetve kivettük belőle a 'Gazdálkodás és innováció' ág kategóriáit. A táblázatot az elkészített XML séma alapján EPrints-be tölthető XML formátumra konvertáltuk az XMLBlueprint² program segítségével.

Az EPrints-en belül a 'subject' típusú mező tartalmazza a tárgyszavakat, melyek módosítására, kiegészítésére az adminisztrátori felületen van lehetőség. A rekord szerkesztési munkafolyamatában egy külön lapon kaptak helyet a szakterületek, melyek közül több is hozzárendelhető egy rekordhoz. A külső, nyilvános felületen a böngészés szakterület szerinti nézetben csak azok a kategóriák láthatóak, melyekhez van már rekord rendelve. Ezek tovább csoportosíthatók szerző és oktatási anyag típusa szerint, valamint minden besorolt rekord tételnézetében is megjelennek a tárgyszavak.

¹ A fordítás Tichy-Rács Ádám munkája.

² XML Editor – XMLBlueprint. <https://www.xmlblueprint.com>

4. ábra – Az átalakított táblázat, valamint a betöltésre kész XML részlete

5. ábra – Szakterületek az EPrints szerkesztői felületén

126

2.1. Omeka – Általános tulajdonságok

Nemrégiben új intézményi kép- és médiaarchívum³ kialakítása vált szükségessé a korábbi MARC alapú képi nyilvántartásaink kiváltására. Erre a célra az Omeka Classic szoftvert választottuk, mely egy könyvtárak, múzeumok, levéltárak számára létrehozott nyílt forráskódú program, amit 2007-től fejleszt a Roy Rosenzweig Center for History and New Media intézet. A projekt támogatói között szerepel többek között a Library of Congress is⁴

A működéshez szükséges környezet: Linux operációs rendszer; Apache HTTP szerver; MySQL adatbázis-kezelő; PHP szkriptnyelv és ImageMagick képszerkesztő.⁵ Kezelőfelülete hasonlít az elterjedt tartalomkezelő rendszerekhez, mint pl. a WordPress, így viszonylag könnyen megtanulható a kezelése.



6. ábra – Az Omeka Classic adminisztrátori felülete

Az Omekát részben a részletes dokumentáció miatt választottuk, nem csak felhasználói⁶, de fejlesztői⁷ szempontból is jól dokumentált, valamint aktív felhasználói közösség virágzik a hivatalos fórumon⁸. A közösségi segítséggel készült fordításokat a Transifex szolgáltatáson keresztül⁹ gyűjtik. Kollégáink segítségével

3 SZTE Képtár és Médiatéka. <https://mediateka.ek.szte.hu>

4 Az Omeka projektről. <https://omeka.org/about/project>

5 Omeka Classic rendszerkövetelmények. https://omeka.org/classic/docs/Installation/System_Requirements

6 Omeka Classic felhasználói dokumentáció. <https://omeka.org/classic/docs>

7 Omeka Classic fejlesztői útmutató. <https://omeka.readthedocs.io/en/latest>

8 Omeka Classic fórum. <https://forum.omeka.org/c/omeka-classic>

9 Omeka Classic fordítások. <https://www.transifex.com/omeka/omeka>



elkészítettük több bővítmény és az alaprendszer nagy részének magyar fordítását, amely elérhető a fent említett oldalon és bekerült az Omeka Classic legújabb 2.7-es verziójába.

A sokrétű tartalomszolgáltatási funkcióknak köszönhetően a szolgáltató intézmények igényeihez jól igazítható a rendszer. Használhatjuk a már elérhető témák egyikét¹⁰, de elkészíthetjük saját, intézményi arculathoz igazodó megjelenésünket is. A különféle bővítmények segítségével térképen ábrázolhatjuk a rekordjainkat és virtuális kiállításokat készíthetünk belőlük. Felhasználóinkat is bevonhatjuk a feldolgozás folyamatába, hozzászólások és egyéb közreműködések formájában. Mivel a rendszer képes többféle felhasználotípus kezelésére, alkalmas crowdsourcing projektek megvalósítására is.

A feltöltési folyamat többlépcsős, illetve egy elemhez több fájl is feltölthető. Első lépés a metaadatmezők kitöltése, majd az elem típusának meghatározása (kép, hanganyag, szöveg stb.). Ezután lehet feltölteni a fájlokat, címkéket hozzáadni és egyéb bővítmények nyújtotta lehetőségeket használni pl. térképen megjelölni a helyszínt.

Egy elem lehet nyilvános vagy privát. Utóbbi nem jelenti feltétlenül azt, hogy a felhasználotól teljesen elzárt lenne. Az Omeka Classicban alapból elérhető négy felhasználotípus egyike a 'Researcher' felhasználó, aki megtekintheti a nyilvános és zárt tartalmakat is, de semmilyen módon nem tudja azokat módosítani. Vagyis megengedjük a felhasználóinknak, hogy regisztráljanak az oldalunkra, ahol adminisztrátori jóváhagyás után a kutatók hozzáférést kaphatnak a teljes gyűjteményhez. A további felhasználotípusok a következők: 'Super' (legmagasabb szintű jogosultság), 'Admin' (új elem feltöltése és a már meglévő elemek szerkesztése, törlése), 'Contributor' (új elem feltöltése, ami csak ellenőrzés után nyilvános). A Guest User bővítmény segítségével megadható egy ötödik típus, az úgynevezett 'Guest' felhasználó, amely szükséges más bővítmények működéséhez (pl. Commenting).

Mivel az Omeka Dublin Core metaadat-struktúrát használ, a meglévő MARC rekordjainkat át kellett alakítani. Ehhez az ingyenes MarcEdit programot használtuk, amivel először egy CSV fájlba konvertáltuk az adatokat, majd a MARC mezőneveket helyettesítettük a megfelelő Dublin Core adatelemre és az így elkészült metaadatokat és fájl elérési utat is tartalmazó CSV fájlt betöltöttük Omekába a CSV Import+ bővítmény segítségével¹¹.

Bekerült az a jelentős képanyag is, melyek digitalizálása megtörtént az évek során, de még nem lettek egy rendszerben sem feldolgozva, így négy fő gyűjteményt alakítottunk ki: Egyetemi Arcképcsarnok; Egyetemi Képgyűjtemény; Shvoy Kálmán fotóalbumai; Szent-Györgyi Albert fotóalbumai (utóbbi három származik

¹⁰ Omeka Classic témák. <https://omeka.org/classic/themes>

¹¹ Részletek: Nagy Gyula, Nagy Dóra, Sándor Ákos: Tömeges adatkonverzió és rugalmas export-import lehetőségek az EPrints, OJS és Omeka szoftverek körében



7. ábra – A betöltött elemek az adminisztrátori felületen

MARC adatbázisból). Utóbb létrehoztunk két algyűjteményt is az Egyetemi Képgyűjteményen belül: Egyetemi könyvtári képek és Egyetemtörténeti fotóalbum. A videófelvételek az Egyetemi videógyűjteményben találhatóak meg. Jelenleg is több, mint 16000 nyilvános és 1500 feldolgozásra váró elemet tartalmaz az adatbázis. A további gyarapodása folyamatos, az újonnan bekerülő és a meglévő papíralapú képanyag digitalizálása révén.

2.2. Omeka – Bővítmények

A hivatalos oldalon több, mint 90 bővítmény¹² található, melyből mi 21-et használunk, ezek közül sorolunk fel néhányat:

- A *Dublin Core Extended* bővítmény hozzáadja az összes Dublin Core metaadat mezőt a már meglévő alap 15 adatelemhez (pl. létrehozás dátuma).
- Az OpenStreetMap alapú *Geolocation* bővítmény használható az egyes elemek térképen való ábrázolására.
- A *CSV Import+* bővítmény nemcsak metaadatokat és fájlokat, hanem egyéb adatelemeket pl. geokódokat is tud importálni. Többféle CSV struktúra áll rendelkezésre, melyek segítségével egy rekordhoz több fájl is betölthető egyszerre; utólagos nagy tömegű módosítás is végrehajtható, illetve az egész betöltés vissza is vonható.
- Az *Exhibit Builder* bővítménnyel lehet több aloldalt is tartalmazó virtuális kiállítást létrehozni, melynek akár a főoldaltól eltérő egyedi témát is készíthetünk. Az *Exhibit Image Annotation* ezen belül a képen szereplő dolgok megjelölésére (tagelésére) alkalmas.

¹² Omeka bővítmények. <https://omeka.org/classic/plugins>



- A *Commenting* bővítmény hozzászólási lehetőséget biztosít a felhasználóknak. A *Guest User* bővítménnyel együttműködve egy olyan felhasználói típust hoz létre, amely csak hozzászólás írására képes, bejelentkezés és adminisztrátori jóváhagyás után. A *Social Bookmarking* bővítmény minden elemhez megosztás gombokat társít (pl. Facebook, Twitter). Míg a *SimpleContactForm* regisztráció nélküli üzenetküldésre használható.

Ezek mellett több bővítmény tesztelése zajlik, ilyen például az *AvantRelationship*, mellyel kapcsolati háló hozható létre az egyes elemek között. A *Simple Vocab Plus* bővítménnyel pedig ajánló tárgyszólista készíthető. A tartalmi bővülést tekintve folyamatos a Szent-Györgyi Gyűjtemény feldolgozása, illetve további hagyatéki és egyetemi vonatkozású kép- és videóanyagok betöltése és feldolgozása.

Tömeges adatkonverzió és rugalmas export-import lehetőségek az EPrints, OJS és Omeka szoftverek körében

Nagy Gyula
SZTE Klebelsberg Könyvtár
gyula.nagy@ek.szte.hu
ORCID: 0000-0002-8391-2851

Nagy Dóra
SZTE Klebelsberg Könyvtár
dora.nagy@ek.szte.hu

Sándor Ákos
SZTE Informatikai és Szolgáltatási Igazgatóság
akos.sandor@ek.szte.hu

Possibilities of massive data conversion and flexible processes of export-import regarding EPrints, OJS and Omeka software

Since we have been using different types of systems to provide our digital contents, we had to migrate our data several times over the years. The lesson to be learned is ensuring interoperability and data exchange between these systems is a constant priority for libraries. Uploading several records at a time have become common practice in our library by now.

Each year since 2012 thousands of degree theses have been received from Modulo, which is an online platform assisting students and staff of the University. We have to convert the original XML file to a structure which is compatible with our EPrints based repository. The handling of SZTE Repository of Papers and Books, Miscellanea and UnivHistória repositories follows a similar procedure, but with different initial conditions. Last year we have started to work with Open Journal System, therefore establishing an efficient data conversion method between EPrints and OJS and vice versa was necessary.

One of our long-term plans was to have a search engine which can discover all of our repositories and we were able to achieve this with the help of the Vufind, which is an open-source search engine especially for libraries. A critical point of the project was to develop a data exchange format for MARC using OAI-PMH in EPrints.

Our previous Marc-based databases (Bodza) were exported to EPrints years ago, and with that experience we were able to start building our new Omeka-based photo archive. In this presentation we demonstrate the above mentioned processes through a few practical examples.

Keywords: repository, data conversion, data import and export, bulk data import, EPrints, Open Journal System, Omeka, VuFind, MARC, OAI-PMH



Bevezetés

Az SZTE Klebelsberg Könyvtár digitális tartalmainak szolgáltatása terén az évek során különféle szoftveres megoldásokat használtunk, így többször előfordult, hogy az adatok migrálására volt szükség. Fontos tanulság, hogy a különböző rendszerek közötti átjárhatóság és adatcsere biztosítása folyamatosan kiemelt feladatként van jelen a könyvtárak munkájában. Ezen migrálások mellett mára bevett gyakorlattá vált az új rekordok tömeges adatbetöltése repozitóriumainkba, amely munkamenetnek mindig az adott archívum sajátosságaihoz kell illeszkednie. Többek között 2012 óta ilyen módon zajlik az évente több ezer kurrens szakdolgozat átvétele a Szegedi Tudományegyetem tanulmányi rendszeréből. A hallgatók diplomamunkáikat a Modulo-ba töltik fel, amelyek a Neptunból származó, összefésült metaadatokkal együtt kerülnek exportálásra. Az így kapott XML fájlt EPrints XML formátumra (EP3 XML) szükséges konvertálni. Az adatkonverzió után történik a szakdolgozatok tömeges betöltése. Hasonló elvek mentén, de más kiinduló feltételekkel történik az SZTE Egyetemi Kiadványok, az SZTE Miscellanea, az SZTE UnivHistória és a Tiszatáj archívumának kurrens és retrospektív gyarapítása is.

A tavalyi évben egy EFOP 3.6.3 projektnek köszönhetően elindult az SZTE OJS folyóirat-platform. A Szegedi Tudományegyetemhez köthető folyóiratok archívumának gyors kiépítése és a hatékony munkavégzés megteremtése miatt szükség volt az EPrints-OJS, majd a repozitóriumok gördülékeny gyarapításának biztosítása miatt az OJS-EPrints automatizált konverziós irányok megteremtésére is. Az adatcsere egy másik megközelítését alkalmazva, repozitóriumaink speciális tagoltsága miatt régi tervünk volt ezek közös kereshetőségének biztosítása, amelyet egy EFOP 3.4.3. projekt keretében a VuFind rendszer segítségével valósítottunk meg. A projekt kritikus eleme volt az EPrints OAI-PMH kimenetén előállított MARC formátumok adatcsere lehetőségének kidolgozása.

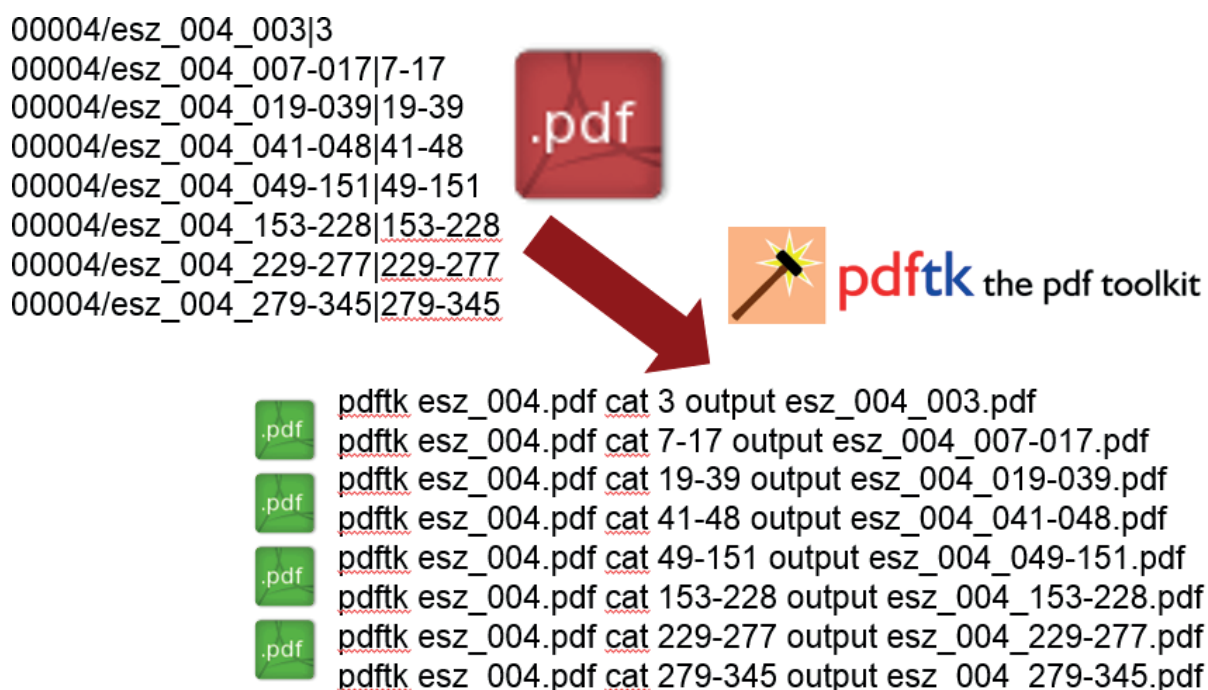
Az eddig említett szoftverek körében szerzett gyakorlat és korábbi MARC alapú adatbázisaink (pl. Bodza) tartalmának EPrints-be való átköltöztetése évekkel ezelőtt megtörtént, amely tapasztalatokat már fel tudtuk használni új, Omeka alapú képgyűjteményünk (SZTE Képtár és Médiatéka) létrehozásakor. Tanulmányunkban a fent említett projektek munkafolyamatai közül a legtanulságosabbakat kíséreljük meg bemutatni, elsősorban a tömeges adatkonverzió és a rugalmas export-import lehetőségek területéről.

1. Repozitóriumi metaadatbetöltési gyakorlat

Érdekes volna alaposabban körüljárni a hazai (és nemzetközi) repozitóriumok gondozói által használt tömeges adatbetöltési gyakorlatot, hogy mely intézményi repozitóriumoknál számít bevett rutinnak a tömeges export-import lehetőségek kihasználása, illetve a legtöbb szabványos repozitóriumrendszer esetében rendelkezésre álló OAI-PMH protokoll által biztosított potenciál kiaknázásának jó példái is tanulságosak lennének. Sajnos erre jelen tanulmány keretében nincs módunk, helyette elsősorban saját, ilyen irányú megoldásainkról tudunk csak beszámolni.

A metaadatok repozitóriumba való injektálásának egyik kézenfekvő módját jelentheti egy adott könyvtár OPAC-jából származó katalógusadatok repozitóriumi betöltése. Erre kiváló példaként szolgál az MTA Könyvtár és Információs Központban az Aleph és az EPrints rendszerek közötti adatcsere kapcsolat megteremtése vagy az Országos Széchényi Könyvtár adatbázis konszolidációs projektje kapcsán alkalmazott megoldások¹. Egy másik kiválóan működő példa lehet a metaadatok automatizált továbbítására az MTMT felőli, SWORD² protokollon át történő intézményi repozitóriumokba való adattovábbítás, ahol valójában nemcsak metaadatok utaznak a hálózaton keresztül, hanem a publikációk teljes szövegét tartalmazó fájlok is részei a továbbított csomagnak.

Saját tömeges adatbetöltési gyakorlatunk alapját a minél automatizáltabb megközelítés adja. Egyes repozitóriumok esetében már a kezdetektől célul tűztük ki a teljes analitikus feldolgozást, amihez az egyes folyóiratokat, egyetemi actákat és tanulmányköteteket magától értetődő módon, fájlszinten is részekre kellett bontani. Ehhez a PDFtk³ nevezetű parancssori eszközt használjuk. Ennek szintaktikája és működési módszere az 1. ábrán látható.



1. ábra – A PDFtk segédprogram alkalmazása: a cikkek logikai és fizikai oldalhatárai és a parancssorban futtatható, kötegelt fájl szintaxisa

- 1 Balázs László. Adatbázis konszolidáció az OSZK-ban. Networkshop 2017, Debrecen, 2016.03.29.-2016.04.01. Hozzáférés: 2019.06.17.
<https://kifu.videotorium.hu/hu/recordings/12965/adatbazis-konszolidacio-az-oszk-ban>
- 2 Allinson, Julie, Sebastien François, and Stuart Lewis. "SWORD: Simple Web-service offering repository deposit." (2008). Hozzáférés: 2019.06.17.
<http://scholarworks.csun.edu/handle/10211.3/118201>
- 3 PDFtk – The PDF Toolkit. <https://www.pdf4me.com/tools/pdfkit-the-pdf-toolkit>



Az ilyen módon előállított PDF fájlok alapját képezik a következő lépésnek, hiszen az akár több ezer soros fájllistából egy ugyanennyi soros XLS vagy CSV fájlt állítunk elő, ahol az egyes mezők fogják tartalmazni az egyes metaadat-elemeket. A módszer segítségével jó néhány adatelem tömegesen, illetve fél-automatikusan kitölthető (pl. számozási adatok, típusra vonatkozó adatok, azonosítók, stb.). Mivel az analitikus feldolgozás mellett fontosnak tartjuk a borítótól-borítóig terjedő teljes kötetek repozitóriumi megőrzését is, ezért ezek is bekerülnek egy, a cikkek metaadatait tartalmazó táblázathoz hasonló listába (a 2. ábrán piros karikával jelöltük a két megközelítés közötti különbségeket). A teljes lapszámokat és köteteket „full”, míg az egyes tanulmányokat, cikkeket „part” néven hivatkozunk. Ezek az elnevezések tükröződnek a táblázatok elnevezésében, illetve a későbbi munkafolyamatokban keletkező különböző kimenetekben is.

A következő lépésben az ilyen módon előállított, akár több tízezer soros táblázatokból elő kell állítanunk az EPrints által az automatikus adatbetöltések esetében preferált EP3 XML fájlokat. Ez egy repozitóriumonként meghatározott XML-skeleton alapján történik, a „full” és a „part” táblázatok eltérő adatelemeit természetesen ez az XML-skeleton is követi. Az elkészített XML-sémából az XMLBlueprint⁴, vagy újabban a saját fejlesztésű CSV2XML Python-alapú segédeszköz segítségével készülnek el az EP3 XML fájlok, amelyeket az EPrints importfelülete már fogadni tud. A betöltések során a teljes szövegű PDF-ek is automatikusan bekerülnek a rekordokba, melyet úgy oldunk meg, hogy a <documents> rész megfelelő <url> tag-jében egy általunk üzemeltetett Apache webszerveren elhelyezett, csak a betöltés idejéig élő URL címek találhatóak. Természetesen ehhez a megoldáshoz a repozitórium oldalán engedélyezni kell a web-import lehetőséget. Az itt röviden felvázolt módszert használjuk évek óta mind a kurrens, mind a retrospektív betöltések és időnként a migrálások esetében is, melynek segítségével immár több százezer rekordot tettünk közzé különböző archívumainkban.

1	2	3	4	5	6	7	8	9	10	11	12
1 - type	2 - title	3 - date	4 - publication	5 - volume	6 - number	7 - issn	8 - isbn	9 - format	10 - security	11 - filename	
book	Aetas - 33. évf. (2018) 1.sz.	2018	Aetas	33	1	0237-7934		full	public	aetas_2018_001.pdf	
book	Aetas - 33. évf. (2018) 2.sz.	2018	Aetas	33	2	0237-7934		full	public	aetas_2018_002.pdf	
book	Aetas - 33. évf. (2018) 3.sz.	2018	Aetas	33	3	0237-7934		full	public	aetas_2018_003.pdf	
book	Aetas - 33. évf. (2018) 4.sz.	2018	Aetas	33	4	0237-7934		full	public	aetas_2018_004.pdf	
book	A Szegedi Alföldkutató Bizottság	1930	A Szegedi Alföldkutató Bizottság	kö 1				full	public	alfoldkutato_001_001	
book	A Szegedi Alföldkutató Bizottság	1928	A Szegedi Alföldkutató Bizottság	kö 2				full	public	alfoldkutato_002_001	
book	A Szegedi Alföldkutató Bizottság	1928	A Szegedi Alföldkutató Bizottság	kö 2				full	public	alfoldkutato_002_002	
book	A Szegedi Alföldkutató Bizottság	1928	A Szegedi Alföldkutató Bizottság	kö 3				full	public	alfoldkutato_002_003	
book	A Szegedi Alföldkutató Bizottság	1928	A Szegedi Alföldkutató Bizottság	kö 4				full	public	alfoldkutato_002_004	
book	A Szegedi Alföldkutató Bizottság	1929	A Szegedi Alföldkutató Bizottság	kö 5				full	public	alfoldkutato_002_005	
book	A Szegedi Alföldkutató Bizottság	1930	A Szegedi Alföldkutató Bizottság	kö 7				full	public	alfoldkutato_002_007	
book	A Szegedi Alföldkutató Bizottság	1930	A Szegedi Alföldkutató Bizottság	kö 8				full	public	alfoldkutato_002_008	
book	A Szegedi Alföldkutató Bizottság	1931	A Szegedi Alföldkutató Bizottság	kö 9				full	public	alfoldkutato_002_009	
book	A Szegedi Alföldkutató Bizottság	1928	A Szegedi Alföldkutató Bizottság	kö 2				full	public	alfoldkutato_003_002	

1	2	3	4	5	6	7	8	9	10	11	12
1 - type	2 - title	3 - date	4 - publication	5 - volume	6 - number	7 - pagerange	8 - issn	9 - isbn	10 - format	11 - security	12 - filename
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	5-18	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	19-39	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	40-60	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	61-82	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	83-95	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	96-107	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	108-123	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	124-140	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	141-159	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	160-180	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	181-190	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	191-202	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	203-210	2064-5929		part	public	belvedere_2018_00
article	feldolgozásra vár	2018	Belvedere Meridionale	30	4	211-215	2064-5929		part	public	belvedere_2018_00

2. ábra – A teljes és rész PDF-ek metaadatait tartalmazó táblázatok

4 XML Editor – XMLBlueprint. <https://www.xmlblueprint.com>

full

```

1 <?xml version='1.0' encoding='utf-8'?>
2 <eprints xmlns='http://eprints.org/ep2/data/2.0'>
3 {repeat}
4 <eprint>
5   <eprint_status>archive</eprint_status>
6   <metadata_visibility>show</metadata_visibility>
7   <type>$(column_1)</type>
8   <full_text_status>public</full_text_status>
9   <title>$(column_2)</title>
10  <date>$(column_3)</date>
11  <publication_full>$(column_4)</publication_full>
12  <volume>$(column_5)</volume>
13  <number>$(column_6)</number>
14  <issn>$(column_7)</issn>
15  <isbn>$(column_8)</isbn>
16  <documents>
17    <document>
18      <format>$(column_9)</format>
19      <security>$(column_10)</security>
20      <files>
21        <file>
22          <filename>$(column_11)</filename>
23          <url>$(column_12)</url>
24        </file>
25      </files>
26    </document>
27  </documents>
28 </eprint>
29 {repeat}
30 </eprints>

```

part

```

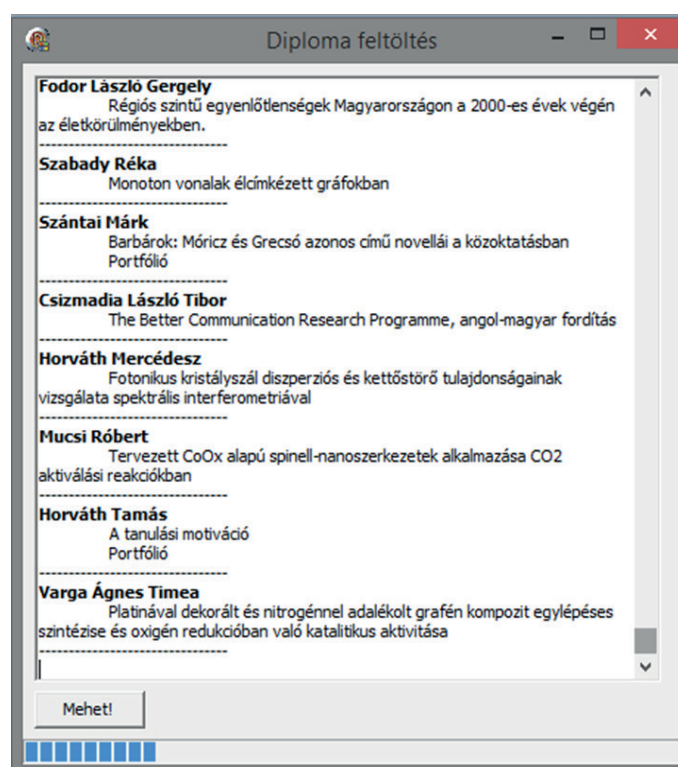
1 <?xml version='1.0' encoding='utf-8'?>
2 <eprints xmlns='http://eprints.org/ep2/data/2.0'>
3 {repeat}
4 <eprint>
5   <eprint_status>archive</eprint_status>
6   <metadata_visibility>show</metadata_visibility>
7   <type>$(column_1)</type>
8   <full_text_status>public</full_text_status>
9   <title>$(column_2)</title>
10  <date>$(column_3)</date>
11  <publication>$(column_4)</publication>
12  <volume>$(column_5)</volume>
13  <number>$(column_6)</number>
14  <pagerange>$(column_7)</pagerange>
15  <issn>$(column_8)</issn>
16  <isbn>$(column_9)</isbn>
17  <documents>
18    <document>
19      <format>$(column_10)</format>
20      <security>$(column_11)</security>
21      <files>
22        <file>
23          <filename>$(column_12)</filename>
24          <url>$(column_13)</url>
25        </file>
26      </files>
27    </document>
28  </documents>
29 </eprint>
30 {repeat}
31 </eprints>

```

3. ábra - A "full" és a "part" betöltéseknél alkalmazott XML-skeletonok, kiemelve az eltérő XML tag-eket

Az SZTE Diplomamunka repozitórium⁵ kurrens gyarapítására egy, a bemutatotthoz elviekben nagyon hasonló, ám más szoftveres megoldáson nyugvó módszert használunk, mivel ennek kimunkálása időben megelőzte a fenti módszer általános bevezetését. A Szegedi Tudományegyetem hallgatói szakdolgozatukat a Modulo adminisztrációs rendszerben adják le, ami szoros kapcsolatban áll a Neptun tanulmányi rendszerrel. Könyvtárunk az ezekből a rendszerekből exportált metaadatokat és teljes szövegű fájlokat kapja meg 2012 óta, amely több tízezres kurrens gyarapodást tett lehetővé a szakdolgozatokat tároló adatbázisunk esetében. Terveink szerint a 2019-es év végére a kurrens és retrospektív projektek keretében gyarapított rekordok száma eléri az 55 ezret.

5 SZTE Diplomamunka Repozitórium. <http://diploma.biblu-szeged.hu>

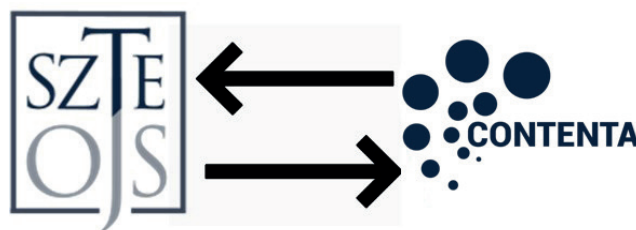


4. ábra - A szakdolgozatok metaadatainak konvertálása

2. A repozitóriumok és az OJS közötti kétirányú adatkapcsolat megteremtése

Mivel a módszer bizonyított, ezért a felmerült igények alapján elkezdtek kiterjeszteni más általunk használt szoftverekre is. A 2018-as év nyarán beindított Open Journal System folyóirat-szerkesztőségi keretrendszer használatakor már kézenfekvő lehetőségként merült fel a fentiekben bemutatott módszeren keresztüli archívumépítés a platformra beköltöző 15 folyóirat mintegy 20 ezer körüli tanulmánya esetében. Ez a következő munkafolyamatot jelentette: EPrints EP3 XML export → EP3 XML táblázattá alakítása⁶ → Manuális korrekciók → OJS natív XML fájl előállítása → OJS XML import.

A munkafolyamat kvázi megfordításával pedig az OJS platform alatt megjelenő kurrens lapszámok EPrints repozitórium alatti archiválásának folyamatos biztosítását tudjuk automatizálni: OJS DOAJ Export Plugin XML → XML fájl táblázattá alakítása⁴ → Manuális korrekciók → EP3 XML fájl előállítása → EPrints XML import.



5. ábra - Az SZTE OJS folyóirat-platform és az SZTE Contenta rendszerek közötti kétirányú adatkapcsolat

6 Convert CSV/Excel To... <http://www.convertcsv.com>

3. Tömeges adatbetöltési lehetőségek az Omeka rendszer esetében

A 2019-es év tavaszán elindított SZTE Képtár és Médiatéka⁷ szolgáltatás alapjául az Omeka Classic⁸ nyílt forráskódú rendszert választottuk. Több tízezernyi fotónk metaadatai rendelkezésre álltak MARC formátumban, ezért mindenképpen meg kellett teremtenünk ezek Omekába való injektálásának lehetőségét. Mivel az Omeka rendszer CSVImport+ pluginja metaadatokat, fájlokat és geokódokat is tud importálni, továbbá egy rekordhoz több fájl is betölthető egyszerre, sőt akár az adatok csoportos módosítására is lehetőség van, valamint az adminisztrátori felületen visszavonható a betöltés, így nagy rugalmasságot biztosítottak ezek a funkciók. Ehhez „csupán” arra volt szükség, hogy a Bodza rendszerből kiexportált MARC rekordokat a MarcEdit ingyenes szoftvercsomag⁹ segítségével – illetve a megfelelő MARC mezők és almezők Dublin Core megfeleltetésével – import-kompatibilis, táblázatos formára alakítsuk. A kurrens, kötegetelt betöltéseknél már eleve alkalmazható ez a táblázatos forma, hiszen az említett CSVImport+ plugin kiválóan fogadni tudja az így betöltött rekordokat.

Dublin Core:Identifier	Dublin Core:Creator	Dublin Core:Title	Dublin Core:Medium	Dublin Core:Date	Dublin Core:Extent
shvoy-001-001-001	Shvoy Kálmán	Autótúra 1930. Ausztria, Németország (dél)	cimlap	1930 p. 1.	32,1x24,5 cm (3702
shvoy-001-003-000	Shvoy Kálmán	1. album, 3. oldal	oldalkép	1930 p. 3.	35x26 cm (3816x25
shvoy-001-003-001	Shvoy Kálmán	Aspang nyaralóhely	fénykép	1930 p. 3.	13x8,1 cm (2656x16
shvoy-001-003-002	Shvoy Kálmán	Aspang nyaralóhely	fénykép	1930 p. 3.	13x8,1 cm (2600x16
shvoy-001-003-003	Shvoy Kálmán	Aspang nyaralóhely	fénykép	1930 p. 3.	13x8,2 cm (3420x21
shvoy-001-003-004	Shvoy Kálmán	Aspang nyaralóhely	fénykép	1930 p. 3.	13x8 cm (3368x209
shvoy-001-004-000	Shvoy Kálmán	1. album, 4. oldal	oldalkép	1930 p. 4.	35x26 cm (3264x23
shvoy-001-004-001	Shvoy Kálmán	Alsó Aspang nyaralóhely	fénykép	1930 p. 4.	13x8,1 cm (3444x21
shvoy-001-004-002	Shvoy Kálmán	Kirchberg am Wechsel, Schneeberg	fénykép	1930 p. 4.	13,1x7,8 cm (3198x
shvoy-001-004-003	Shvoy Kálmán	Kirchberg am Wechsel A. Ausztria	fénykép	1930 p. 4.	13,4x8,6 cm (3231x
shvoy-001-004-004	Shvoy Kálmán	Kirchberg am Wechsel A. Austria	fénykép	1930 p. 4.	13,6x8,7 cm (3260x
shvoy-001-005-000	Shvoy Kálmán	1. album, 5. oldal	oldalkép	1930 p. 5.	35x26 cm (3012x22
shvoy-001-005-001	Shvoy Kálmán	Kirchberg am Wechsel	fénykép	1930 p. 5.	13,5x8,6 cm (2890x
shvoy-001-005-002	Shvoy Kálmán	Kirchberg am Wechsel : Kilátás a Kernstockwarteről a	fénykép	1930 p. 5.	13,5x8,6 cm (1853x
shvoy-001-005-003	Shvoy Kálmán	Kirchberg am Wechsel : St. Wolfgang templom	fénykép	1930 p. 5.	8,7x13,5 cm (1470x
shvoy-001-005-004	Shvoy Kálmán	Kirchberg am Wechsel	fénykép	1930 p. 5.	13,8x8,7 cm (3168x

6. ábra - Az Omeka rendszerbe szánt Dublin Core adatelemek táblázata

4. A MARC formátum és az EPrints lehetséges kapcsolódási pontjai

A támogatását vesztett Bodza keretrendszer kiváltása során szembekerültünk azzal a problémával, hogy nagy tömegű (összességében százazres nagyságrendű rekordszámról van szó) MARC rekordot kellett EP3 XML formátumra konvertálni. Szintén a Bodza alól való kiköltözés igényét erősítette a Java Applet támogatásának kivezetése a böngésző programokból, mivel így megszűnt a Bodza MARC szerkesztői felülete. Néhány érintett adattárunk: SZTE Egyetemi Kiadványok, SZTE Miscellanea, SZTE UnivHistória, DélmagyarArchiv, Földrajzi Közlemények, Magyar Nyelvű Filozófiai Irodalom. A Bodzában tárolt nagy számú MARC21 alapú rekord konverziója EP3 XML formátumra egy saját fejlesztésű Java alkalmazás segítségével történhetett meg.

7 SZTE Képtár és Médiatéka. <https://mediateka.ek.szte.hu>

8 Sirhán Bálint. Repozitóriumiépítés: válasszuk az Omeka open source rendszert! Tudományos és Műszaki Tájékoztatás 64. 12. sz. (2017) 619-622.

9 MarcEdit. <https://marcedit.reeset.net>

100 \$a Tóth Sára

```

<creators>
  <item>
    <name>
      <family> Tóth </family>
      <given> Sára </given>
    </name>
  </item>
</creators>

```

7. ábra – Egy kiragadott MARC-EP3 XML konverziós példa részlete

5. A repozitóriumok közös kereshetőségének megteremtése VuFind alapokon

A korábban alkalmazott Bodza-keretrendszer nemcsak metaadatokat és teljes szövegű állományokat szolgáltatott, hanem egyúttal több különálló adatforrás közös keresőjeként is működött, amely funkció a Bodza kivezetésével ellátatlanul maradt, ugyanakkor érzékelhető módon erre a szolgáltatásra folyamatos igény mutatkozik a felhasználók részéről. Ezért a korábban alkalmazott megoldást megkíséreltük kiváltani az EPrints repozitóriumaink VuFind alapú közös kereshetőségének biztosításával. A megfelelő szoftveres háttér kialakításához az ötletet egy 2016-os Networkshop előadás adta¹⁰. Informatikusaink különböző megfontolások miatt a MARC szabvány mellett tették le voksukat az adatcsere formátumát illetően, ezért már a kísérletezés elején megvizsgáltuk a GitHub-on megtalálható EPrints MARC export-import eszközt¹⁰. Sajnos ennek használata körül adódtak nehézségek, mivel egy több mint tíz éves kódról van szó, melyet az akkori EPrints verziókhoz fejlesztettek¹¹. Ilyen nehézség volt például, hogy a megfelelő MARC mezőbe a rekord exportálásának időpontja íródott az eredeti létrehozási dátum helyett, illetve a tesztek során az OAI exportot követően az Apache webszerver többször lefagyott. Az eszköz szerencsére parancssorból is működőképesnek bizonyult, ami végül eredményre vezetett. Ütemezett feladatok segítségével így is biztosítható a VuFind alapú közös kereső friss metaadatokkal való folyamatos ellátása.

¹⁰ EPrintsMARC. <https://github.com/eprintsug/EPrintsMARC>

¹¹ Neugebauer, Tomasz, and Bin Han. [Batch Ingesting into EPrints Digital Repository Software](#). Information Technology and Libraries 31. 1. sz. (2012) 113-125.


```
<collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns="http://www.loc.gov/MARC21/slim" xsi:schemaL
ocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
<record>
<leader>01329nam a2200265 i 4500</leader>
<controlfield tag="001">1567</controlfield>
<controlfield tag="005">20180522121405.0</controlfield>
<controlfield tag="008">
120725s2008 hu om 0|| hun d
</controlfield>
<datafield tag="040" ind1=" " ind2=" ">
<subfield code="a">SZTE Doktori Repozitórium</subfield>
<subfield code="b">hun</subfield>
</datafield>
<datafield tag="100" ind1="1" ind2=" ">
<subfield code="a">Tóth Sára</subfield>
</datafield>
```

8. ábra – Az SZTE Doktori Repozitórium egyik rekordjából konvertált MARC XML egy részlete

Mivel célunk többféle adatforrásból egy közösen kereshető adathalmaz létrehozása volt, amely célkitűzés szinte magában hordozza a duplumok problémáját, ezért foglalkozni kellett az esetleges duplumrekordok kérdéskörének kezelésével is. Ezzel kapcsolatban szintén a GitHub-on találtunk megoldást a RecordManager nevezetű fejlesztés formájában¹². Az algoritmus leírása és az ellenőrzési szempontok részletesen megtalálhatóak a hivatkozott oldalon.

A VuFind alapú kereső létrehozásával a metaadatok közös kereshetőségének biztosítása mellett természetesen a teljes szövegű indexelést is szeretttük volna megoldani. Ezt különböző kiegészítő szoftverkomponensek segítségével lehet megvalósítani, az egyik ilyen például az Apache-projekt részét képező Tika elnevezésű megoldás¹³. A teljes szövegű kereséshez a Tika megfelelő telepítése és paraméterezése mellett a VuFind konfigurációs fájljaiban is el kellett végezni néhány beállítást.

12 RecordManager Deduplication.
<https://github.com/NatLibFi/RecordManager/wiki/Deduplication>

13 Apache Tika - a content analysis toolkit. <http://tika.apache.org>



A Contentas névre hallgató közös kereső jelenleg tesztüzemben működik, két repozitóriumunk (SZTE Doktori Repozitórium és SZTE Publicatio Repozitórium) teljes állománya található meg benne. A rendszer a <http://contentas.bibl.u-szeged.hu> URL címen keresztül nyilvánosan kipróbálható. Terveink szerint hamarosan a Contenta rendszer (amely jelenleg 12 független adattárból áll) tartalmának teljes egésze be fog kerülni.

The screenshot displays the Contentas search interface. At the top, the Szegedi Tudományegyetem logo and the Klebelsberg Kuno Könyvtára name are visible. A search bar contains the text 'Csapó Benő', with a dropdown menu showing 'Összes mező'. A 'Keresés' button is present. Below the search bar, the search results are displayed. On the left, there is a sidebar for 'Keresés szűkítése' (Refine search) with filters for 'Adatforrás' (Data source), 'Formátum' (Format), and 'Szerző' (Author). The 'Adatforrás' filter shows 'SZTE Publicatio Repozitórium' with 368 results and 'SZTE Doktori Repozitórium' with 65 results. The 'Formátum' filter shows 'Cikk' (Article) with 179 results, 'Fejezet, tanulmány' (Chapter, study) with 143 results, and 'Könyv' (Book) with 111 results. The 'Szerző' filter shows 'Szegedi Tudományegyetem' with 357 results. The main search results area shows two results. Result 1 is titled 'Klasszikus énekesi hangképzés empirikus kutatása, az orr és melléküregei bekapcsolhatóságának vizsgálatára' by Szerző Altörjay Tamás, published in 2019. It is a book (Könyv) and has three 'Dokumentum-elérés' (Document access) links. Result 2 is titled 'How to Make Learning Visible through Technology The eDia-Online Diagnostic Assessment System /' by Szerző Molnár Gyöngyvér, published in 2019. It is a chapter/study (Fejezet, tanulmány) and has one 'Dokumentum-elérés' link. The search results are sorted by 'Dátum (csökkenő)' (Date (decreasing)).

9. ábra – Találati lista a VuFind alapú Contentas közös keresőben

**Magánvállalkozás és közszolgálat:
A biográfiai index műfaji megújítása a Magyar Életrajzi Kalauzban**

Németh Tibor
Kemenesaljai Művelődési Központ és Könyvtár
mabi@cellkabel.hu

Private company and public service : The renewal experiment of the genre of the biographical index in the Hungarian Biographical Guide. I summarise 34 years of experience in the light of the results: database of the 'Hungarian Biographical Guide' currently maintaining the data of 400,000 people. In 1985 I began to make catalogue cards of the data of people connected to my native land, Kemenesalja, and soon came across a lot of persons unknown for the locals. I have convictions that by broadening the coverage, it is worth proceeding to a comprehensive, a national registry. The creation of a representative national digital biography index seemed for a long time to be an impossible mission with work done during leisure time and without research background.

My experience entitles me to inspect the problem of the genre and the critical analysis of the initiatives. In my lecture I concentrate on the content elements and the conceptual relationships of this complex form. I scrutinise the loosely defined concept, then I analyze the novelties of the "Hungarian Biographical Guide" genre. I discuss the utilisation opportunities and finally, I supplement my study with personal remarks.

Keywords: Hungarian Biographical Guide, biography index, national digital database

Bevezetés

Előadásomban megpróbálom vázlatos körképet adni a „Magyar Életrajzi Kalauz” reprezentatív nemzeti digitális névkataszter közel három és fél évtizedes történetéről, fejlődésének legfontosabb állomásairól, jelenlegi helyzetéről illetve a jövő lehetséges forgatókönyveiről¹.

¹ A részletek iránt érdeklődőknek: A biográfiai index műfajának megújítási kísérlete a Magyar Életrajzi Kalauzban. 1-2. rész. *Könyvtári Figyelő*, 27. (63.) 4. sz. (2017), 519-539. (<http://ki2.oszk.hu/kf/2018/01/a-biografiai-index-mufajanak-megujitasi-kiserlete-a-magyar-eletrajzi-kalauzban/#more-11291>) (Hozzáférés: 2019. március 14.) + 28. (64.) 2. sz. (2018), 229-245. (<http://ki2.oszk.hu/kf/2018/07/a-biografiai-index-mufajanak-megujitasi-kiserlete-a-magyar-eletrajzi-kalauzban-2/#more-11658>) (Hozzáférés: 2019. március 14.) További tájékozódáshoz javasolom egyetemi szakdolgozatomat: A biográfiai index, mint az életrajzi tájékoztatás integratív műfaja, különös tekintettel a Magyar Életrajzi Kalauzra. Szombathely, 2014. 150 p. (http://cellbibl-digit.cellkabel.hu/nemeth_tibor_-_biografiai_index.pdf) (Hozzáférés: 2019. március 14.)



1. Előzmények

Csöngéről, egy kis kemenesaljai faluból származom, amely születésem előtt fél évszázaddal egy másik könyvtáros gyermekkorának is helyszíne volt. Weöres Sándor azonban nem a bibliotékákban dolgozva alkotott maradandót, hanem a líra területén. 18 éves korom óta dolgozom a celldömölki könyvtárban „mezei” könyvtárosként és egy másik kemenesaljai híresség, Berzsenyi Dániel egyik munkájának címét citálva, az általam létrehozott névkataszter semmi másról, csupán a „mezei szorgalomról” szól... A nyolcvanas évek közepén a szerény számú lexikonokat lapozgattam a helyi személyeket gyűjtve és rengeteg helyben számon sem tartott érdekes emberre bukkantam. Úgy döntöttem, hogy módszeresen elkezdem kicédulázni a forrásokat immár a teljes körű feltárás igényével katalóguscédulákra, ahol minden adattípusnak meghatározott helye volt a kartonon.

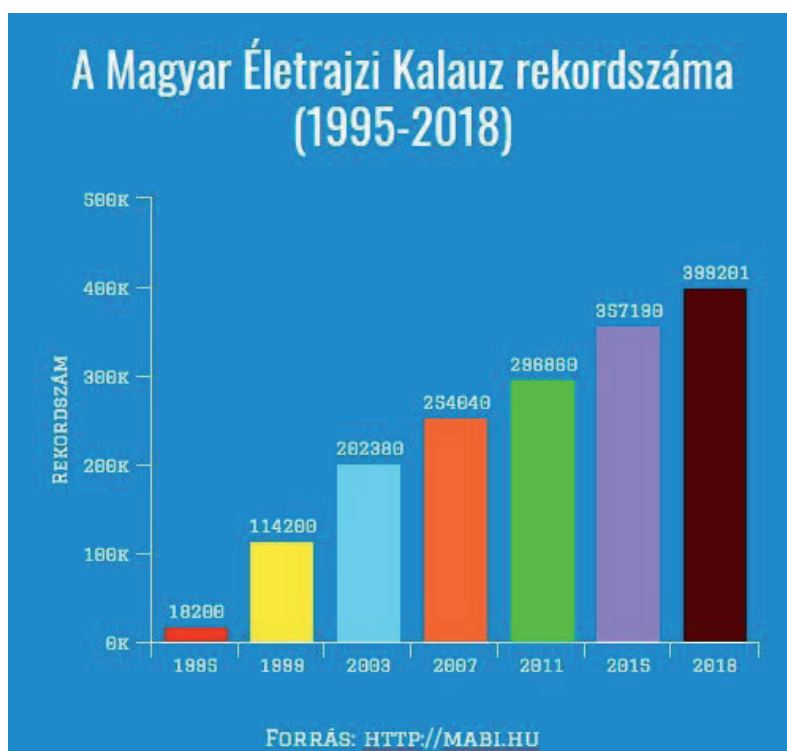
2. Műfaji dilemmák

Már jócskán benne jártam az adatgyűjtésben, amikor elkezdett érdekelni, hogy mi lehet a műfaji meghatározása az általam lelkesen végzett munkának. Sikertől beazonosítanom, hogy adattáram a legközelebb a biográfiai indexhez áll. Bizonytalanságom azonban nem tűnt el, csak fokozódott, hiszen e tájékoztatási formának eléggé pongyola definíciója létezik, amely egzakt módon nem rögzíti, csupán körülírja a feltárandó adattípusokat. A minimálisan szükséges 4-6 alapadat azonban jóval kevesebb volt az általam alkalmazott 16 féle szempontnál, másrészt jelentősebb adattömeg esetén – mint majd látni fogjuk – már a beazonosítást sem teszi egyértelműen lehetővé.

Az alap szempontok áttekintése után nézzük meg, hogy mit jelentenek az elvek a gyakorlatban. A műfaj nemzetközi reprezentánsa, a World Biographical Index nem tartalmaz születési helyet, a születési és halálozási dátumnál a MÉK napi pontosságú regisztrációjától eltérően csak az évet adja meg. A puritán alapfilozófiából adódóan a születési hely mellett figyelmen kívül hagy olyan, a Magyar Életrajzi Kalauzban feltárt adattípusokat, mint az álnevek, adateltérések, kitüntetések és egyéb (kiegészítő) adatok. Az összehasonlításból nem túl merész következtetések vonhatók le a két adatbank információs potenciáljára vonatkozóan...

3. Az adatrögzítés eredményei

A cédulákon lévő anyag retrospektív számítógépes rögzítése 1994-ben kezdődött el, miután a Lakitelek Alapítvány támogatása jóvoltából sikerült vásárolnom egy akkoriban korszerűnek számító számítógépet. Az előkészített tételek ütemes feldolgozásával dinamikusán növekedett a rekordszám. Az ezredfordulón meghaladta a 100.000-es határt és a legfontosabb, ún. magforrások CD-n is hozzáférhetővé váltak. Napi könyvtárosi munkám mellett szabadidőmben reggel és este, hétvégén, szabadságom alatt otthon végeztem az adatgyűjtést és rögzítést 2003-tól kezdődően már online, azonnal hozzáférhető formában. 2019 januárjában sikerült átlépnem a 400.000 rekordot.



1. kép: A Magyar Életrajzi Kalauz rekordszáma (1995-2018)

Az eredmények elérésében Ősze László informatikus segítette a kezdetektől munkámat illetve az internetes hozzáférést a Cellkabel nevű helyi kábelszolgáltató cég térítésmentesen bocsátotta rendelkezésemre, ezzel támogatva a nemzeti névkataszter fejlesztését.

A „Magyar Életrajzi Kalauz” feltöltöttségi áttekintését 400.000 rekord elérésekor bemutató táblázat léte már önmagában kuriózumnak számít a számítógépes adatbázisok világában. A fejlesztők ilyen transzparens módon nem szokták közzétenni az egyes adatmezők teljességét vagy egy rekordra vonatkozó tételátlagát. Bőven lehetne elemezni az adatsorokat, amelyek közül kettőre hívom fel a figyelmet. A születési helyek között több mint 16000 települést tartunk nyilván, ami elképesztő szám akkor is, ha rendszerünk nem csak a mindenkori Magyarországon születetteket, hanem a világon bárhol megtalálható magyar származásúakat is gyűjtőkörébe tartozónak érzi. A hivatkozások vonatkozásában az élő hivatkozások számát jelentősen, mintavételeink alapján 20-25%-kal meghaladja a valaha feldolgozott forrásoké, mert az időszakonként frissített segédanyagok tekintetében mindig csak az aktuális, legújabb kiadást adjuk meg, amely elvileg magában foglalja az előzményeket is.

A Magyar Életrajzi Kalauz átvilágítása 400000 személy elérésekor (2019.01.29.)

Keresési szempont (mezőnév)	Törzs file (témák száma)	Feltöltöttség (összes rögzítés)	Teljesség (% vagy tételátlag)
Szül. hely (megye)	177	296992	74,24%
Születési helység	16033	296762	74,19%
Műk. hely (megye)	144	61761	15,44%
Működési helység	8070	58486	14,62%
Foglalkozás	978	434867	1,0871
Felsőoktatási tey.	9	18493	4,62%
Tudományos fokozat	56	17505	4,37%
Akadémiai tagság	5	2523	0,63%
Kitüntetés	1692	44490	11,12%
Névváltozat	96832	96832	24,20%
Álnév	9543	9543	2,38%
Adateltérés	28382	33506	8,37%
Egyéb adat	17992	154198	38,54%
Hivatkozás	796	734839	1,8370

2. kép: A Magyar Életrajzi Kalauz átvilágítása (2019.01.29.)

4. A "Magyar Életrajzi Kalauz" hat műfaji újdonsága

Az összegző táblázat rávilágít a „Magyar Életrajzi Kalauz” első műfaji újdonságára, ami az adattípusok széles terjedelmét és sokszínűségét jelenti. Második műfaji novumként jelenik meg a nemzeti digitális névkataszter átláthatósága és feltöltöttségének dokumentálása.

A transzparencia kérdésköre átvezet a forrásanalízis kulcskérdéséhez: ki kerülhet be az adattárba az egyes forrásokból? Adatfelvételi filozófiánknak megfelelően a forrásokba felvett személyek esetében további szelekciós technikát nem alkalmazunk. Ez az a harmadik műfaji újdonság, amellyel minden eddigi hasonló jellegű kezdeményezéstől markánsan különbözik adatbankunk. Keresési módszerünk során egyértelműen jelezzük, hogy egy adott forrásból hány személy került feldolgozásra. Két külön analitikus mellékadattárat is építünk: MÉKSAJT elnevezéssel 2001. január 1-től napilapok alapján végezzük a halálozások és a kitüntetések karbantartását, a MÉKADAT pedig 2007. január 1-től az internet monitorozásával ugyanezeket az adattípusokat aktualizáljuk.

A sok műfaji újdonság mellett a legfontosabb alapelv: meggyőződésünk, hogy nincs jogunk felülbírálni az adott művet összeállítók szakmai kompetenciáját. Az életrajzi dokumentációs vagyont feltáró tevékenységünkkel megpróbáljuk az önálló szócikkkel vagy adatsorral rendelkező személyeket teljességgel beépíteni adatbázisunkba.

Az analitikus feltárás mélységének hasznosságára egyetlen példa: ha egy Nagy Ilona névváltozattal rendelkező illetőt kell beazonosítanunk, akkor jelenleg rendszerünkben 41 ilyen személyt találunk. A névváltozat adatmezőben láthatjuk, hogy közülük öten is 1950-ben születtek. A hónapra / napra történő precíz feltárás

nélkül vajon megkülönböztethetők lennének? A negyedik műfaji újdonságunk: az egyes adattípusokon belül mindig le kell hatolni a legkisebb azonosítható egységig.

KERESÉS

Keresett személy neve : Sz

Név hasonlítás : Mező elejével egyező

Születés : 1901 .év 01 .hó 01 .nap

Után : .év .hó .nap

☒ és ☐ vagy ☐ de nem

Nincs kiválasztva : .év .hó .nap

Nincs kiválasztva : .év .hó .nap

Kiválasztott feltételek : Előző Következő

K 1. [---] Születési_hely_(megye/ország) [] Vas

K 2. [Es] Foglalkozás [] történész

K 3. [Es] Születési_hely_(helység) [Nem] Szombathely

K

K

K

K

K

K

K

Maximum találatok száma : 250

TALÁLATOK

Találatok új ablakba

Találatok: 3 (1-3)

Keresett személy neve	Születési dátum	Info	Halálozási dátum	Info
Szabó Béla Lőrinc	1923.08.10		1982.02.01	
Szakonyi István	1911.07.10		1974.09.28	
Szemere Vera	1916.07.02		1992.05.07	

3. kép: A Magyar Életrajzi Kalauz kereső űrlapja és találati listája

A hatékony visszakeresésre szolgál bizonyítékként az a példa, amelyben „Sz” betűvel kezdődő nevű személyeket keresünk, akik a XX. században, azaz 1901. január 1. után születettek Vas megyében, de nem Szombathelyen és történész volt a foglalkozásuk.

Nem túl bonyolult a minta, de nehéz lenne pár másodperc alatt megtalálni azt a 3 személyt, akit rendszerünk a feltételeknek megfelelően lekeresett. A Boole-algebra hatékony alkalmazása mellett érdemes figyelni arra, hogy a keresési stratégia kiválasztásának megkönnyítése érdekében olyan megjelenítést alkalmazunk, amelyben az adott ismerv mellett megtalálható a hozzá tartozó személyek folyamatosan aktualizált száma. Ezzel olyan komplex keresési teret nyitunk meg, amely variabilitásával és rugalmasságával a tematikus keresőrendszerek között rendszerünk ötödik műfaji újdonságát jelenti.

Nem a fő csapásirányt jelenti a feltárás során, de ha az adott személynél olyan érdekes tényre bukkanunk, amelynek visszakereshetőségét fontosnak érezzük, akkor az egyéb adat kategóriában hozzáférhetővé tesszük. Ennek klasszikus megjelenési formája, ha valaki valamiben az első magyar volt. Jelenleg 264 ilyen megjelölés szerepel adattárunkban s talán ez az adatbányászati gyöngyszem jól jelzi, hogy mennyire nem végezhető mechanikusan az adatok rögzítése.



Látványosan tükrözi rendszerünk hatodik műfaji újdonságát a foglalkozások hierarchikus rendszerének összehasonlító táblázata a World Biographical Index és a Magyar Életrajzi Kalauz között. Ég és föld a különbség a magyar adattár javára a sport kategória feltártságában. Adatbankunk saját fejlesztésű, rugalmasan hierarchizált foglalkozási tárgyszórendszert használ, amely hatékonyan segíti a használót a tematikus keresések megtervezésében. Az indokolt esetben többes besorolásoknak lehetnek vitatható elemei, de ezek eltörpülnek a rendszer alkalmazásának gyakorlati előnyei mellett.

World Biographical Index		Magyar Életrajzi Kalauz	
<input type="checkbox"/> ART (600)	990320	6. SPORT	6.10.9. Ökölvívó
<input type="checkbox"/> Musicians (610)	262730	6.1. Aréna	6.10.10. Szumó
<input type="checkbox"/> Composers (611)	66393	6.1.1. DOBÓATLÉTA	6.10.11. Tekvandó
<input type="checkbox"/> Conductors, Musical Directors (612)	30534	6.1.1.1. Diszkoszvető	6.10.12. Vívó
<input type="checkbox"/> Instrumentalists, Orchestral Musicians (613)	68701	6.1.1.2. Gerelyhajító	6.10.12.1. Kardvívó
<input type="checkbox"/> Singers (614)	62344	6.1.1.3. Kalapácsvető	6.10.12.2. Parbajtorvívó
<input type="checkbox"/> Performing Artists (620)	161456	6.1.1.4. Súlylökő	6.10.12.3. Törvívó
<input type="checkbox"/> Stage Managers, Directors (621)	22111	6.1.2. Gyulogó	6.10.12. Zen Bu Kan Kempo harcművész
<input type="checkbox"/> Actors (622)	110179	6.1.3. Haromszázó	6.11. LABDAJÁTÉKOK
<input type="checkbox"/> Dancers (623)	11683	6.1.4. Hetpróbázó	6.11.1. Asztaliteniszező
<input type="checkbox"/> Other Stage Occupations (624)	10040	6.1.5. Kalapácsvető	6.11.2. Biliárdozó
<input type="checkbox"/> Plastic Artists (630)	566134	6.1.6. Ötpróbázó	6.11.3. Fallabdázó
<input type="checkbox"/> Painters, Draughtsmen, Graphic Artists (631)	361736	6.1.7. Súlylökő	6.11.4. Floorballzó
<input type="checkbox"/> Sculptors, Carvers, Founders, Modelers (632)	83327	6.1.8. Tízpróbázó	6.11.5. Golfzó
<input type="checkbox"/> Book Artists and Calligraphers (633)	19330	6.1.9. UGRÓATLÉTA	6.11.6. Gyeplabdázó
<input type="checkbox"/> Handicraft Artists (634)	71591	6.1.9.1. Hármassugró	6.11.7. Kézilabdázó
<input type="checkbox"/> Photographers (636)	9391	6.1.9.2. Magassugró	6.11.8. Kosárlabdázó
<input type="checkbox"/> FILM, RADIO AND TELEVISION (650)	48610	6.1.9.3. Rúdsugró	6.11.9. Labdarúgó
<input type="checkbox"/> RELIGION (700)	843811	6.1.9.4. Távolugró	6.11.10. Látnesiszező
<input type="checkbox"/> Christian and Jewish Religion (702)	801738	6.2. Edző	6.11.11. Lovaspólozó
<input type="checkbox"/> Theologians, Scholars of Religion (705)	60231	6.3. Erősportoló	6.11.12. Rögbijátékos
<input type="checkbox"/> Priests, Preachers, Ministers, Rabbis (710)	389492	6.3.1. Súlyemelő	6.11.13. Röplabdázó
<input type="checkbox"/> Church Dignitaries, Representatives, Administrators (715)	150450	6.3.2. Testépítő	6.11.14. Teniszező
<input type="checkbox"/> Members of a Religious Order (720)	163288	6.4. GEPIARMUSPORT	6.11.15. Tollaslabdázó
<input type="checkbox"/> Saints, Martyrs (724)	14172	6.4.1. Autómodellező	6.11.16. Vízilabdázó
<input type="checkbox"/> Christian and Jewish Religion: Others (725)	24105	6.4.2. Autóversenyző	6.12. LOVASSPORT
<input type="checkbox"/> Islamic Ecclesiastics (730)	33539	6.4.3. Gokartversenyző	6.12.1. Díjlovagló
<input type="checkbox"/> Hindu Religion (740)	245	6.4.4. Kerékpárversenyző	6.12.2. Díjugrató
<input type="checkbox"/> Buddhist Religion (750)	8052	6.4.5. Motorcsónakversenyző	6.12.3. Fogathajtó
<input type="checkbox"/> Other Religions (790)	237	6.4.6. Motorkegypárversenyző	6.12.4. Lovas
<input type="checkbox"/> SPORTS AND LEISURE (800)	86787	6.5. GIMNASZTIKA	6.12.5. Lovaspólozó
<input type="checkbox"/> Sportsmen (801)	66550	6.5.1. Aerobicversenyző	6.12.6. Lovastornász
<input type="checkbox"/> Toreros (810)	5212	6.5.2. Fitnessversenyző	6.12.7. Lovasszázó
<input type="checkbox"/> Coaches, Managers (815)	6382	6.5.3. Kötelugró	6.13. LOVÉSZET
<input type="checkbox"/> Sports: Others (820)	8643	6.5.4. Művészeti tornász	6.13.1. Íjász
		6.5.5. Sportakrobata	6.13.1.1. Szímszeríjász
		6.5.6. Tornász	6.13.2. Sálvó
		6.5.7. Versenytáncos	6.13.3. Sportlövő
		6.6. GORSPORT	6.14. Mászó
		6.6.1. Gorkorcsolyázó	6.15. MODELLEZŐSPORT
		6.7. Játékvezető	6.15.1. Autómodellező
		6.8. KARTYAJÁTEK	6.15.2. Hajómodellező
		6.8.1. Búdzsező	6.15.3. Radiómodellező
		6.8.2. Tarokkozó	6.16. Sportorvos
		6.9. Kerékpárversenyző	6.17. Sporttörténész
		6.9.1. Motorkegypárversenyző	6.18. Sportújságíró
		6.9.2. Salákmotorkerékpáros	6.19. Sportvezető
		6.10. KÜZDŐSPORT	6.20. TABLAJÁTEK
		6.10.1. Aikidó	6.20.1. Dartsversenyző
		6.10.2. Birkózó	6.20.2. Sakkozó
		6.10.3. Bó jitsu kai versenyző	6.20.2.1. Hexasakkkozó
		6.10.4. Cselgacssozó	6.20.2.2. Sakófeladványyszerző
		6.10.5. Karetézó	6.21. TEKEJÁTEK
		6.10.6. Kendó	6.21.1. Bocsaversenyző
		6.10.7. Kickboksoló	
		6.10.8. Kung-fu	
			6.21.2. Bowlingozó
			6.21.3. Petanque-versenyző
			6.21.4. Tekező
			6.22. TERMÉSZETI SPORT
			6.22.1. Hegymászó
			6.22.2. LÉGISPORT
			6.22.2.1. Ejtőernyős
			6.22.2.2. Légömbrepülő
			6.22.2.3. Radióvárász
			6.22.2.4. Repülőmodellező
			6.22.2.5. Sárkányrepülő
			6.22.2.6. Síkiernyős
			6.22.2.7. Sportrepülő
			6.22.2.8. Vitorlázórepülő
			6.22.3. Szkiing
			6.22.4. Tájfutó
			6.22.5. TELISPORT
			6.22.5.1. Bobversenyző
			6.22.5.2. Gyorskorcsolyázó
			6.22.5.3. Jégkorongozó
			6.22.5.4. Kutyaszánkversenyző
			6.22.5.5. Műkorcsolyázó
			6.22.5.6. Sielő
			6.22.5.7. Sífutó
			6.22.5.8. Sílövő
			6.22.5.9. Síugró
			6.22.5.10. Snowboardozó
			6.22.6. VÍZISPORT
			6.22.6.1. Buvár
			6.22.6.2. Búvárúszó
			6.22.6.3. Evezős
			6.22.6.4. Hajómodellező
			6.22.6.5. Horgász
			6.22.6.6. Kajakozó
			6.22.6.7. Kajakpólos
			6.22.6.8. Kenu
			6.22.6.9. Kite versenyző
			6.22.6.10. Motorcsónakversenyző
			6.22.6.11. Műugró
			6.22.6.12. Szinkronúszó
			6.22.6.13. Szörföző
			6.22.6.14. Úszó
			6.22.6.15. Vitorlázó
			6.22.6.16. Vízilabdázó
			6.22.6.17. Vízisíz
			6.23. Testnevelő
			6.24. TÖBBPRÓBÁZÓ
			6.24.1. Hetpróbázó
			6.24.2. Ötpróbázó
			6.24.3. Öttusó
			6.24.4. Quadriatlionista
			6.24.5. Tízpróbázó
			6.24.6. Triatlionista

4. kép: A Magyar Életrajzi Kalauz és a World Biographical Index összehasonlító foglalkozási listája (sport)

A nyomtatott források mellett egyre inkább fókuszálnunk kell az internet biográfiai erőforrásainak rendszerszemléletű feltárására is. E folyamatos pásztázás egyrészt a már említett MÉKADAT segédállomány építését jelenti, másrészt a csak a világhálón hozzáférhető források feldolgozását (pl. Fejér Megyei Életrajzi Lexikon²), harmadrészt a kéziratos formában, kis példányszámban megjelent adattárak digitális forrásainak feltérképezését (pl. Evangélikus lelkészek Dunántúlon, 1793-1920³), negyedrészt pedig a már feldolgozott források új, internetes elérhetőségének

² <http://konyvtar.vmk.hu/fejerlex/fejerlexikon.htm> (Hozzáférés: 2019. március 14.)

³ <https://medit.lutheran.hu/site/konyv/1567> (Hozzáférés: 2019. március 14.)

feltüntetését. A világhálós források rendkívül illékonyak, ezért türelmet és óvatosságot igényel a kanalizálásuk, de fontosságuk egyre nő, ezért nem kerülhető meg, hogy a névkataszter aktualizálása, karbantartása érdekében e forrástípus is feltárássra kerüljön.

Hogyan foglalható össze a Magyar Életrajzi Kalauz műfaji újításainak filozófiája? A lényeg: **magas fokú adatintegráció a használoközpontúság érdekében.**

5. Felhasználási területek

34 év, több mint 12000 nap kitartó építkezésének eredményeként jutottam el a mai szintre. Kilencedik négyéves olimpiai ciklusom felénél járok és továbbra is hiszek egykori álmom megvalósításának lehetőségében. 1982 óta dolgozom a celldömölki városi könyvtárban és 1985-től kezdve a mindenkori béremre vetítve kiszámoltam, hogy az adattárra fordított munkám az inflációs indexet is figyelembe véve reálértéken 2018. december 31-ig 94 millió 545 ezer 647 Ft-ot tett ki. Ebben benne van a kezdő 1500 Ft-os havi fizetésem és számításba veszi, hogy 32 év után 2018. április 1-től heti egy kutatónapot vehetek igénybe...

A Magyar Életrajzi Kalauz komplex jellegéből adódóan sokféle célra használható, amelyek közül ízelítőként néhányat megemlítek. Az utolsó átfogó magyar ki kicsoda, az MTI Ki kicsoda 2009 megjelenése előtt megkerestem az MTI-t, mert az előző kiadásokban azt tapasztaltam, hogy sok már évekkel, sőt évtizedekkel korábban elhunyt személyt tartalmaz az adattár. Felajánlottam a segítségemet ezek kiszűrésében illetve a legfontosabb születési adatok pontosításában. A kért összességében több ezer adatnyi információt az együttműködési szerződésben rögzítettek szerint rendelkezésre bocsátottam, de a hivatalos dokumentumban rögzítettek ellenére nem tüntették fel nevemet a szerkesztői névjegyzékben... A példát csak azért említem, mert a feldolgozás során látom, hogy milyen sok forrás esetében jelentene minőségjavulást, ha megjelenés előtt ellenőriznék adataikat egy kontroll adattárban, ami lehet a Magyar Életrajzi Kalauz vagy a Petőfi Irodalmi Múzeum grandiózus adatbáziscsaládjának egyik ékköve, a Magyar Életrajzi Index⁴.

Egy ideig személyesen is részt vettem a Libinfo tájékoztató szolgálat működésében, majd az adatbázis ingyenes rendelkezésre bocsátásával segítetttem közérdekű tevékenységüket.

Büszke vagyok, de egyúttal fáj is a szívem, ha a „Nemzeti Évfordulóink” című sorozatra gondolok. 11 esztendőn keresztül, a 2008-as kötettől a 2018-ig állítottam össze a következő év kerek, azaz 25-tel osztható évfordulós személyiségeinek alap szócikk-listáját, amelyre építve a szerkesztőség elkészítette az évkönyvet.⁵ Minden esztendőben az egyre bővülő adattáram szisztematikus átszűrésével nyaranta hónapokat foglalkoztam ezzel az előkészítéssel. A Balassi Intézet átalakulása a 2019-es esztendő kimaradt, de 2020-tól a Petőfi Irodalmi Múzeum égisze alatt némileg megváltozott koncepcióval folytatódni fog a színvonalas kalendáriumok sora.

4 Ld. Murányi Péter: Két biográfiai index adatbázis formájában. A Magyar Életrajzi Kalauz és a Magyar Életrajzi Index. *Könyvtári Levelező/Lap*, 13. 7-8. sz. (2001), 23-30.
(http://epa.oszk.hu/00300/00365/00001/pdf/KLL_EPA00365_2001-07-08_023-030.pdf)
(Hozzáférés: 2019. március 14.)

5 <http://www.balassintezet.hu/hu/kiadvanyok/nemzeti-evforduloink/>
(Hozzáférés: 2019. március 14.)



Ha a „Nemzeti Évfordulóink” szívszerelem, akkor mit mondhatnék az általam szerkesztett és jórészt írt Kemenesaljai Életrajzi Lexikonról? A városi könyvtárunk műhelyében régóta készült szócikkek eleinte a világhálóra kerültek fel és mindig vártam a pályázati lehetőséget, hogy nyomtatott formában is napvilágot láthasson. A Leader program keretében 2014 végén 2000 példányban úgy jelenhetett meg az elegáns kivitelű kötet⁶, hogy az érdeklődők ingyen juthattak hozzá. Az már csak hab volt a tortán, hogy kiadója a szülőfalum Kemenesaljai Regionális Ifjúsági Szervezete volt. Csöngei impresszummal nem hiszem, hogy túl sok kötet jelent meg valaha....

Az adattár karbantartása mellett állandóan keresem a folytatás és az együttműködések lehetőségeit. A legtermészetesebb módon a „Nemzeti Névtér”-rel⁷ alakult ki a kapcsolat. Megtiszteltetés számomra, hogy közreműködhetek ebben a vállalkozásban és bízom abban, hogy hosszú távon, erőforrásainkat koncentrálna dolgozhatunk együtt közös céljaink megvalósításáért.

A együttműködési lehetőségek következő területei még képlékenyek, de jóleső kihívásként tekintek rájuk. A Magyar Életrajzi Kalauz jellegéből adódóan a magyarságkutatás, az identitás feltárásának fontos terepe lehet. A lokális információk bőséges tárházaként természetes módon lehet együttműködő partnere a honismereti-helytörténeti mozgalomnak.

Ha nyitottak vagyunk a jövőre, akkor nem kerülhető meg a közösségi média csatornáinak felhasználása. Biztos vagyok abban, hogy az adatbank kontrollált társadalmasításával a közösségi platformok erőforrásai felszabadíthatók a minőségi együttműködés érdekében.

6. Személyes záró gondolatok

Szilárd meggyőződésem, hogy személyes áldozatvállalás nélkül nem valósulhat meg egyetlen nemzeti jelentőségű ügy sem. A feladat nagyságának jellemzéséhez talán elég arra utalnom, hogy jelen pillanatban annyi elvégzendő retrospektív munkát látok magam előtt, ami a következő 35 évre elég lenne a friss források feldolgozása nélkül is...

2014 szeptemberében részt vettem a halálának centenáriuma alkalmából megrendezett Szinnyei József-émlékkonferencián. A rendezvény előtt kilátogattam Szinnyei sírjához a Kerepesi temetőbe és az üres sírra elhelyeztem emlékező sírcsokromat. A zuhogó esőben lépten-nyomon csigákra kellett ügyelnem illetve mókuskok cikáztak a sírok között. Projektorientált világunkban furcsa mókus vagyok magam is, hiszen 34 éve munkaidőn kívül, ingyenes munkával mintegy digitális gőzhangyaként dolgozom az általam felvállalt nemes ügy érdekében, bár jómagam csigalassúságúnak tekintem az előrehaladást... Reményeim szerint a modern technológia eszköztárával sikerrel próbálom ötvözni a nagy tömegű adatkezelést a Szinnyei-Gulyás-Viczián triász⁸ által fémjelzett "kézműves alapossággal"...

6 Kemenesaljai életrajzi lexikon. Csönge, 2014. 222 p.

7 Ungváry Rudolf: A névtér, mint kulturális szükséglet. A nemzeti aggregátor kitüntetett feladata. Tudományos és Műszaki Tájékoztatás, 59. 8. sz. (2012), 320-326.
(<http://tmt.omikk.bme.hu/tmt/article/view/718/10593>) (Hozzáférés: 2019. március 14.)

8 Czigány Lóránt: Szinnyei-Gulyás-Viczián. Magyar Könyvszemle. 113. 1. sz. (1997), 104-108.

Saját sorsom és a "Magyar Életrajzi Kalauz" összefonódó históriáján keresztül élem meg azt, hogy "a Big Data már nem csupán a professzionális felhasználók kiváltsága, a tárhely és a számítási teljesítmény növekedésével és ezek árának csökkenésével a mindennapi felhasználók is részeseivé válhatnak a Big Data paradigmájának"⁹.



9 Szűts Zoltán-Jinil Yoo: Big Data, az információs társadalom új paradigmája. Információs Társadalom. 16. 1. sz. (2016), 12. (https://informaciostarsadalom.infonia.hu/index.php/inftars/article/view/6/it_2016_1_1_szuts_yoo) (Hozzáférés: 2019. március 14.)

Az elektronikus szakirodalmi információforrások hazai lelőhelyadatbázisa, a COMPASS

Sütő Péter

MTA Könyvtár és Információs Központ EISZ Titkárság
(grid.496758.1)

COMPASS, the database of scientific e-resources available in Hungary

Hungarian university- and research libraries purchase most of their electronic scientific content in the framework of consortium-level agreements via Electronic Information Service National Programme (EISZ). Apart from acquisitions, it is also the goal of EISZ to ensure the utilisation of the available content for the wide community. For this reason, the development of COMPASS database was started in autumn 2016 in order to collect, organise, and make available the scientific databases and electronic documents purchased by the consortium or the institutions individually.

In September 2017, a newly developed version of COMPASS was launched on a new platform. The database provides multiple options for searching and interacting with scientific contents available in Hungary. In addition, it offers direct access to the databases for eligible users, as well as links to the full text of Open Access articles.

The paper covers how much the goals and plans of the project have been achieved, and describes the new functionalities and the experiences of the operation and usage since the database was launched.

Keywords: union catalogue, e-resources, electronic document delivery, open access

Bevezetés

Napjainkban a tudományos szakirodalom egyre nagyobb mértékben elektronikus formában jelenik meg, amely magával hozza a tudományos közösség igényét is az elektronikus formátumban rejlő előnyök iránt. Ez a folyamat Magyarországon is jól érzékelhető, a hazai felsőoktatási- és kutatóintézetek is nagyobb részt elektronikus formában szerzik be a szakirodalmat és teszik elérhetővé a felhasználók számára.¹

A hazai kutatóközösség szakirodalommal való ellátását az Elektronikus Információszolgáltatás Nemzeti Program segíti elő azzal, hogy lebonyolítja az elektronikus tudományos információforrások beszerzését konzorciumi keretek között a hazai felsőoktatási és non-profit kutatóintézetek számára. 2019-ben az EISZ programon keresztül több mint 230 intézmény szerzi be az elektronikus információforrásait, közel 180 adatbázist, e-folyóirat- és e-könyvcsomagot, ami azt jelenti, hogy többszáz ezer tudományos publikáció érhető el elektronikusan Magyarországon konzorciumi előfizetések keretében.

¹ Dér Ádám, Lencsés Ákos. Az EISZ Nemzeti Program szerepe a könyvtárak külföldi szakirodalommal való ellátásában. Tudományos és Műszaki Tájékoztatás 64, 5. sz. (2017): 241–246.

Óriási lehetőségeket rejt magában ennek a hatalmas mennyiségű publikációnak a rendelkezésre állása, azonban felmerülnek a következő kérdések:

- Hogyan képesek ezt az információhalmazt a kutatók és a könyvtárosok áttekinteni és hatékonyan felhasználni?
- Hogyan tud ez az egyes intézményeken keresztül rendelkezésre álló forrás az ország teljes tudományos közössége számára hasznosulni?

Ezeket a kérdéseket az EISZ Titkárság munkatársai már 2016-ban is feltették maguknak és válaszként megszületett a COMPASS² továbbfejlesztésének terve, azaz egy országos elektronikusforrás lelőhelyadatbázis létrehozása.

A fejlesztés céljai³

A fejlesztés kezdetén, 2017. elején kiemelt célként fogalmaztuk meg, hogy a lelőhelyadatbázis segítse elő a szakirodalmi információkhoz való legális hozzáférés lehetőségének kiszélesítését, a források elérésének meggyorsítását, a hazai intézmények konzorciumi előfizetései mellett az egyéni előfizetések láthatóságát, mint fontos felhasználói szempontokat.

Természetesen a gyarapítási, fenntartói és tudományszervezési szempontok is hangsúlyosan megjelentek a célok között, úgymint a költséghatékonyság elősegítése, az országos tudományos tartalombeszerzés áttekinthetősége, a fejlesztésre szoruló területek feltérképezése és az intézményi kapcsolatok erősítése. Ezekkel a szempontokkal a COMPASS tervezett célja volt, hogy támogatást nyújtson akár stratégiai tervezési szinten a döntéselőkészítő folyamatokhoz is.

A fejlesztés eredménye

A célok elérése érdekében a már létező COMPASS felülethez egy PHP alapú MySQL adatbáziskezelőt használó keresőmotort fejlesztettünk, amely a már meglévő szűkös keresési lehetőségek⁴ (adatbázis, intézmény és település keresők) hatékonyságát jelentősen növelte és lehetővé tette a keresési lehetőségek kibővítését.

Figyelembe kellett ugyanis venni azt, hogy a felhasználók elsősorban az őket érdeklő folyóirat, vagy cikk adatainak vannak a birtokában, nem pedig annak, hogy az melyik adatbázisban érhető el. Tehát csak akkor működhet hatékonyan a lelőhelyadatbázis, ha a címszintű keresési lehetőség is rendelkezésre áll.

A tervezés során a következő szint, a cikkekre való keresés biztosítása is felmerült, melyet a lehetőségek vizsgálata után a Crossref OpenURL használatának segítségével sikerült megvalósítani. Ez azt eredményezte, hogy a COMPASS-on

² A COMPASS elérhetősége <http://compass.mtak.hu>

³ Dér Ádám, Nyiscsák Sándor. A COMPASS adatbázis új verziójának fejlesztése. (Networkshop 2017, Szeged) 2017. április 20. URL: <https://conference.niif.hu/event/7/session/5/contribution/52/material/slides/0.pdf>

⁴ Páll Zoltán. COMPASS - Iránytű az információhoz (az MTA KIK EISZ új információszolgáltató adatbázisa). (Networkshop 2015, Eger) 2015. április 1. URL: <https://conference.niif.hu/event/3/session/14/contribution/80/material/slides/1.pdf>



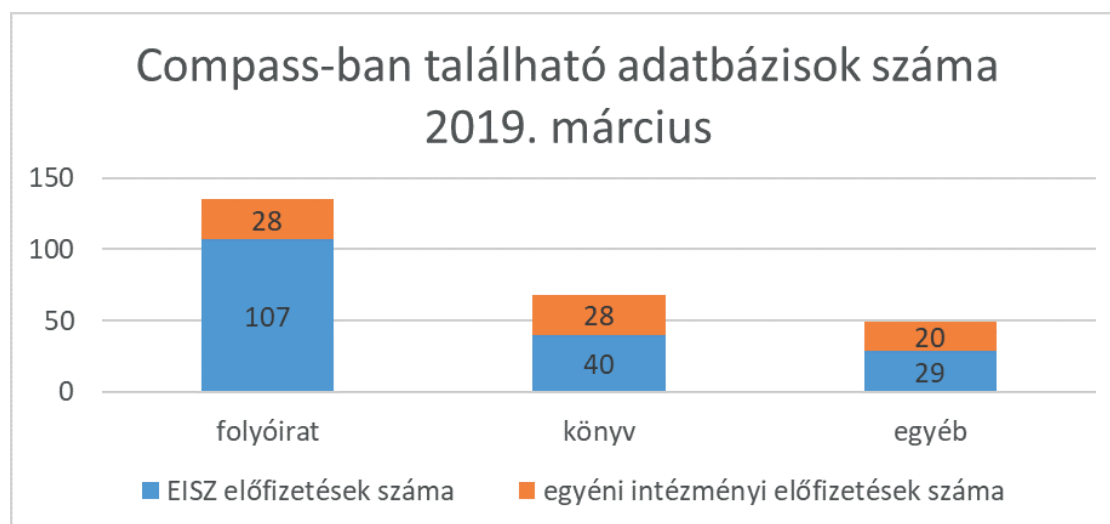
keresztül lehetővé vált a keresés a DOI azonosítóval rendelkező publikációk közel teljes halmazára.

A lelőhelyadatbázis hatékonyságának kiteljesedését pedig a cikkszintű kereséshez kapcsolódóan az open access cikkek közvetlen hozzáféréseinek biztosítása jelentette. Ez a funkció az Unpaywall (korábban oaDOI) API-n⁵ keresztül éri el a cikkek open access státuszára vonatkozó metaadatokat, amelyek automatikus feldolgozásával úgy a green, mint a gold open access cikkek teljes szöveghez vezető linkjei is megjelennek a COMPASS találati listájában.⁶

Tapasztalatok a használati adatok tükrében

A COMPASS jelenlegi verzióját 2017. szeptemberében élesítettük, azaz 1,5 év használati adatai állnak rendelkezésre.

A COMPASS jelenlegi tartalmának adatbázistípusok szerinti megoszlását (1. ábra) áttekintve jól látszik, hogy milyen nagy arányban kerültek be az intézmények egyéni előfizetései is az adatbázisba. Ez egyrészt mutatja azt, hogy jelentős mértékben sikerült bevonni az intézményeket a COMPASS tartalmi bővítésébe, egyúttal megerősíti azt a kezdeti célkitűzést, hogy a COMPASS tartalma nőjön túl a konzorciumi portfólión és törekedjen teljes országos lefedettségre.

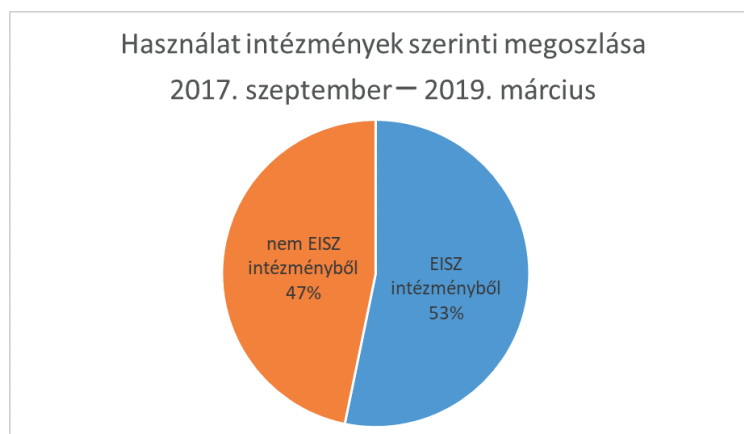


1. ábra: COMPASS-ban található adatbázisok száma, 2019. március

Mivel a COMPASS platform szabadon hozzáférhető mindenki számára ezért érdemes vizsgálni azt is, hogy honnan érkeznek a használatok (2. ábra). A célokkal összhangban azt tapasztaljuk, hogy a COMPASS-t az EISZ tagintézményeken kívüli IP tartományokból is nagy arányban használták, azaz lényegesen szélesebb használói kört ér el a lelőhelyadatbázis, mint a konzorciumi tagintézmények kutatói.

⁵ Az Unpaywall API elérhetősége <http://unpaywall.org/products/api>

⁶ Dér Ádám, Nyiscsák Sándor: A COMPASS adatbázis új verziójának fejlesztése. (Networkshop 2017, Szeged) 2017. április 20. URL: <https://conference.niif.hu/event/7/session/5/contribution/52/material/slides/0.pdf>

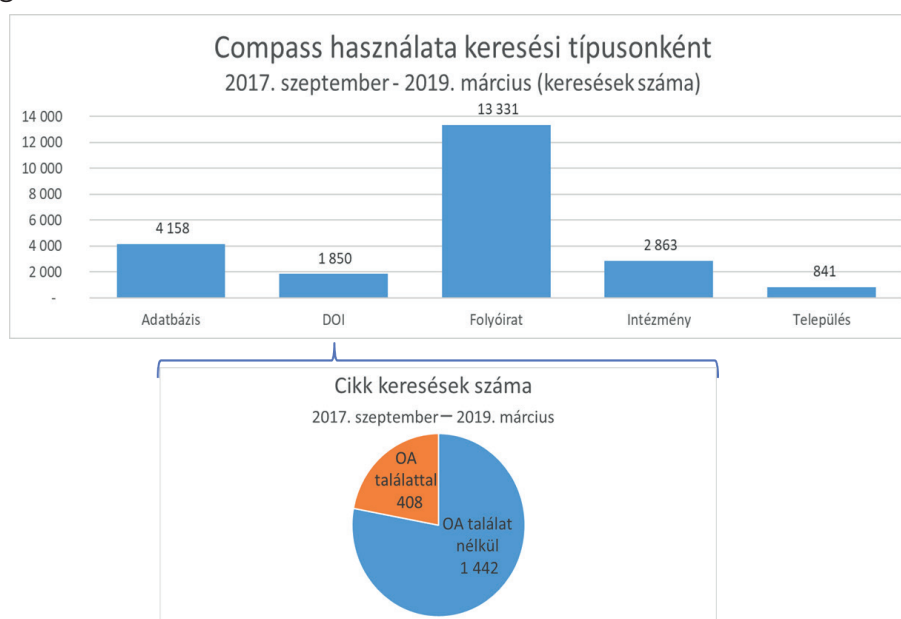


2. ábra: COMPASS használat intézmények szerinti megoszlása, 2017. szeptember – 2019. március

A COMPASS használata (3. ábra) az első 1,5 év alatt az elvárásoknak megfelelően alakult. Nem meglepő, hogy a folyóirat/könyv címre való keresés kimagaslík a többi keresési típus közül, hiszen a lelőhelyadatbázisnak ez az elsődleges funkciója. Egyúttal igazolja azt, hogy a fejlesztés egyik kulcs feladata volt a címszintű kereséssel való bővítés.

Azonban az adatok azt mutatják, hogy a cikkszintű keresések (DOI keresések) száma alacsony. Kutatói és könyvtáros visszajelzések alapján ennek lehetséges oka az, hogy a DOI használata, mint keresőkifejezés nem olyan kézenfekvő, mint egy folyóiratcím használata, még akkor sem, ha nem ugyanazon a szinten ad találatokat.

A DOI keresés alacsony használatából következik, hogy az Open Access cikk találatok száma is alacsony, és a részarányát tekintve is elmarad az Open Access cikkek globális arányától⁷. Ez rámutat arra, hogy mindenképpen növelni kell a felhasználók tudatosságát a cikkszintű keresésre.



3. ábra: COMPASS használata keresési típusonként, 2017. szeptember – 2019. március

7 Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S.: The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. PeerJ. 6:e4375 (2018) <https://doi.org/10.7717/peerj.4375>



Ha megvizsgáljuk a keresések hatékonyságát (1. táblázat), azaz a találatot adó keresések számát, akkor a keresési típusok között nagy eltéréseket tapasztalunk. A DOI keresésnek nagyon magas a találati aránya, ami a DOI keresőbe beépített intelligenciának köszönhető, mely felismeri a keresőmezőbe bevitt karaktersorozatban a DOI azonosítót. Az adatbázis és folyóirat keresések alacsonyabb hatékonysága viszont arra enged következtetni, hogy vagy a felhasználók figyelmetlenek a keresés során, vagy a felhasználói igények szélesebbek, mint a COMPASS lefedettsége.

1. táblázat: Találati hatékonyság⁸ keresési típusonként,
2017. szeptember – 2019. március

DOI keresés	91%
Intézmény keresés	91%
Település keresés	75%
Folyóirat/könyv keresés	61%
Adatbázis keresés	50%

Ahhoz, hogy jobban megismerjük az alacsonyabb keresési hatékonyság okait, elemeztük a sikertelen keresések keresési kifejezéseit a leggyengébb mutatóval rendelkező adatbázisra történő keresésnél (2. táblázat).

A vizsgálat során azt láttuk, hogy a sikertelen keresések 29%-a a hibás keresés közvetlen megismétléséből adódik, azaz a felhasználó újra lefuttatja változatlan keresőkifejezéssel ugyanazt a keresést, ami nem adott találatot, mert nem hiszi el, hogy nincs találat. Ezeket a duplumokat a hatékonyság vizsgálata szempontjából figyelmen kívül hagyhatjuk.

Az elírásból adódó sikertelen keresések alacsony aránya (2%) feltételezhetően a kereső mezőbe beépített automatikus kiegészítő (autocomplete) funkciónak köszönhető, ami mindenképpen igazolja ennek a fejlesztési elemnek a létjogosultságát.

Gyakori hiba (10%), hogy az adatbázisra kereső fülön folyóíratra, esetleg intézményre keresnek a felhasználók, illetve hasonló félreértés a kiadókra való keresés (4%). Ez részben figyelmetlenségből adódik, azaz a felhasználó elfelejt átkattintani a megfelelő fülre, másrészt félreértésből, miszerint a kiválasztott fül nem az elérni kívánt eredményt, hanem a beírandó keresőkifejezést specifikálja. Ezeket a mutatókon a keresőfelület felhasználókat segítő információinak pontosabb megjelenítésével, valamint a keresési algoritmus kiadóra és alternatív névre való kiszélesítésével várhatóan tudunk javítani.

8 (az adott kereséstípus esetén találatot adó keresések száma) / (adott kereséstípus esetén lefuttatott összes keresés száma)

A legnagyobb halmazt a hibás keresések között a tárgyszóra történő keresés (34%) teszi ki. Ez azt mutatja, hogy a felhasználók egy része nem lelőhelyadatbázisként, hanem egy általános tudományos információforrásként próbálja használni a COMPASS-t. Ezt a félreértést mindenképpen el kell oszlatni. A lelőhelyadatbázis célja nem az hogy szakirodalmat keressen általa a felhasználó a kutatásához, hanem az, hogy a már különböző forrásokban megtalált szakirodalmaknak megtalálja a magyarországi elérhetőségét, illetve amennyiben van nyílt hozzáférésű elérhetősége, akkor biztosítson közvetlen elérést a cikk teljes szövegéhez.

A COMPASS-ból hiányzó keresett adatbázisok egy része nem releváns adatbázisra történt keresés volt – például Matarka, ODR –, más részük pedig tényleg hiányzott az adott időben az EISZ portfóliójából. Azonban sok közülük azóta már bekerült az elérhető adatbázisok körébe.

2. táblázat: Sikertelen adatbázisra történő keresések megoszlása a kereső kifejezések elemzése alapján⁹, 2017. szeptember – 2019. március

újrakutató keresés	29%
nem értelmezhető kifejezés	0%
elírás	2%
kiadóra keresés	4%
alternatív név	4%
más keresőmezőnek megfelelő kifejezés	10%
hiányzik az adatbázis a COMPASS-ból	17%
tárgyszó	34%

Felhasználói tapasztalatok

A felhasználói tapasztalatok megismerése érdekében a partnerintézmények több mint 40 elektronikusforrás-menedzser könyvtárosával konzultáltunk és kaptunk visszajelzést közvetlenül a COMPASS egy elsődleges használói csoportjától.

Ennek során kiderült, hogy a könyvtári munka mely területein használják a COMPASS-t. A leghangsúlyosabb terület egyöntetűen a könyvtárközi kölcsönzés volt, míg meglepő módon az olvasószolgálat és a gyarapítás kevésbé jelentős felhasználási területek. Ez alapján úgy tűnik, hogy a COMPASS elérte az intézményeket, de elsősorban az elektronikusforrás-menedzserek és a könyvtárközi kölcsönző kollégák ismerik, úgyhogy további időt és energiát kell ráfordítani, hogy a Compass egy alapvető eszköz legyen minden hazai könyvtáros eszköztárában.

A felhasználói élményhez kapcsolódó visszajelzések alapján a kollégák a platformot könnyen kezelhetőnek tartják, továbbá a keresési lehetőségeket és a keresés hatékonyságát is nagyon jónak gondolják. Ez utóbbi ellentmond a korábban látott hatékonysági adatoknak, azonban feltételezhető, hogy a sikertelen keresések nem a gyakorlott könyvtárosokhoz köthetők.

⁹ (az adott keresési hibával lefutott adatbázisra keresések száma) / (összes sikertelen adatbázisra keresés száma)



Abban teljes volt az könyvtárosok egyetértése, hogy a COMPASS-nak ismertebbnek kellene lennie a könyvtáros és a kutatói közösségben egyaránt, amit az EISZ is fontos feladatának tart a jövőre nézve is.

Összegzés

A Magyarországon elérhető elektronikus tudományos információforrások mennyisége átlépte azt a kritikus tömeget, hogy a köztük lévő eligazodáshoz és a hatékony országos szintű hasznosulásukhoz szükségessé vált egy lelőhelyadatbázis létrehozása.

Az elektronikus források konzorciumi beszerzését lebonyolító EISZ Titkárság által 2017-ben továbbfejlesztett nyílt hozzáférésű COMPASS platform lehetővé teszi a források lelőhelyének visszakeresését adatbázis, folyóirat/könyvcím, vagy cikk szinten egyaránt, továbbá nyílt hozzáférésű cikktalálatok esetén a teljes szöveghez is közvetlen hozzáférést biztosít.

Az adatbázis használati adatainak elemzése rámutat arra, hogy a keresőplatform sokkal szélesebb használói kört ér el, mint az adatbázisokra előfizető intézmények kutatói és igazolta a fejlesztés egyik kulcselemének, a címszintű keresésnek a fontosságát. A keresések hatékonyságvizsgálata segített feltárni a még fejlesztendő elemeket, melyeken jelenleg is dolgozunk.

A könyvtáros felhasználói tapasztalatok pozitívak a könnyű kezelhetőség, a funkcionalitás és a hatékonyság szempontjából egyaránt, viszont szükséges a COMPASS ismertségének növelése, hogy egy mindennapos referenzs eszköz legyen a könyvtárosok kezében és egy használt eszköze legyen a kutatóknak.

Bibliográfia

Dér Ádám, Lencsés Ákos. Az EISZ Nemzeti Program szerepe a könyvtárak külföldi szakirodalommal való ellátásában. Tudományos és Műszaki Tájékoztatás 64, 5. sz. (2017): 241–246.

Dér Ádám, Nyiscsák Sándor. A COMPASS adatbázis új verziójának fejlesztése. (Networkshop 2017, Szeged) 2017. április 20. URL: <https://conference.niif.hu/event/7/session/5/contribution/52/material/slides/0.pdf>

Páll Zoltán. COMPASS - Iránytű az információhoz (az MTA KIK EISZ új információszolgáltató adatbázisa). (Networkshop 2015, Eger) 2015. április 1. URL: <https://conference.niif.hu/event/3/session/14/contribution/80/material/slides/1.pdf>

Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. PeerJ. 6:e4375 (2018) <https://doi.org/10.7717/peerj.4375>

E-Learning rendszer és elektronikus tananyagfejlesztés a Könyvtári Intézetben

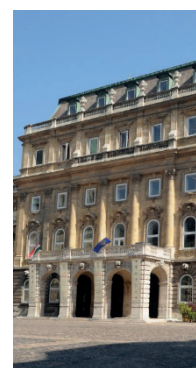
Szepesi Judit
Országos Széchényi Könyvtár, Könyvtári Intézet
szepesi.judit@oszk.hu

We would like to present the plans of an e-learning system for the Hungarian Library Institute. Within the OKR project of the National Széchényi Library we develop also electronic training materials for librarians.

Keywords: e-learning, e-learning system, blended learning, Hungarian Library Institute, professional training.

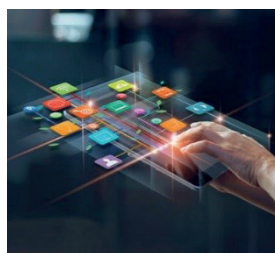
Bevezető

A teljes életpályára kiterjedő képzés igénye és az oktatási médiumok (nyomtatott, elektronikus) változatainak fejlődése révén egyre szélesebb körben alkalmazzák a hagyományos oktatási formák mellett az elektronikus eszközökkel támogatott blended képzéseket. E képzési forma rugalmas módon támogatja a munka melletti tanulást, figyelembe véve a képzésben résztvevők eltérő felkészültségét, tanulási szokásait és időbeosztását.



Az Országos Széchényi Könyvtár, az Országos Könyvtári Rendszer projektjének részeként a Könyvtári Intézet távoktatási rendszerének a megújításán és multimédiás tananyagok fejlesztésén dolgozik.

A 2018-as Workshop-on beszéltünk először a fejlesztési tervekről, jelen összefoglaló célja, hogy az azóta történt munkát bemutassa.



1. E-learning rendszer

Jelenleg az e-learning rendszerünk közbeszerzési folyamata még nem zárult le, ezért a terveket csak röviden – a lényeges elemeket kiemelve – mutatom be.

Tervezési munkánk során meg kellett állapítanunk, hogy az e-learninghez kapcsolódó szóhasználat és a meghatározások ellentmondásokat, illetve néhol átfedéseket mutatnak. Ez főleg abból adódik, hogy magát az e-learning rendszert sokszor azonosítják azzal a szoftverrel, alkalmazással, keretrendszerrel, amely támogatja az online oktatást, a tanulást és az oktatásszervezési feladatokat. Véleményünk szerint a tanulási folyamatokat támogató keretrendszer csak egy eleme lehet az e-learning környezetnek, és mellette más online és/vagy offline alkalmazások is szükségesek.

A projekt keretén belüli tervezés során a tanulási környezet kialakítása mellett további modulokat is tervezünk, amelyek megkönnyítik az oktatásszervezők számára az egységes adatkezelést, átláthatóvá teszik a dokumentációkat.



A tervezés során fontosnak tartottuk a szabványok használatát, ezzel biztosítva a rendszer átjárhatóságát.

Célkitűzésünk

A fejlesztési feladat célja olyan weben elérhető e-learning portál (szolgáltató rendszer) kialakítása, amely támogatja a Könyvtári Intézet továbbképzési tevékenységéből fakadó adminisztrációs és dokumentációs feladatok ellátását. Újdonság lesz, hogy a tervek szerint az Országos Széchényi Könyvtár belső képzéseit is támogatni fogja a portál. A rendszer használatával lehetőséget szeretnénk adni más könyvtáraknak is, hogy – megfelelő felkészítés után – saját szakmai oktatási tevékenységüket ezen keresztül szervezhessék.

Az oktatási tartalmak egy böngészőn keresztül lesznek hozzáférhetők. A tanulási folyamatokhoz web 2.0-ás környezetet hozunk létre, amely alkalmazásaival közvetlenül és közvetetten is támogatja a kapcsolatok kiépítését és a közösségi tanulás folyamatát. A tanulási környezetben mini blogolási lehetőséget kínálunk, google alkalmazásokat építünk be (pl. Google calendar eseményértesítővel).

Az elkészült tananyagokat egy tananyagtárban helyezzük el, ahol gondoskodunk a tananyag-elemek és a tananyagok feldolgozásáról, tárolásáról. Elképzelhetők tartjuk egy nyílt forráskódú dokumentumkezelő vagy tartalomkezelő rendszer beépítését a rendszerbe (pl.: Liferay, Dispace).

A tervezett rendszer elemei

- webes szolgáltató felület

Az e-learning portál egy szolgáltató felület, amely mögött moduláris felépítésű programrendszer támogatja az oktatásszervezési, tanulási, tanulás-irányítási, oktatási feladatok elvégzését, valamint a digitális tananyagtár kezelését.

- oktatásszervezési modul

Háttéradatbázisok és űrlapok segítségével teljes körűen képes ellátni a képzések **(OKJ-s képzés, akkreditált tanfolyamok, nem akkreditált belső képzések)** szervezéséhez szükséges adminisztrációs és dokumentációs feladatokat.

- tanulási és oktatási folyamatokat támogató modul

A tanulásirányítási oldal olyan eszközöket nyújt, amelyek segítségével könnyen megtervezhető és összeállítható egy továbbképzés tartalma, a tanfolyamok lebonyolíthatók, a tanulási folyamatok nyomon követhetők, ellenőrizhetők. Törekszünk arra, hogy a tananyagok mobil eszközökön is elérhetők legyenek.

- digitális tananyagtár

A tananyagtár olyan – az aktuális képzéseket támogató – gyűjtemény, amely biztosítja az interaktív, multimédiás tananyagok, tanulási elemek és objektumok szabványos módon történő feltöltését, feldolgozását (Dublin Core és/vagy LOM alapú leíró adatok hozzárendelését), kereshetőségét és használatát.

2. Multimédiás tananyag



A közbeszerzés során a multimédiás tananyagok kialakítására többen is pályáztak, a nyertes pályázó az OnlearnTech Kft. lett. A fejlesztés során négy tananyag születik meg, amelyek a későbbiekben blended képzésekbe épülnek be. (temakörök: webarchiválás, állományvédelem, digitalizálás, az e-learning módszertana, szerzői jogi tudnivalók az elektronikus tananyag használatáról)

Mielőtt azonban a fejlesztés lépéseit röviden bemutatnám, szeretném fogalommeghatározással kezdeni. Ezen a területen is különböző megvalósítási formákba, típusokba botlottunk, ezért a magunk számára így értelmeztük az elektronikus tananyag fogalmát:

Olyan pedagógiai elvek alapján, az informatika eszközeivel felépített oktatási tartalom, amely önálló, módszertani szempontból zárt egységet alkot, de akár egy vagy több képzéshez is illeszkedhet. Felépítését és alkotóelemeit szabvány rendszerezi. Alapvető célja egy adott kompetencia kialakítása, fejlesztése.

A fejlesztés során az a célunk, hogy weben elérhető, a könyvtári szakmai továbbképzést támogató, szabványos, multimédiás, interaktív tananyagok készüljenek, amelyek internetes böngészőkben, mobil eszközökön és egy nyílt szabványokat és open source megoldásokat támogató rendszerben is lejátszhatók, megtekinthetők.

Általános formai elvárásaink a multimédiás tananyaggal szemben

Egységes, a tartalomhoz illő design, amely felhasználóbarát és könnyen kezelhető
Egységes megjelenés, témakörök szerinti színválasztás. Felhasználóbarát, menürendszere átlátható és könnyen kezelhető, a navigálás egyszerű. A szép design a felhasználó számára eszköz, amely segíti a tartalom megtalálásában (pl. egyértelmű ikonok segítségével).

Szabványos megoldásokat alkalmaz (pl.: HTML5, SCORM)

Egy jól működő elektronikus tananyag alapja az e-learning szabványok használata, ezek **biztosítják** a tananyagok **megtekinthetőségét**, lejátszhatóságát és kereshetőségét a különböző típusú, tanulási folyamatokat menedzselő rendszerekben. Az ily módon elkészített tananyagok elemei, objektumai **újrahasznosíthatók** egy új képzés kialakítása során. A legismertebb e-learning szabvány a SCORM, amelynek több verziója is napvilágot látott már.

Interaktív elemeket tartalmaz

Az interaktív elemek kiegészítik a szöveges tananyagot, segítik és támogatják a tanulás hatékony megvalósulását. Egyszerű, játékos, interaktív megoldás lehet pl. drag and drop tesztek használata. A teljesítmények elismerésének az egyik formája lehet pl. kitűzők (badges) adása.

Reszponzív megjelenítést biztosít

A felhasználói élmény biztosítása, a tananyagok jó olvashatósága, hordozhatósága, lejátszhatósága okos eszközökön (pl.: okostelefon, tablet) elengedhetetlen. A részponzív megjelenítés jellemzője, hogy eszközfüggetlenül, rugalmasan alkalmazkodik a böngészők felületéhez azért, hogy optimális megjelenést adjon. Szabványa a HTML5-t kiegészítő CSS3.

A tananyagok felépítése

A tananyagok felépítése moduláris. Az egy modulba tartozó leckék jellemzője, hogy közöttük tartalmi-logikai kapcsolat van.



A **modulok** elején rövid bevezető található, amely leírja, hogy mi az adott modul célja és tartalma, milyen kompetenciákat fejleszt.

A modulok végén – a lényeges információk kiemelésével – **rövid összefoglalást és önellenőrző kérdéseket**, feladatokat helyeztünk el, amelyek segítségével a képzésben résztvevők ellenőrizhetik a tudásukat az adott modul témájában. A helyes válaszok azonnal megtekinthetők.

A fejlesztés folyamata

A multimédiás tananyag tervezésekor meghatározott célrendszer rögzítése után és arra épülve készítettük el a kéziratokat az adott téma szakértőivel.

A kéziratírás során jött létre az az oktatási tartalom, amellyel a továbbiakban a tananyagfejlesztők dolgoztak. Ez tekintjük a végleges tananyag háttéranyagának, szakmai dokumentációjának. (A teljes kéziratot az egyes moduloknál letölthetővé tettük.)

Ez az oktatási tartalom alakul tovább, ebben a formában még nem rendelkezik az online tananyagok jellemzőivel, lineáris és statikus. Ebben a fázisban már együtt dolgoztunk a szerzőkkel és a fejlesztőkkel. Közös emeltük ki a lényeges gondolatokat, az alapfogalmakat és minden olyan tanulási egységet, amely a tananyag sikeres előállításához és használatához tartozik.

Ebben a fázisban gondoskodtunk azoknak az elemeknek a megtervezéséről is, amelyek az oktatási tartalmat kiegészítik, illusztrálják (pl.: kép, videó, animáció, hanganyag, design elemek).

A következő lépésként elkészültek a forgatókönyvek, amelyek tükrözik a kézirat szerkezeti felépítését, de azon túlmutatnak a didaktikai elemek beépítésével. Elkészítésük során jött létre a végleges tartalom és a médiaelemek listája, egy adott szerzői rendszerre specializálódva. A digitális tananyag összeállítása a forgatókönyvben található szövegek, módszertani elemek és multimédia állományok jelölése alapján, azok pontos meghatározásával történt.

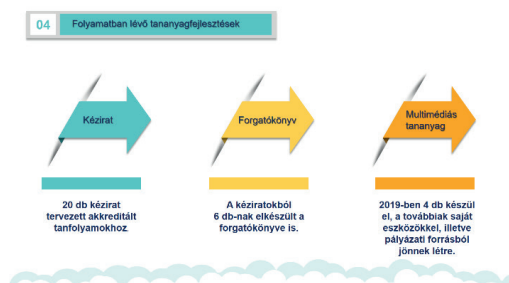
Az elektronikus tananyag tervezési folyamatában a forgatókönyv elkészítéséhez prototípusokat alakítottunk ki a fejlesztőkkel, amelyek támogatják a szöveg egységes megjelenését, stílusát és az oktatási tartalom tagolását.

A fejlesztők az elkészült tananyagokat úgy adják át, hogy lehetőségünk legyen a későbbiek során a tartalmak átdolgozására, frissítésére.

A tananyag tesztelésébe a kéziratok szerzőit is bevontuk, velük együttműködve, véleményük, javaslatuk alapján végezzük el a módosításokat, változtatásokat. A végső változatnak stilisztikai, műfaji és működési szempontból kifogástalannak, hibátlanak kell lennie.

Folyamatban lévő tananyagfejlesztések

A Könyvtári Intézet módszertani vezetésével a szakmai továbbképzésekhez további témákat választottunk ki, amelyekről szakértők bevonásával 20 db kézirat készült el. Multimédiás tananyaggá alakításukhoz jelenleg forrást keresünk.



Képzők képzése

A Könyvtári Intézet oktatóit és a projektben fejlesztés alatt álló tananyagok szerzőit egy rövid képzés során készítettük fel a kéziratok megírására.

A képzés célja, hogy a résztvevők a tananyag elsajátítása során megismerjék az e-learning és az elektronikus tananyag fogalomrendszerét, felépítését, képesek legyenek arra, hogy – saját témakörükhöz – önállóan létrehozzák egy elektronikus tananyag kéziratát, valamint tanácsadóként részt vegyenek az e-learning oktatási tartalom kialakításában.

A projektünk az OKR fejlesztéssel is kapcsolatban van, ezért a későbbiek során az ott jelentkező oktatási igényeket is figyelembe vesszük, illetve a partner könyvtárakban dolgozó kollégákat is igyekszünk felkészíteni a rendszerünk használatára, valamint elektronikus tananyagok készítésére.

A képzések további célja a szemléletváltás támogatása, a szakmai kompetenciák fejlesztése és új ismeretek átadása.

4. A multimédiás tananyagok helye és szerepe a továbbképzésben

Online oktatás

Munkánk során több alkalommal is belefutottunk abba, hogy az online tanulási és oktatási környezet meghatározása mennyire sokszínű. Az online oktatásmódszertan esetében még több változatot és lehetőséget tapasztaltunk, ezek közül pár fogalmat emelnék ki: távoktatás, e-learning, blended learning, online oktatás,



konnektivizmus. A meghatározásokat tovább árnyalhatja az, hogy melyik szakmán belül használják.

Számunkra – meghatározása és felhasználási módja miatt – a mai magyarországi tanulási és oktatási módszerek közül a blended learning látszik legmegfelelőbbnek. A „kevert képzés” olyan oktatásmódszertani megoldás, amely egymásba integrálja a kontakt tevékenységeket az online tevékenységekkel, az irányított tanulást az önszabályozó tanulással, vagyis az online oktatást a jelenléti oktatás kiegészítéseként értelmeztük.

Ennek fényében a kontakt (gyakorlat) tevékenységek részben változatlan formában maradnak (tartalmi és módszertani szempontból), de az órák egy része helyettesíthető online elérhető, multimédiás tananyagokkal, amelyek kiegészítik a jelenléti oktatást és többlet ismeretet adnak át. (információk, tudásanyagok, tesztek, feladatok stb.).

Összefoglalás

A Könyvtári Intézet – azon belül az Oktatási és Humánerőforrás-fejlesztési osztály – olyan oktatást támogató rendszer létrehozását tervezi, amely eszközöket biztosít az oktatásszervezéshez, támogatja az oktatási és tanulási tevékenységet és szabványos környezetet biztosít az elektronikus tananyagok szolgáltatásához, lejátszásához, tárolásához.

Bibliográfia

ANTAL Péter: Interaktív elektronikus tananyagok tervezése. Eger, Eszterházy Károly Főiskola, 2013. 175 p. (Médiainformatikai kiadványok)

Islam, Md. Shiful: How library and information science academic administrators perceive e-learning in LIS schools : a qualitative analysis. In: IFLA journal. – 40. (2014) 4., 254-266 p.

Kane, Kathryn: [Tapping into social networking : collaborating enhances both knowledge management and e-learning](#). In: VINE. – 40. (2010) 1., 62-70p.

Komenczi Bertalan: Az E-learning lehetséges szerepe a magyarországi felnőttképzésben. Bp., Nemzeti Felnőttképzési Intézet, 2006. 60p.

Nagy Vitéz: [„E-learning ABC”](#). In: Vezetéstudomány, 2016. XLVII. Évf. 12. szám, 6-15p.

Noesgaard, S. S. es Orngreen, R. es Foundation, K.: The Effectiveness of E-Learning : An Explorative and Integrative Review of the Definitions, Methodologies and Factors that Promote e-Learning Effectiveness. In: Electronic Journal of e-Learning, 13. volume, 2015. p. 278-290.

OKTATÁSTERVEZÉS, digitális tartalomfejlesztés. Szerk.: Ollé János, Kocsis Ágnes, Molnár Előd. Eger : Líceum K., 2015. 266 p

Besorolási, szabványosított, normatív vagy „autorizált”

Ungváry Rudolf
Országos Széchényi Könyvtár
ungvaryr@gmail.com

In English at the introducing of the MARC the former „heading” name had to be changed because in the description of the resource it denominated the data from the point of view of its place. In Hungarian this change was not necessary because the Hungarian name does not mean heading, it names the data based on its purpose. Disregarding the logic counter-arguments, at present the well accepted former phrase has been replaced – it was introduced the “authorized” (“autorizált”) data phrase. The present study describes the pros and contras in the present debate, points out the misinterpretation of the phrase which has been successfully used until these days, and demonstrates why the suggested phrase is misleading and inaccurate.

Angol nyelven azért kellett a MARC megjelenésével megváltoztatni a korábbi besorolási adatot jelentő „heading” nevet, mert az a besorolási adat elhelyezése szempontjából nevezte meg az adatot a forrás leírásában. A magyarban azért nem kellett ezt megtenni, mert a „besorolási adat” név nem a helye, hanem a rendeltetése szempontjából nevezi meg az adatot. Az ellenérveket semmibe véve most bevezették helyette az „autorizált” adat kifejezést. A tanulmány részletesen ismerteti az érveket, az ellenérveket, és rámutat a „besorolási adat” kifejezés félreértelmezésére, a javasolt kifejezés félrevezető és hibás voltára. A „besorolási” jelző a maga tömörségével és szépségével nyelvi telitalálat volt.

Egy terminológiai kérdés háttere

Azért kellett angolul és németül a MARC előtti időkben kialakult „heading” és „Ansetzungsform” kifejezéseket megváltoztatni, mert ezek a helye szempontjából nevezték meg a tételt. Annak a helynek a szempontjából, amelyet a mindenkor dokumentumleírásban elfoglalt. A MARC-rekordban magát a „heading” és németül az „Ansetzungsform” kifejezést azonban továbbra is használták a rendszói részre és a közvetlenül hozzá kapcsolódó kiegészítő elemekre, és ez a magyarban továbbra is ugyanígy van. De maga az egész MARC-rekord nem csak a tételfejből áll (lásd a 4. ábrát).

Ezzel szemben a „besorolási adat” kifejezés nem tételfej jelent, holott a manuális gyakorlat idején valóban csak a tételfej képviselte ezt az adatot. Amikor tehát megjelent angolul az új „authority record” kifejezés, magyarul nem volt szükséges megváltoztatni a „besorolási adat” nevet. Már csak azért se, mert maga az MSZ 3440-es besorolási szabvány megszületésekor, 1979-ben (lásd a 3. ábrát) már létezett mint a Magyar Nemzeti Bibliográfia számítógéppel kezelt állománya. Ennek rekordjai lényegében a MARC-nak feleltek meg, még ha annak a magyar könyvtári szabványokhoz alkalmazott bibliográfiai MAMARC volt is. A HUNMARC, és annak besorolási formátuma 1988-ban jelent meg [5, 6, 7]. Hosszú ideig a „besorolási adat”,



„besorolási rekord” kifejezés semmiféle értelmezési problémát nem okozott. 1998-ban, tehát amikor a MARC már régen létezett, és a magyar HUNMARC besorolási adatsere-formátum is készen volt, a „besorolási” jelző magyar használatát a MARC-rekordra senki se kifogásolta, holott angolul már évtizedek óta létezett az „authority” név.¹

Besorolási adat helyett „autorizált”. Érvek

Körülbelül két-három év óta Magyarországon az „autorizált formátum”, „autorizált adat” és az „autorizálás” kifejezés használatára akarnak áttérni.²

Pártolói azt állítják, hogy a „besorolási adat”, „besorolási rekord” elavult, mert

- (1) ma már nincs szó besorolásról;
- (2) csak a könyvtári besorolási szereppel rendelkező adatokat tartalmazza, nem tartalmazza a tartalmi feltárással kapcsolatos újabb metaadatokat és relációkat, az adatkezeléssel kapcsolatos metaadatokat
- (2.1) a MARC21 „authority” rekord sokkal több adatelemet tartalmaz, mint amennyit a korai ISBD és MSZ szabványokban rögzítettek;
- (2.2) a MARC21 „authority” nemcsak több, de komplexebb, sokféleséget kifejező adatok értendők, és az „autorizáláson” komplexebb tevékenység értendő, mint a besorolási adatok szabványosított formáinak kialakítása;
- (2.3) mivel a digitális információs térben sokféle forrásból kell összetereelni és összehangolni a metaadatokat, ezért összetettebb kognitív folyamatok zajlanak le;
- (2.4) nem alkalmasak arra, hogy tágabb kontextusban is helyt álljanak. Nem beszélve olyan gyökeres ellentmondásokról, hogy az egységesített besorolási adatok nem egységesek [8],
- (2.4) „nem egyszerűsíthetjük le a kérdéskört az egységesítés és a besorolás nézőpontjaira [...] a reprezentálás elemibb adatok szintjén történik, a „besorolás” tényleg értelmezhetetlen ebben a közegben” [8];
- (2.5) „Ma már egyenrangúak a névvariánsok, amelyek közül kiválasztható a kitüntetett névalak, a többi kezelése pedig az authority rekord feladata.” [6]
- (3) Az FRBR, FRAD, RAK a források kezelésének többszintű modelljén alapszik, ezen belül a megjelenési forma csak az egyik szint. [6]

1 A „besorolási adat” magyar kifejezés tehát 1998-ban is elfogadott volt, nem Vajda Erik találta ki egyedül, csak abban az évben regisztrálta egy elemzésében a név magyar használatának tényét a „besorolási rekord” esetére is. A HUNMARC esetében ugyanis azt már eleve használták azok, akik a változatait kidolgozták (és ezek a szakemberek, mint Sipos Márta, Szűcs Jenőné nem azok közé tartoztak, akik az MSZ 3440 magyar szabványt 1976-ban létrehozták, hanem azok közé, akik a tényleges könyvtárinformatikai gyakorlatban ezt a rekordtípust tevőlegesen a Magyar Nemzeti Bibliográfiához előállították. Nekik semmiféle problémájuk nem volt a kifejezéssel – pedig ekkor már nem csak a tételfejtet kellett meghatározni folyamatosan, hanem magát a rekordot is.

2 Az ellenérveket kifejtette Dudás Anikó [1, 2, 3] és a Katalisten Tóvári Judit [8].

- (4) a besorolási adat csak a hagyományos „heading element”-et, azaz a korai ISBD és MSZ szabványokban rögzített tételfejtet jelenti, nem pedig a teljes MARC21-rekordot.

Ellenérvek

Egyik indoklás sem igaz.

- (1.1) A fentiekben már kimutattuk, hogy **a besorolás nem csak azt jelenti**, hogy valami konkrétat (esetünkben egy katalóguskártyát) egy rendezési elv alapján sorrendbe állítanak, elhelyeznek egy konkrét sorban (egymás mögé vagy elé, azaz sorrendben). Van egy általánosabb jelentése, amely szerint valamit egy adott fogalom terjedelmébe sorolnak be. Márpedig szaktárgyszavakkal (ahogy az osztályozási rendszerek jelzeteivel is) a forrásokat az általuk megnevezett konkrét dologhoz kapcsolják vagy az adott fogalom terjedelmébe sorolják. Ehhez nem szükséges semmiféle cédulakatalógus.

A nyelvhasználat is ezt bizonyítja:

„Hogyan érhetem el, hogy a Google minél egyszerűbben besorolja az oldalamat a keresőbe?”

- (1.2) Attól kezdve, hogy a MARC-formátum Magyarországon ismertté vált, az Országos Széchényi Könyvtár (OSZK) által bevezetett és használt besorolási rekord és adat minden félreértést kizárva használható volt a szakmában. 1970 óta, mióta MARC-formátumot használnak Magyarországon is (kezdetben HUNMARC formájában), senki sem emelt ellene kifogást, és nem számított félreérthetőnek. Ha egy szabványban rögzítik, hogy az „authority data” besorolási adatot jelent, akkor teljesen egyértelmű a szakmai jelentése.

Ráadásul az MSZ 3440 besorolási szabványt csak 1979-ben adták ki. Kialakításakor már egyrészt régen ismert volt a USMARC, másrészt már a hetvenes évektől kezdve használták a MAMARC-ot is, majd a HUNMARC-ot a Magyar Nemzeti Bibliográfia számítógéppel történő előállítására. Magában a szabványban egyetlen egyszer nem fordul elő a katalóguscédula szó, vagy az, hogy besorolási adat alapján cédulák közé sorolják a katalógustételt. Azaz a szabvány készítői – akik közé a számítógéppel készített Magyar Nemzeti Bibliográfia szerkesztője is tartozott – úgy hozták meg a döntésüket a szabványos „besorolási adat” megnevezéséről, hogy egyáltalán nem hagyták figyelmen kívül azt a tényt, hogy a gépesítés körülményei között már szó sincs cédulák közé sorolásról.

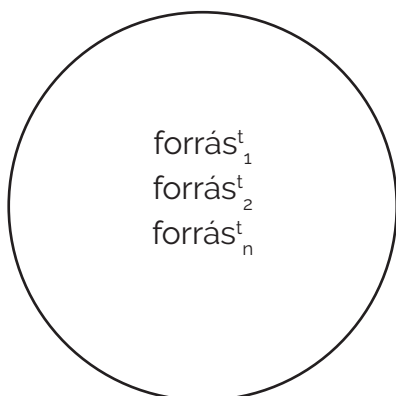
- (1.3) A besorolási adat, vele a besorolási rekord kifejezés **egyáltalán nem akadályoz** annak, hogy azt ne csak az ISBD és MSZ szabványokban leírt formára, hanem az újabb metaadatokat és relációkat, valamint az adatkezeléssel kapcsolatos többi metaadatot tartalmazó teljes MARC21-rekordra alkalmazzák.

Ezen se a FRAD, se az RDA nem fog változtatni, mivel csak az adattartalmat strukturálják.

Az 1. ábrán modell szinten látható a rendeltetés szempontjából a MARC előtti és utáni helyzet.

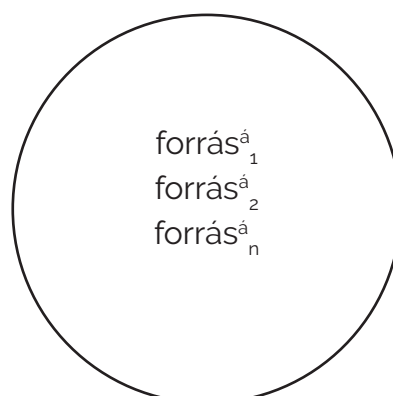
a MARC előtt:

tulajdonnév esetében a
eredményhalmaza:



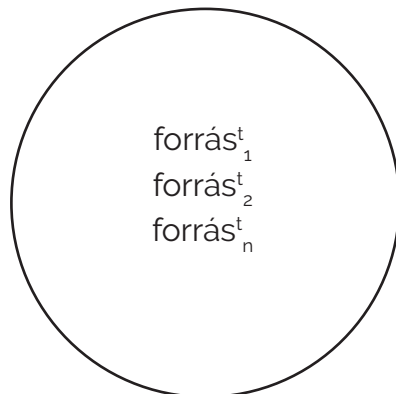
általános név (szaktárgyszó)

esetében a besorolási adat
eredményhalmaza:



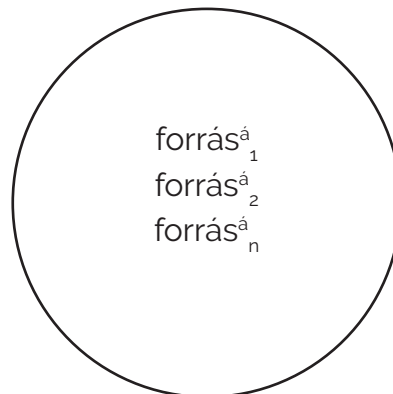
a MARC után:

tulajdonnév esetében a
eredményhalmaza:



általános név (szaktárgyszó)

esetében a besorolási adat
eredményhalmaza:



1. ábra Az „authority data”, a „Normdaten” és a „besorolási adat” rendeltetési mintázata. Mindegyik esetben azonos forráshalmaz keletkezik!

Ez nem a programozás és a gépi tárolás szintje, hanem a koncepcionális modell szintje; továbbá nem a nevekről önmagukban van szó, tehát szótárakról, névterekről), hanem a logikai szinten létező MARC-rekordról

- (2.1) **A MARC21 „authority” rekord valóban sokkal több adatelemet tartalmaz,** mint amennyit a korai ISBD és MSZ szabványokban rögzítettek – csak éppen *ettől semmit se változik a rekord szerepe*. Ugyanúgy arról van szó, hogy a MARC-rekord és az ISBD és MSZ szabványos besorolási tétel lényege, hogy a rekord/tétel tételfejével indexeljenek, „tárgyszavazzanak”, osztályozzanak forrásokat úgy, hogy azok a mindenkori állományban hozzáférési pontok legyenek. Az, hogy a MAR21 esetében a tételfejen (annak rendszaván, kiegészítő adatain) kívül még számos további, a tételfejre vonatkozó metaadat is van, a tényen semmit sem változtat.
- (2.2) A többletadatok valóban sokfélék. Megállapításuk valóban több munkával jár. Ez azonban egyáltalán nem változtat azon a tényen, hogy a rekord ezen adatelemei egyrészt a *besoroláshoz szükséges metaadatai*, másrészt pedig a besorolás eredményeként *a hozzáférési pontot alkotó tételfej metaadatai* („visszajelző referenciális” adatai). Márpedig miért ne lehetnének ilyen adatok egy olyan rekordban, melyet magyarul besorolásinak nevezünk? E többlet metaadatok léte *mit változtat azon*, hogy a rekord továbbra is hozzáférési pontként szolgál, azaz vagy egy megnevezett fogalom terjedelmébe sorolja a forrást, vagy egy tulajdonnévhez kapcsolva sorolja azt be?
- Az „egységesített forma megszerkesztését, összevetését stb.” pedig a besorolási adatok megállapításakor ugyanúgy el kellett végezni, ezen az, hogy a MARC21-rekordban több az adatelem, semmit sem változtat, hiszen a kevesebb vagy a több adatelem megállapítása egyaránt „komplex tevékenység”.
- (2.3) Az „információk összeterelése”, a „nem azonosakra vonatkozó megkülönböztetése” egyáltalán nem változás a MARC előtti időkhöz képest. Ez semmi mást nem jelent, mint amit a (2.2) pontban leírtunk: hogy a besorolási/„autorizált” adattal vagy meghatározott névhez kapcsolják, vagy meghatározott, szakkifejezéssel megnevezett fogalom terjedelmébe sorolják a forrást. A „névterek integrált, szemantikus, többcélú kezelését” pedig *egyáltalán nem zárja ki, hogy az adattal besorolnak*, hiszen éppen az ilyen, besorolt, tehát tartalmilag egyértelműen azonosított adattal lehetséges elvégezni ezeket a kezeléseket. A „kognitív folyamatok” a hagyományos, ISBD szerinti besorolási adatok használata és a velük végzett keresések alkalmával is ugyanúgy lejátszódnak, mint a mostani, gépesített körülmények között (hiszen ezek nem a külvilágban, nem a technikában, hanem az elmében zajlanak).
- (2.4) Attól, hogy a szabványosságot (mert az egységesítés ezt fejezi ki) különféleképpen értelmezik, és ezért még a szabványos MARC21-rekordok se teljesen egyformák ugyanazon entitás esetében, nem változik az a tény, hogy a forrásokat ezekkel a MARC21-rekordokkal *besorolják*. Hogy a besoroláshoz használt rekordban több vagy kevesebb „elemibb” adat (értsd adatelem) van, attól még az állítás igazságtartalma változatlan marad.



- (2.5) *Tóvári Juditnak az az ellenérve, hogy „ma már egyenrangúak a névvariánsok”, azt jelenti, hogy a névvariánsok (tehát a „lásd/helyett” relációkkal kapcsolódó gépi rekordok) mindegyike – helyesen, a számítástechnika jóvoltából – hozzáférési pont lehet. Azaz mindegyik szerint besorolva van a forrás adott fogalom terjedelmébe, illetve adott névhez. Akkor meg miért ne lenne használható rájuk a besorolási adat?*³
- (3) Attól, hogy az FRBR, FRAD és RAK modellje megkülönbözteti a mű, a kifejezési forma, a megjelenési forma és a példány szintjét, abból a szempontból, hogy mi a MARC21-rekord feladata, semmi se változik: ezeken a szinteken *egyaránt be kell besorolni a forrást*.
- (4) **Az angol „heading” (tételfej)** valóban csak az ISBD és MSZ szabványokban rögzített tételfejét jelenti, nem pedig a teljes MARC21-rekordot. Éppen ezért kellett angol nyelvterületen a MARC-rekordok megjelenésével egy új kifejezést bevezetni. Az *új kifejezés* („authority”) azonban egyáltalán nem utal arra, hogy a korábbi „heading”-nél nagyobb adatról van szó. A *feladata kizárólag az, hogy kizárja a MARC-rekord azonosítását a tételfejjel (a „heading”-gel)*. Csak éppen magyar nyelven nem kell a szakkifejezést megváltoztatni a MARC-rekord esetén, mert a besorolási rekord eleve nem jelenti, hogy csak tételfejről van szó.

A MARC21 besorolási rekord tételfejen kívüli többi adatelemét azért nem kell a tételfejjel (vagy annak akár csak egyetlen elemével is) együtt megjeleníteni hozzáférési pontként, mert azok a tételfej másodlagos avagy metaadatai. Ugyanakkor gépi rendszerben – ha olyan az adatbázis-kezelő rendszer – a besorolási rekord tetszőleges adateleme kereshetővé válik, azaz mindegyik besorolási adat lehet.

A felsorolt ellenérvekben nem általában „a magyar nyelv védelméről” van szó, hanem egy javasolt szakkifejezés történeti és szemantikai elemzéséről. Elsősorban szakmailag érveltem amellett, hogy ez a névváltoztatás téves értelmezés eredménye.

A szaknyelv nyilvánvalóan más, mint a beszélt magyar nyelv, nem az a probléma, hogy egy kifejezés természetes nyelvű-e vagy sem. A besorolási adat is szaknyelvi kifejezés.

Az elemzéssel a döntés eredményére, azaz az „autorizált” kifejezés félrevezető voltára kívánok rámutatni szakszerű érvek alapján! Arra, hogy teljesen más okból választottak az angol-amerikai szakterületen a „heading” helyett „authority data” és „authority record” nevet, mint amit a magyar névváltoztatás javaslói állítanak. Azért

3 Azt a tényt, hogy „ma már egyenrangúak a névvariánsok” már 1998-ban Vajda Erik is megállapította. „Akkor” még senki se kifogásolta, hogy ezért helytelen a besorolási rekord kifejezése, és nem javasolta, hogy a besorolási rekord helyett az „autorizált rekord” kifejezést kellene használni! Ha „akkor” mindenkinek egyértelmű volt ez tény, akkor „ma”, 21 évvel később (tehát több mint két évtized múlva) Tóvári Judit állítása miért érvényesebb, mint „akkor”?

választották, mert a hagyományos angol ISBD-kifejezés (a „heading”) nem volt alkalmas az új adat és rekord nevének azonosítására, és nem önmagában azért, mert a MARC-rekordban sokkal több adat jelent meg, mint az ISBD-tételben, és mert összetettebbé vált a „szétszórt adatok összeterelése”. A magyar „besorolási adat” és „besorolási rekord” név ezzel szemben nem alkalmatlan arra, hogy a megnövekedett adattartalmú tételt azonosítsa – *mert nem pusztán tételfejt jelent* – és a használatát illetően valóban továbbra is besorolásról van szó. Az elfogadott „autorizált” jelző szolgálai átvétele az angoltól szakszerűtlen, és kritikájának jelenlegi hiánya fényt vet a könyvtári szakma jelenlegi helyzetére.

Az „autorizáció” egy másik szaknyelvben – bankszerktörében is megjelenik, ott hitelesítést, feljogosítást, felhatalmazást, engedélyezést jelent. Szakkifejezésként ma az autorizációt csak az automatán végzett kártyaművelet esetében használják, és ez megfelel a „hitelesítési” jelentésének. Azt értik rajta, hogy az elfogadó (az automata) hitelesként ismeri fel a bankkártyát és engedélyezi a további műveleteket. [4] Márpedig a MARC21 tételfejei és a hozzájuk kapcsolódó metaadatok, akárcsak az ISBD „heading”-je és „heading element”-jei *nem felhatalmazás, engedélyezés*, hanem szabványosítás nyomán keletkező adatok. Egyszerűen azért, mert itt – ahogy a szabványos tárgyszavak esetén is – szabványosításról, és nem engedélyezettségről van szó.

A „besorolási” helyett az „autorizált” névváltozta következtében egy idő múlva a szakma új nemzedékei számára érthetlenné válik a korábbi szakirodalomban használt, „besorolási” jelzővel ellátott összes szakkifejezés.

Németül se az autorisierte Date kifejezést használják! A német gyakorlathoz hasonlóan *legfeljebb átkeresztelhető szabványos névadatra, szabványos névrekordra*:

authority data	Normdaten	szabványos névadat
authority record	Normsatz	szabványos névrekord
authority file	Normdatei	szabványos névállomány
authorized	normiert	szabványos

A változtatás így is azzal járna, hogy egy idő múlva a szakma új nemzedékei számára érthetlenné válik a korábbi szakirodalomban használt, „besorolási” jelzővel ellátott összes szakkifejezés. A „besorolási” jelző a maga tömörségével és szépségével nyelvi telitalálat volt. Hozzáértő szakemberek hozták létre, évtizedekig nem okozott problémát, és eleve annak figyelembe vételével született, hogy a MARC (és később annak magyar megfelelője) létezett.

Epilógus

Nem remélhetem, hogy ezzel az elemzéssel meg tudom változtatni a „besorolási” jelző megszüntetéséről szóló végzetesen hibás döntést. Ennek a tanulmánynak a célja, hogy felsorolja az érveket az „autorizált” kifejezés ellen, és ezt párhuzamba állítsa azzal, hogy mekkora félreértés ez névváltoztatás.



Amikor a múlt évben vita kezdődött a Katalisten a besorolási adat kérdéséről, az érveimre – melyek lényegében megegyeztek azokkal, melyeket ebben az értekezésben felsoroltam – nem érkezett viszontválasz.

Azon a nemrég tartott megbeszélésen, melyen eldöntötték, hogy az „autorizált” kifejezést fogják használni, nem vehettem részt, érveimet nem adhattam elő.

Irodalom

- [1] Dudás Anikó. Az autorizálás az információszerzés és –hozzáférés feladatkörében – definiálás és nemzetközi tapasztalatok. = Könyvtári Figyelő, 52. évf. 4. sz. (2006). <http://www.ki.oszk.hu/kf/kfarchiv/2006/4/dudas.html#top>
- [2] Dudás Anikó: Nevek, antik nevek – autorizálás: egy magyarországi felmérés eredményei. = Tudományos és Műszaki Tájékoztatás, 52. évf. (2005) 9. sz.
- [3] Dudás Anikó: Forrásleírás és hozzáférés. Az új angol-amerikai katalogizálási szabályzat (RDA) és kritikája. = Könyvtári Figyelő, 2013. január 3. 4. sz. <http://ki2.oszk.hu/kf/2013/01/forrasleiras-es-hozzaferes-az-uj-angol-amerikai-katalogizalasi-szabalyzat-rda-es-kritikaja/>
- [4] Gál Erzsébet: Hitelkérelmek, banki ismeretek. 4.3.3. Autorizáció (felhatalmazás) folyamata. (2013) https://www.tankonyvtar.hu/en/tartalom/tamop412A/0007_d3_hitelkerelem_jav_scorm/4_3_3_autorizacio_felhatalmazas_folyamata_knqQLg6m4m26OXpo.html
- [5] HUNMARC, a besorolási rekordok adatcsere-formátuma. Összeáll. Sipos Márta, Szabó Julianna, Ungváry Rudolf. Budapest, 2010. Tervezet.
- [6] HUNMARC, a besorolási rekordok adatcsere-formátuma. Összeáll. Sipos Márta. Budapest, 1998 március. Tervezet.
- [7] KSZ/4.1 HUNMARC. A bibliográfiai rekordok adatcsere formátuma. [Összeáll. Sipos Márta]. Budapest, 2002. március. http://www.ki.oszk.hu/107/e107_files/public/hunmarc.pdf
- [8] Katalist. [Vita az „autorizált” jelző alkalmazásáról] [Dancs Szabolcs és Tóvári Judit hozzászólása]. 2018. augusztus 24. http://katalist-to-feed.blogspot.com/2018/08/re-katalist-uj-rendszerhez-uj-zsargont_30.html
<https://listserv.niif.hu/pipermail/katalist/2018-August/038933.html>
<https://listserv.niif.hu/pipermail/katalist/2018-August/038886.html>

Köznevelés, felsőoktatás, elektronikus tanulási környezetek

Digitalizáció és sport: az IKT kihívásai a testnevelő képzésben

Antal Péter

Eszterházy Károly Egyetem, Médiainformatika Tanszék

antal.peter@uni-eszterhazy.hu

Digitalization and Sports: ICT-related challenges in Physical Education teacher training

The principal challenge of the 21st century is the digitalization process requiring the educational sphere to meet the demands of the new instructional environments. While digital devices have been virtually fully integrated in the instruction of natural science or humanities subjects, different expectations apply to Physical Education teachers as digital boards do not play a significant role in such instruction programs. At the same time the increasing prevalence of digital and communication technology both in competitive and recreational or amateur sports calls for the instruction of new competences in Physical Education training programs.

Suffice to mention portable analytical devices assessing and analyzing the actual state and the changes of athletes' physiological conditions in real time and in real contexts. Furthermore, tactical analysis, efficiency surveillance, or movement analysis programs along with devices monitoring performance and physiological conditions pervading non-professional sport as well can provide crucial help to athletes and coaches. Since students enrolled in Physical Education teacher training programs are expected to have different ICT skills from their other colleagues in the profession special training schemes are required for them.

My presentation at the juncture of sports and ICT introduces the information technology aspects of the Physical Education teacher training programs at Eszterházy Károly University and compares the results of a survey performed among Physical Education teachers with that of the international trends.

Keywords: ICT, Digitalization, Sport, Performance, Competences.

Bevezetés

A 21. század a digitalizációról szól, melynek fontos kihívása, hogy az oktatás is megfeleljen az új oktatási környezetnek. Míg a természettudományos vagy humán tantárgyak esetében egyre természetesebb a digitális eszközök használata addig a testnevelés tárgy oktatásában teljesen mások az elvárások, hiszen a digitális tábla, mint eszköz, nem játszik fontos szerepet egy testnevelő repertoárjában. Ezzel szemben a digitális- és kommunikáciotechnológia hihetetlen mértékben beszivárgott a verseny- és amatőr sportba egyaránt és új kompetenciák igényét generálja a testnevelők felkészítésében.

Gondoljunk csak azokra a hordozható analitikai eszközökre, amelyek, a sportolók fiziológiai állapotának, változásait valós időben, valós körülmények között képesek mérni és analizálni. Ugyanilyen fontosak a taktikai elemzéseket végző,

hatékonyságot figyelő, vagy mozgáselemző programok, vagy amatőr sportban is használt teljesítményt és fiziológiai állapotot monitorhozó eszközök, amelyek segíthetik a sportolók és edzők munkáját. Éppen ezért fontos a testnevelő képzésben résztvevő hallgatók, speciális felkészítése, hiszen teljesen más IKT eszközökkel kerülnek szembe, mint a közoktatás egyéb szereplői.

Előadásomban a testnevelő szakos hallgatók információtechnológiai felkészítését és képzési rendszerét mutatom be, az Eszterházy Károly Egyetemen, illetve egy testnevelők körében végzett felmérés eredményeit hasonlítom össze a nemzetközi trendekkel a sport és IKT témakörében.

A globális sporttechnológiai piac alakulása a napjainkban

A Transparency Market Research által közzétett új piaci jelentés szerint 2016-ban [6] a globális sporttechnológiai piacot 21.609 millió dollárra értékelték. Ez az érték a becslések szerint, 2019 és 2027 között 14,7% -os éves növekedési rátával számolva, eléri a 93.825 millió dollárt 2027-re. Ez elsősorban a technológiai fejlesztésektől és a sporttechnológiai háttér informatikai támogatásának a növekedésétől függött, főleg a kiemelt sportok körében.

A legkeresettebbek, az analitikai megoldások, a hordozható eszközök, a csapatmenedzsmenttámogató alkalmazások, az egyéni sportolói teljesítményeket és a csapatteljesítmény adatait feldolgozó szoftverek és szolgáltatások.



1. ábra: A sporttechnológiai piac összetevői



A sporttechnológiai piac összetevőit többféle szempontból elemezhetjük.

- Komponens szerint
 - Viselhető eszközök és sporteszközök
- Szoftver
 - Felhőalapú alkalmazások
 - A helyszíni adatfelvétel
- Szolgáltatások
 - Képzés és megvalósítás
 - Támogatás és karbantartás
 - Irányított / kiszervezett szolgáltatások
- Alkalmazás szerint
 - Aktivitás követés
 - Analitika és statisztika
 - Taktika és szimuláció
 - Döntéshozatali támogatás
 - Edzésterv
 - Játék teljesítmény elemzése
 - Csapatelemzés és -kezelés
 - Sérülés és egészségügyi elemzés

A fent látható portfólióból kiderül, hogy a sport és a technológia összefonódása egyre intenzívebb folyamat lesz, gyakorlatilag a jövő egyik leghasznosabb teljesítménynövelő eszközévé maga a technológia válik.

Ezen belül a viselhető eszközök jelentik majd a legnagyobb piacot, hiszen ezek közvetlenül tudnak majd adatokat szolgáltatni az elemzőknek. Az Ericsson cég 2016-os kutatásai szerint [5], 2020-után éri el a technológia azt a szintet, hogy komoly orvosi diagnosztikára is alkalmasak lesznek, akár a mobiltelefonokat is helyettesíteni tudják. Megjelenik a viselhető eszközök internete, ami segít a felhasználóknak interakcióba kerülni más fizikai tárgyakkal, ami nem feltétlenül eszközöket jelent. A felhasználók 60 százaléka gondolja úgy, hogy a következő öt év során a tabletták és a bőr alá ültetett chippek el fognak terjedni - és ezek nemcsak az egészségügyi adatok nyomon követésére lesznek használhatók, hanem ajtók nyitására, tranzakciók és személyazonosság hitelesítéséhez, és tárgyak vezérléséhez. Már ma is az okostelefon-használók 25 százaléka használja az okostelefonját arra, hogy távolról vezéreljen digitális eszközöket otthonában, és 30 százalékuk használ hangalapú keresést az okosóráján.



2. ábra: A viselhető jövő trendjei [5]

A technológia szerepe a testnevelés oktatásában

A technológia jelenlétét és szükségességét a sportban ma már nem kérdőjelezhetjük meg és a bevezető részből is kiderül, hogy a profi sport már kihasználja ezeket a lehetőségeket. A testnevelő képzésben viszont komoly elmaradások vannak a viselhető eszközök iskolai alkalmazásának módszertanában [3].

Még Európában is ritkák azok a kezdeményezések, ahol a közoktatásban használnák ki a viselhető eszközök lehetőségeit. Egyik ilyen példa a spanyolországi Barcelona-Párizs virtuális futás tapasztalatai [2], illetve hasonló még az olaszországi „The Portable Devices as Means to Promote Children's Active” kísérlet, amely a gyerekek iskoláig történő gyaloglását próbálta meg segíteni és inspirálni mobil alkalmazás segítségével. [1]

Magyarországon az utóbbi években kiemelt szerepet kapott az iskolai tevékenységekben a testmozgás, hiszen a 2011/12-es tanévtől, heti 5 órában, kötelező a mindennapos testnevelés a hazai közoktatási intézményekben, melynek eredményessége fontos a felnövekvő nemzedékek szempontjából. A Magyar Diáksport Szövetség 2013-ban indította útnak a „A testnevelés új stratégiájának és a fizikai állapot új mérési rendszerének kialakítása és az önkéntes részvétel ösztönzése a komplex iskolai testmozgásprogramok szervezésében” –T.E.S.I. elnevezésű kiemelt projektjét.



A projekt célja volt, hogy létrehozza a magyar iskolarendszerben egységesen működtethető, egészségközpontú fittségmérési-értékelési rendszert, a NETFIT-et. A NETFIT szoftveres alkalmazása olyan személyre szabott értékelési lehetőséget kínál minden tanuló számára, amely vizuálisan is szemlélteti a fittségi állapotot, valamint konkrét ajánlásokat fogalmaz meg a fejlesztés érdekében. A NETFIT rendszert közel 3700 iskola, 800 ezer tanuló és 13 ezer pedagógus használja. A rendszer hatékony használatához elengedhetetlen a testnevelők részéről a nyitottság a digitális kompetenciák megszerzésére. A testnevelő képzésben viszont nem történik meg a felkészítés a rendszer használatára [4].

IKT innováció a testnevelő képzésben

A fenti trendekből jól látszik, hogy egy nagyon dinamikusan fejlődő informatikai szegmenssel van dolgunk, amit a képzési struktúra sem hagyhat figyelmen kívül.

Az Eszterházy Károly Egyetemen korábban is volt a testnevelők részére a technológiát oktató kurzus, Oktatás- és kommunikációtechnológia néven, de ez elsősorban a tanári pályán általánosan előforduló IKT eszközök használatára és módszertani lehetőségeire korlátozódott.

A 2017/18-as tanévtől azonban kísérleti jelleggel elindítottunk egy speciális kurzust új tartalommal a testnevelők részére IKT innovációk néven.

A kurzus célja, hogy megismertesse a hallgatókat a 21. század IKT eszközeinek hatékony alkalmazásához szükséges ismeretekkel, különös tekintettel azok sport és rekreációs tevékenységek során történő alkalmazására. Fontos az új sporttechnológiák és módszerek megismertetése a hallgatókkal, illetve ezek alkalmazási módszertanának kidolgozása a szakiránynak megfelelően.

A kurzus két kredit értékű, heti két órás gyakorlatot takar egy féléven keresztül.

A legfontosabb témakörök a következők:

- Az információs és kommunikációs technológiák (IKT) hatása a 21. században
- Az IKT fogalomrendszere, az információs- és hálózattársadalom alapjai
- A 21. század kommunikációs és oktatási formái: az elektronikus tanulási környezet fogalomrendszere.
- A technológiák alkalmazási lehetőségei az amatőr és versenysportban (sportanalitikai, teljesítményelemző rendszerek)
- Viselhető technológiák, kreatív médiatechnológiák (okoseszközök használata a sport és a rekreációs tevékenységek során).
- Az iskolai alkalmazás eszközei és szoftverei, adatmegosztás a felhőben
- A hordozható eszközökre épülő ismeretszámítás, mobil kommunikációs eszközök és a tartalomipar, és az iskolai teljesítménymérés lehetőségei
- Szemléltetés, szemléletesség. Az elektronikus publikáció sajátosságai, a prezentálás alapjai.
- On-line prezentációkészítés- Prezi a gyakorlatban
- A digitális technológiára épülő kutatástervezés alapjai.

A tantárgy relevanciájának vizsgálata

A tantárgy tartalmi relevanciájának és a hallgatói vélemények mérésének érdekében egy kérdőíves felmérést és mini interjúk készítését végeztem el a nappalis és levelező hallgatók körében, összesen 38 fő válaszolt.

A kérdőíves kérdések a hallgatók technológiai tájékozottságára, az eszközhasználatra, illetve a tantárgyi relevanciára vonatkoztak.


A miniinterjúk során elsősorban arra a kérdésre voltam kíváncsi milyen ismeretekkel rendelkeznek azok a hallgatók, az információs társadalomról és annak hatásairól, akik a koruk alapján már a digitális bennszülöttek táborába tartoznak.

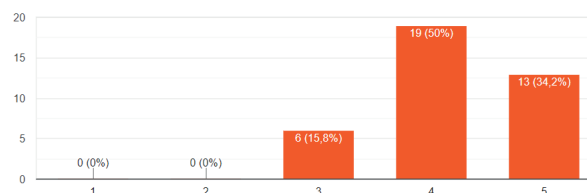
A beszélgetés témakörei a következők voltak:

1. Ön szerint milyen hatással van a technológiai fejlődés a korosztálya életére (pro- és kontra).
2. Hogyan képzei el a modern technológiák felhasználását a saját sportágában?
4. Milyen hatással van média a sportra, és a sport a médiára, milyen előnyök, hátrányok származnak ebből?
6. Hogyan képzei el a jövő IKT eszközeit? Mit gondol milyen új eszközök fognak megjelenni?
8. Hogyan segítheti a sportot a technológia, milyen eszközöket látnának szívesen, illetve hogyan tudnának ezek segíteni a sportban?
10. Hogyan ítélik meg, hogy általános- és középiskolában illetve a felsőoktatásban felkészítették-e Önöket az információs társadalom kihívásaira és megfelelő szintűnek tartják-e ezt?

Jelen cikkben csak a legérdekesebb kérdéseket elemzem a következőkben.

A kérdőív eredményei

Mennyire tartja fontosnak a fejlett információtechnológia megjelenését a sportban? 
38 válasz

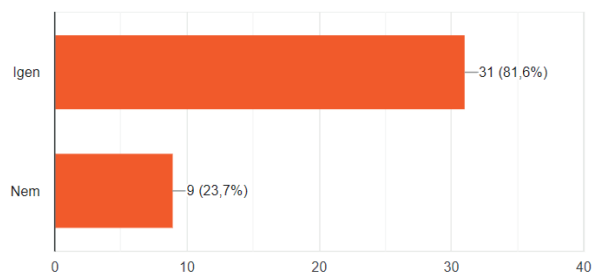


A kérdésre mindenki pozitív választ adott, gyakorlatilag a hallgatók 80%-a érzékeli a modern technológiában rejlő lehetőségeket, hogy segítséget nyújthat a sportban. Amint az eredményből látszik, még mindig van 20% akik elutasítják a technológiai vívmányok használatát a sportban. Ennek oka többféle lehet, az elutasítók többsége olyan eszközökre gondolt, amelyek illegálisan segítik a sportolók eredményességét, illetve nem élsportolók és nem látták bizonyítottan a technológia szükségességét a sportágukban.



Használ(t)-e valamilyen okos eszközt sporttevékenység közben?

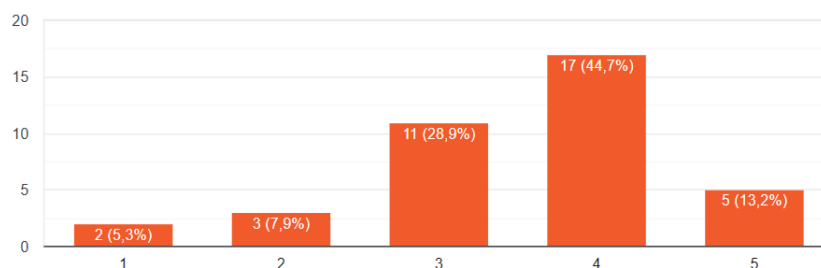
38 válasz



Az eszközhasználatban is pozitív a hallgatók véleménye, sokan használtak vagy használnak viselhető eszközöket. Legtöbbjük okosórát, testszenzort használt már telefonnal tablettel kombinálva. Speciálisabb eszközöket (okosruha, okoscipő, kardiológiai adatokat mérő) eszközöket azonban nem. Az is kiderült, hogy nem mindenki alkalmazza őket megfelelően, illetve nem használja ki ezek lehetőségeit maximálisan, ami elsősorban abból adódik, hogy nem is ismerik megfelelően az eszközöket.

Milyen motiváló erővel hathat a technológia használata a sportban?

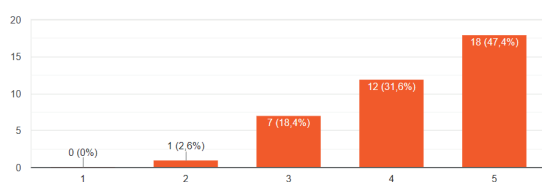
38 válasz



Arra a kérdésre, hogy van-e motivációs ereje a viselhető eszközöknek a sportban zömében pozitívan válaszoltak, elsősorban a gyors feedback-et jelölték meg pozitívként. A kicsit szkeptikusok inkább az eredmények megbízhatóságát vonták kétségbe, vagyis nem bíznak az eszközök pontosságában, aminek szintén lehet alapja.

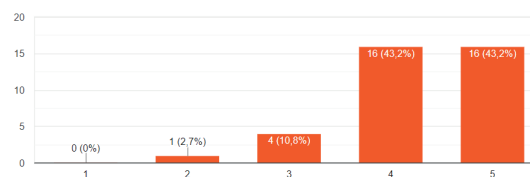
Fontosnak tartja-e a testnevelő képzés technológiai irányú fejlesztését a jövőben?

38 válasz



Fontosnak tartja-e a korszerű teljesítménymérő és analitikai eszközök, programok megismerését az egyetemi képzésben?

37 válasz



A képzés relevanciájára és tematikájára vonatkozó kérdések közül mindkettőre pozitívan reagáltak. A válaszadók mindkét esetben 70% felett nyilatkoztak a képzés technológiai megújítása és ezek megismerésének fontosságát illetően.

A miniinterjú eredményei

A mininterjúkérdései közül a két legérdekesebbet emelném ki, amelyek a digitális világ kihívásait boncolgatja.

1. *Ön szerint milyen hatással van a technológiai fejlődés a korosztálya életére (pro- és kontra)?*

A válaszok nagyon eltérőek voltak, és nem biztosak a technológia mindenhatóságában nagyon furcsának és ellentmondásosnak tartják azt a világot melyben élnek.

Volt, aki azt nyilatkozta a „*tamagochi én magam vagyok*” minden percét az elektronikus naptárja szerint éli szervezi. Volt, aki azt emelte ki, hogy a szűk időbeosztás mellett nagyon kedveli az elektronikusan elérhető tananyagokat, hiszen időben és térben független lehet a tanteremtől. Akadt olyan válasz is, aki a közösségi oldalak szerepét emelte ki elsősorban azt, hogy példaképeit tudja követni és ez inspirálja őt is nagyobb teljesítményre.

2. *Eddigi tanulmányaik során felkészítették-e Önöket az információs társadalom kihívásaira és megfelelő szintűnek tartják-e ezt?*

Az egyéni felkészültségük szintjéről az információs társadalomban nem nyilatkoztak pozitívan, sokan érzik az eddigi felkészültségüket hiányosnak, nem ismernek igazán jó, megbízható információforrásokat, jellemzően gyors célirányos kérdésekre kevesen kapnak releváns választ, nem tanulták a keresési stratégiákat, kevés olyan alkalmazást ismernek, amely megkönnyíti az életüket. (Pl: Neptun rendszerről órarend szinkronizálása mobilra.)

Összegzés

A kutatás eredményei bizonyítják, hogy érdemes a testnevelő képzésben az információ technológiai kérdésekkel foglalkozni, hiszen ezek az életük részei és mint kiderült, nagyon sok hallgatónak van mit pótolnia a kérdéskörben. Felmerül a kérdés ha bele tudjuk csempészni a mobiltechnológiát az iskolai testnevelés órákra is, lesz e motiváló hatása a mobil eszközök használatának az iskolai testnevelés órákon, illetve lesz- e ennek pozitív következménye? Például egy virtuális Balaton körbefutás során, ahol csapatban mindenki digitális térképen megjelenítve tudja majd a saját teljesítményét a többiekkel megosztani. Ezek a kérdések még nyitottak mindenképpen megéri kipróbálni őket.



Irodalom

- [1] ARDUINI, BORGOGNI, CAPELLI (2016): [The Portable Devices as Means to Promote Children's Active Lifestyle: The Case of a Walk To School Action in Italy](#) in: *Physical Education and new technologies* (2016) szerk: Novak, Antala, Knjaz, Zagreb 2016. pp: 19-25. ISBN: 978-953-7965-05-1
- [2] ARÉVALO, HERNANDO, MÓN, CATASÚS (2016): Physical Education And Ict: An Unstoppable Combination in: *Physical Education and new technologies* (2016) szerk: Novak, Antala, Knjaz, Zagreb 2016. pp: 61-69. ISBN: 978-953-7965-05-1
- [3] KNJAZ, RUPČIĆ, ANTEKOLOVIĆ (2016): Application of Modern Technology in Teaching and Training with Special Emphasis on Basketball Contents in: *Physical Education and new technologies* (2016) szerk: Dario Novak, Branislav Antala, Damir Knjaz, Zagreb 2016. pp: 112-113 ISBN: 978-953-7965-05-1
- [4] NETFIT koncepció: A 2015/2016. tanév NETFIT mérés eredményei tudományos elemzése
https://www.netfit.eu/public/pb_media.php?media=dokumentumok
(2019. 09.07.)
- [5] A viselhető eszközök trendjei: <https://www.ericsson.com/49e8cf/assets/local/trends-and-insights/consumer-insights/consumerlab/infographs/wearable-technology-infograph-ericsson-consumerlab-2016.pdf>
(2019. 09.07.)
- [6] Sports Analytics Market - Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2018 – 2026
<https://www.transparencymarketresearch.com/sports-analytics-market.html> (2019. 03.06)

Az informatikai gondolkodás (computational thinking) fogalmi keretei

Csernai Zoltán
Eszterházy Károly Egyetem
csernai.zoltan@uni-eszterhazy.hu

The conceptual framework of computational thinking

The Fourth Industrial Revolution radically changes the system of abilities and skills expected of the citizen of the 21st century. As priorities are shifted from knowledge to competence heretofore unseen demands emerge at the labour market. One of the current priorities is the improvement of STEM competences, that is, the acquisition of knowledge and skills related to (S)cience, (T)echnology, (E)ngineering, and (M)athematics via interdisciplinary and transversal instruction schemes.

According to experts the demand related to STEM competences and skills is expected to increase. Analysts forecast a 3% overall growth in employment in the EU until 2020, and this figure could reach up to 9% in STEM-impacted professions (Szegedi, 2014). Simultaneously computational thinking (Wing, 2008) is gaining increasing popularity as an umbrella term for the aforementioned four areas.

Subscribing to a deductive approach utilizing a specifically elaborated criteria system, I explore various interpretations of the given concept in the international community. Furthermore, I will analyse the respective manifestations in the DigComp 2.1. framework system (Carretero, Vuorikari & Punie, 2017) approved by the European Union.

This effort is a preliminary stage for further inquiry into the domestic situation especially the examination of the documentation of specific education management initiatives [MDOS, 2016 (Digital Education Strategy of Hungary); NAT (National Curriculum) 2018].

In the next phase of my research I will explore computation thinking-related attitudes in teacher training by the combined paradigm method (Sántha, 2014) My research tools include a self-developed measurement device along with questionnaires and interviews as well.

Keywords: Computational thinking, DigComp 2.1

Problémafelvetés

A 4. ipari forradalom hatására megváltozik a 21. századi állampolgártól elvárt képességek rendszere, a tudásról egyre inkább a kompetenciákra helyeződik a hangsúly, a munkaerőpiacokon a korábbtól eltérő elvárások jelennek meg. Napjaink egyik fő irányvonala a STEM területek fejlesztése, amely a (S) természettudományos, a (T) technológiai, a (E) mérnöki tudományokat és a (M) matematikát helyezi előtérbe, annak interdiszciplináris és transzverzális oktatása révén.

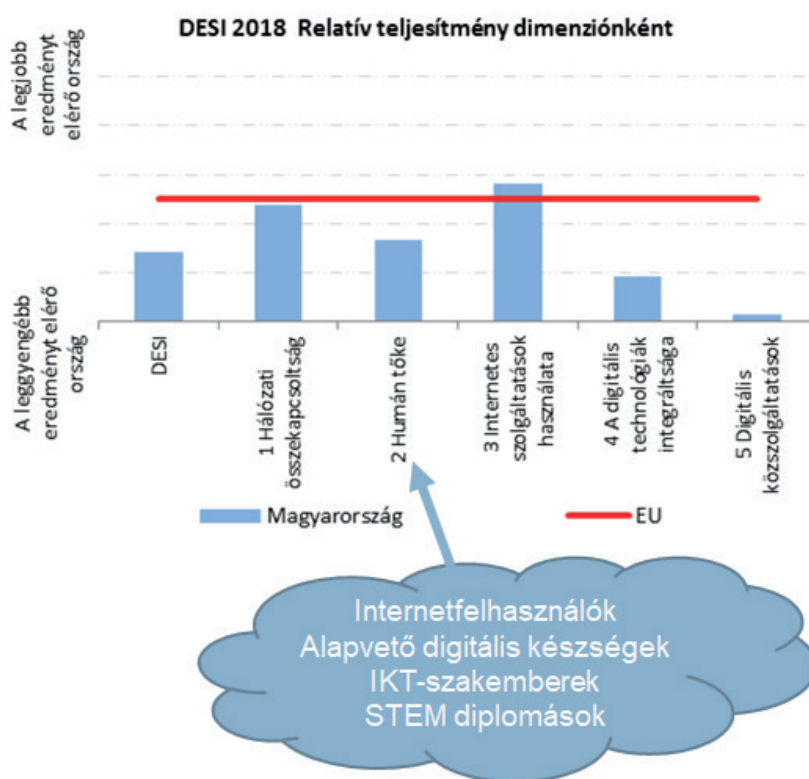


A STEM területekhez kapcsolódó képességek iránti kereslet ugyanis növekedni fog, az elemzők az 2020-ig az EU-ban átlagosan 3%-os foglalkoztatottsági növekedés várnak, a STEM és a társult szakmákon belül ez az érték 9% lehet.¹

A digitális gazdaság és társadalom fejlettségét mérő DESI-index (Digital Economy and Society Index –DESI) azt vizsgálja, hogy mennyire állnak készen az Európai Unió tagállamai a digitális átállásra.

A DESI egyik pillére (2. Humán tőke) azt méri, hogy az egyénnek milyen képességekre van szüksége a digitális társadalomban való boldoguláshoz, amelyhez az alábbi indikátorok szükségesek (1. ábra):

- Internetfelhasználók
- Alapvető digitális készségek
- IKT-szakemberek
- STEM diplomások



1. ábra: A digitális gazdaság és társadalom fejlettségét mérő mutató (DESI), 2018, Magyarországról szóló országjelentés

(Forrás: Digital Economy and Society Index 2018 – Magyarország. URL.: http://europa.eu/rapid/press-release_MEMO-18-3737_en.htm)

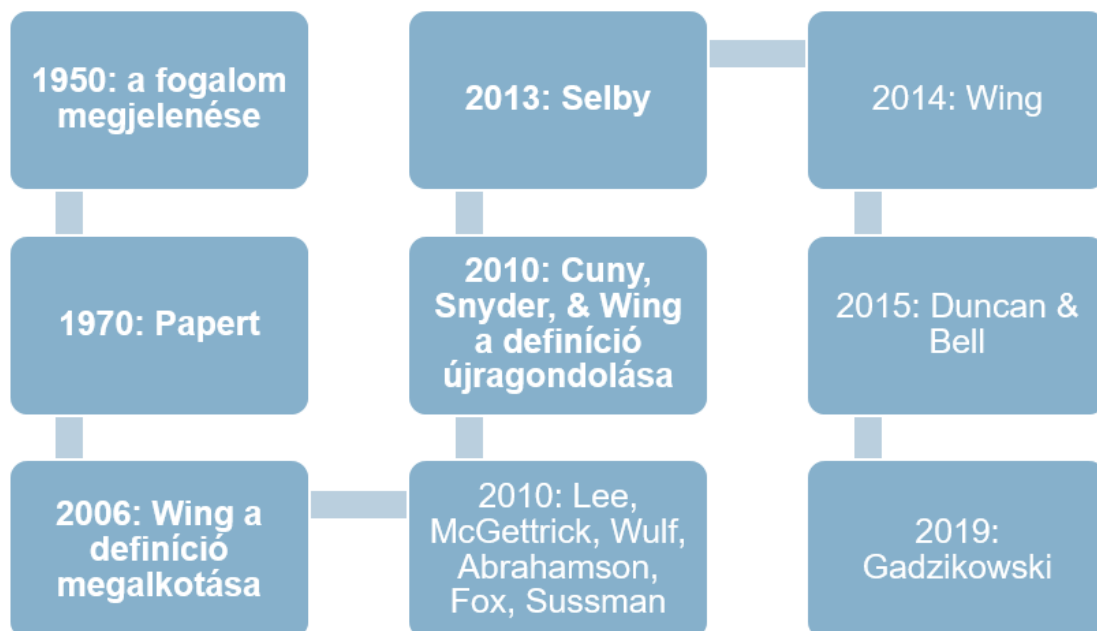
¹ Szegedi Eszter. 'Miért került világszerte fókuszba a STEM területek oktatása?' pp. 9-14., 2014. <http://tka.hu/docs/palyazatok/belestem.pdf>

Az oktatás digitális átállása ugyan elkezdődött, azonban nincsen egységes fogalomkészlete. Bizonyos esetekben, tévesen szinonimaként utalnak az emberek az informatikai gondolkodásra, a digitális kompetenciára és az algoritmikus gondolkodásra.

Kutatásom célja, hogy deduktív módon megvizsgáljam, hogy nemzetközi szinten hogyan értelmezik e fogalmat, és egy kidolgozott szempontrendszer alapján elemezzem az Európai Unió által elfogadott² keretrendszerben való megjelenését.

A Computational Thinking fogalmának értelmezései

A Computational Thinking (CT) fogalmát a nemzetközi szakirodalomban nem egységesen értelmezik, többnyire a fogalom egyes részterületeiben rejlő fejlesztési lehetőségekre fókuszálnak. Kutatásom során a CT fogalmi fejlődésének legfontosabb állomásait mutatom be. (2. ábra)



2. ábra: A Computational Thinking (CT) fogalom „evolúciójának” legfontosabb állomásai
(Forrás: a szerző saját ábrája)

Az 1950-es években jelent meg a CT fogalma a számítógép-tudomány jellemzőjeként. Az 1970-es években Seymour Papert, az MIT professzora, a Logo programozási nyelvet a CT-n keresztül mutatta be a diákoknak.

2 Stephanie Carretero, Riina Vuorikari, Yves Punie. "DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2017.
[http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_\(online\).pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_(online).pdf)



A Computational Thinking fogalmát Jeanette Wing a 2006-ban megjelent cikkében határozta meg. Véleménye szerint az informatikai gondolkodás magába foglalja a problémák megoldását, a rendszerek tervezését, és az emberi viselkedés megértését, a számítástudomány alapelvei alapján.³

Jeanette Wing meg van arról győződve, hogy a XXI. század közepére az írás, az olvasás és a számtan mellett az informatikai gondolkodás megjelenik, mint a 4. alapvető készség.

Jeanette Wing a CT fogalmát 2010-ben újragondolta, és megalkotta az egyik leggyakrabban idézett meghatározást, miszerint az informatikai gondolkodás a problémák megoldására irányuló gondolkodási folyamat, amely segíti, hatékonyabbá teszi az információ feldolgozásának műveletét.⁴

Cynthia C. Selby értelmezésében az informatikai gondolkodáshoz szükséges képességek: absztrakt fogalmakban gondolkodás, részekre bontás a gondolkodás során, algoritmikus gondolkodás, értékelésben való gondolkodás és általánosítás képessége a gondolkodás során.⁵

Gerald Jay Sussman professzor meghatározása alapján az informatikai gondolkodás szigorú elemzést és eljárásokat foglal magában egy meghatározott feladat hatékony végrehajtásához.⁶

Ann Gadzikowski kutatási alapján az informatikai gondolkodás négy készségkategória kombinációja: mintafelismerés, algoritmusok létrehozása és használata, elemi részekre bontás, az absztrakciók megértése.⁷ A mintafelismerés során az óvodások elsajátítják az alakzatok formák és színek szerint történő szétválogatását. Az algoritmus szó valamilyen műveletsort, tevékenységet jelent, amellyel egy adott probléma megoldását adjuk meg. Olyan egyszerű feladatok megoldására kell itt gondolni, mint pl. egy gomb felvarrásának vagy egy pite sütésének lépései. Az elemi részekre történő bontással pl. a matematikában találkozhatunk, amikor egy háromjegyű számot helyiérték szerint fel kell bontanunk százasokra, tízesekre és egyesekre. Az absztrakciók megértése magában foglalja az általánosítások

3 Jeannette M. Wing. "Computational Thinking," Communications of the Association for Computing Machinery Viewpoint, March 2006, pp. 33-35.
<http://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf>

4 Jan Cuny, Larry Snyder, and Jeannette M. Wing. "Demystifying Computational Thinking for Non-Computer Scientists" work in progress, 2010.

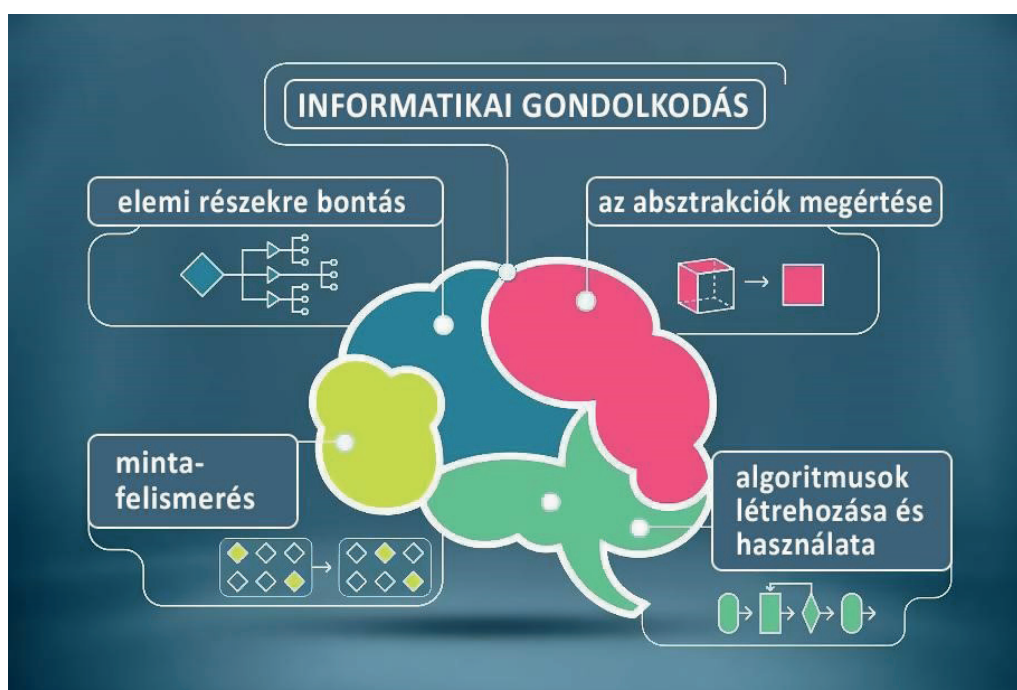
5 Pluhár, Zsuzsa és Torma, Hajnalka és Törley, Gábor. "Hallgatói teljesítményértékelés az algoritmikus gondolkodás tükrében." In: InfoDidact 2018. Webdidaktika Alapítvány, Budapest, pp. 1-10., 2019. <http://realmtak.hu/92129/>

6 National Research Council. "Report of a Workshop on the Scope and Nature of Computational Thinking." Washington, DC: The National Academies Press, 2010.
<https://doi.org/10.17226/12840>

7 Ann Gadzikowski. "Planting the Seeds of Computational Thinking in Early Childhood", 2019.
<http://anngadzikowski.com/planting-the-seeds-of-computational-thinking-in-early-childhood/>

készítését, a következtetések levonását és más problémamegoldást célzó gondolkodási folyamatok tervezését.

Összegzésként kijelenthetjük, hogy az informatikai gondolkodás az 1950-es években kialakult felfogáshoz képest jelentős tartalombővülésen ment keresztül az elmúlt évtizedekben, hiszen a számítástudományi gyökerektől elindulva napjainkra egy átfogóbb, az egyén információ feldolgozása közben végbemenő kognitív és metakognitív folyamatokra fókuszáló, összetett fogalmat értünk alatta.⁸ A CT tehát azonban a programozással, hiszen ez egy funkcionális gondolkodási alap(készség), amelyben az egyén gondolkodásmódja legyen az előtérben, nem pedig az, ahogyan a gépek gondolkodnak.



3. ábra: A Computational thinking (CT) készségkategóriái

A kép angol nyelvű forrása:

<https://www.nextgurukul.in/thenextworld/self-learning/what-is-computational-thinking/>

Az informatikai gondolkodás alatt problémák elemei egységekre történő felbontását értjük, a minták megfigyelését és tendenciák felismerését, a problémamegoldás lépéseinek meghatározását, valamint az absztrakciót, azaz a részletekről történő elvonatkoztatást, a releváns információkra koncentrációt. E fogalom tehát ötvözi és kiegészíti a matematikai és mérnöki gondolkodást, tartalmazva azon mentális eszközöket, amelyek a számítógép-tudományterületén használatos szakkifejezések széles skáláját tükrözik.⁹

8 Romero, M., Lepage, A. & Lille, B. "Computational thinking development through creative programming in higher education.", 2017. <https://doi.org/10.1186/s41239-017-0080-z>

9 Pluhár Zsuzsa. "Az informatikai gondolkodás és a hód." InfoDidac2016 Módszertani konferencia, 2016. https://people.inf.elte.hu/szlavi/InfoDidac16/Manuscripts/PZs_Hod.pdf



A Computational Thinking (CT) tehát ernyőfogalomként hatja át a STEM egyes területeit, hiszen részei többek között az algoritmikus gondolkodás, a problémamegoldás, a programozás oktatása és a szimulációs játékok alkalmazása, szoros kapcsolatot mutatva az információs műveltséggel és a digitális kompetenciával.

A Computational Thinking fogalmának megjelenése a DigComp 2.1-ben

Napjainkban a digitális kompetencia nemcsak az IKT-hoz való hozzáférést és használatát jelenti, hanem magában foglalja a szükséges ismeretek, készségek és attitűdök birtoklását is.

A 2013-ban megjelent DigComp 1.0 (DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe) öt kompetenciaterületet jelölt meg: információ, kommunikáció, tartalomkészítés, biztonság, problémamegoldás.¹⁰

A tanulmányban beazonosításra kerültek a digitális állampolgárság komponensei is, amelyet 21 kompetenciaelemre bontottak.

A DigComp 2.0 (DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model) az első keretrendszert további elemekkel bővítette, illetve kitért a kompetenciaelemekben való jártasság szintjeire is.¹¹

A cikk írásakor a legfrissebb, 2017-ben megjelent DigComp 2.1 (DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use) nyolc jártassági szintet és példát mutat be a tanulás és a foglalkoztatás területein.¹²

Az European Commission weboldalán megtalálható az Európai Unió saját definíciója a Computational Thinking-ről, miszerint: az informatikai gondolkodás egy „számítógépes tudósként” való gondolkodás rövidítése, vagyis a számítógépes

10 Ferrari, Anusca. "DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2013. <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC83167/lb-na-26035-enn.pdf>

11 Riina Vuorikari, Yves Punie, Stephanie Carretero, Lieve Van den Brande. "DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2016. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101254/jrc101254_digcomp%202.0%20the%20digital%20competence%20framework%20for%20citizens.%20update%20phase%201.pdf

12 Stephanie Carretero, Riina Vuorikari, Yves Punie. "DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2017. [http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_\(online\).pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_(online).pdf)

tudomány fogalmának használata a problémák megfogalmazásához és megoldásához."¹³

A kutatás folytatása

E munka egy előkészítő szakaszt képez a további elemzésekhez, amelyben a hazai helyzetet vizsgálom meg, olyan tekintetben, hogy az oktatásirányítási dokumentumokban (MDOS, 2016; NAT 2018) hogyan jelenik meg e fogalom, valamint annak elemei.

A kutatásom következő fázisában egy attitűdkutatásra kerül sor, amely során az informatikai gondolkodással kapcsolatos vélekedéseket vizsgálom meg, kombinált paradigma módszerével (Sántha, 2014) a pedagógusképzésben, egy saját fejlesztésű mérőeszközzel, kérdőív, majd interjú formájában.

Irodalom

Ann Gadzikowski. "Planting the Seeds of Computational Thinking in Early Childhood.",2019.

<http://anngadzikowski.com/planting-the-seeds-of-computational-thinking-in-early-childhood/>

Digital Economy and Society Index 2018 – Magyarország.

http://europa.eu/rapid/press-release_MEMO-18-3737_en.htm

European Commission. <https://ec.europa.eu/jrc/en/computational-thinking>

Ferrari, Anusca. "DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2013. <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC83167/lb-na-26035-enn.pdf>

Jan Cuny, Larry Snyder, and Jeannette M. Wing. "Demystifying Computational Thinking for Non-Computer Scientists" work in progress, 2010.

Jeannette M. Wing. "Research Notebook: Computational Thinking - What and Why? The Link." Pittsburgh, PA: Carneige Mellon, 2011.

<https://www.cs.cmu.edu/~7ECompThink/resources/TheLinkWing.pdf>

Jeannette M. Wing. "Computational Thinking," Communications of the Association for Computing Machinery Viewpoint, March 2006, pp. 33-35.

<http://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf>

Kampylis, P., Punie, Y. "The Computational Thinking Study." EU Science Hub, 2016.

<https://ec.europa.eu/jrc/en/computational-thinking>

¹³ Kampylis, P., Punie, Y. "The Computational Thinking Study." EU Science Hub, 2016.
<https://ec.europa.eu/jrc/en/computational-thinking>



National Research Council. "Report of a Workshop on the Scope and Nature of Computational Thinking." Washington, DC: The National Academies Press, 2010. <https://doi.org/10.17226/12840>

Pluhár Zsuzsa. "Az informatikai gondolkodás és a hód." InfoDidac2016 Módszertani konferencia, 2016. https://people.inf.elte.hu/szlavi/InfoDidact16/Manuscripts/PZs_Hod.pdf

Pluhár, Zsuzsa és Torma, Hajnalka és Törley, Gábor. "Hallgatói teljesítményértékelés az algoritmikus gondolkodás tükrében." In: InfoDidact 2018. Webdidaktika Alapítvány, Budapest, pp. 1-10., 2019. <http://real.mtak.hu/92129/>

Racsko Réka. "Digitális átállás az oktatásban." Gondolat Kiadó, Veszprém. Iskolakultúra-könyvek 52., 2017. http://misc.bibl.u-szeged.hu/46196/1/iskolakultura_konyvek_052.pdf

Riina Vuorikari, Yves Punie, Stephanie Carretero, Lieve Van den Brande. "DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2016. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101254/jrc101254_digcomp%202.0%20the%20digital%20competence%20framework%20for%20citizens.%20update%20phase%201.pdf

Romero, M., Lepage, A. & Lille, B. "Computational thinking development through creative programming in higher education.", 2017. <https://doi.org/10.1186/s41239-017-0080-z>

Selby C. "Computational Thinking: The Developing Defenition." Submitted for ItiCSE Conference 2013. <http://people.cs.vt.edu/%7Ekafura/CS6604/Papers/CT-Developing-Definition.pdf>

Stephanie Carretero, Riina Vuorikari, Yves Punie. "DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use." Luxembourg, Sweden: Joint Research Centre of the European Commission, 2017. [http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_\(online\).pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_(online).pdf)

Szegedi Eszter. "Miért került világszerte fókuszba a STEM területek oktatása?" pp. 9-14., 2014. <http://tki.hu/docs/palyazatok/belestem.pdf>

A felhasználói viselkedés, mint információbiztonsági kockázat becslése

Leitold Ferenc

Secudit Kft.

fleitold@secudit.com

A felhasználói viselkedés sokkal veszélyesebbé teszi egy szervezet működését, mint bármely technikai sebezhetőség. Számos technikai megoldás létezik a technikai lehetőségeket kihasználó támadásokkal szemben. Az informatikai biztonság technikailag egy jól felkészült védelmet képes biztosítani, ugyanakkor az emberi tényező kezelése még mindig a gyerekcipőben jár. Ez az előadás a felhasználói viselkedés értékelésének lehetséges módszereivel foglalkozik, alapvetően néhány hasznos megfigyelési lehetőség kerül előtérbe, melyek alkalmasak (akár automatikus módon) a felhasználói viselkedés mérésére. A bemeneti források, amiket használhatunk, az adott felhasználó által használt munkaállomásból, a hálózati forgalomból és az alkalmazásnaplókban (különösen a védelmi rendszerek naplóiból) származhatnak. Ezeket a bemeneti forrásokat használva néhány nagyon hasznos metrikát is definiálhatunk a felhasználók viselkedésének megítéléséhez, illetve a felhasználók besorolásához. Miután meg tudtuk mérni a felhasználói magatartás szintjét, használhatjuk azt az adott szervezet informatikai biztonságának javítására is.

A tanulmány a következő részeket tartalmazza:

- Bevezetés a felhasználói viselkedés kezelésébe (mérés és cselekvés),
- Hasznos eszközök a felhasználók viselkedésével kapcsolatos néhány alapvető jel automatikus méréséhez,
- A mért adatok kiszámítása a mért jelekből,
- Mit tehetünk a metrikákkal?

Keywords: vulnerability assessment, user behavior assessment, human factor

Bevezetés

A DVA (Distributed Vulnerability Assessment) technológia a Dunaújvárosi Egyetem és a Secudit közös kutatási munkája alapján jött létre. A DVA részletes leírást ad egy szervezet internetes támadási sebezhetőségeiről. A módszer szerint első lépésként az egyedi felhasználók és az informatikai infrastruktúra elemeinek sebezhetőségét az egyes ismert fenyegetésekre vonatkozóan kell felmérni, majd ezeket az eredményeket kombinálni az adott szervezet számára releváns fenyegetésekre vonatkozóan. A módszer egy adott szervezet integrált kiber-támadási sebezhetőségét a jelenleg ismert fenyegetések elterjedtségét és hatékonyságát; a felhasználók biztonságtudatos viselkedését; és az informatikai infrastruktúra gyengeségeit alapul véve értékeli. Matematikai módszereket alkalmazva az integrált sebezhetőség felbontható arra, hogy az egyes felhasználók, illetve az egyes IT infrastruktúra elemek milyen mértékben járulnak hozzá az integrált sebezhetőséghez, a teljes szervezet fenyegetettségéhez. A DVA-eredményekből a fenyegetettség mennyiségi szempontból hozzárendelhet a



különböző belső hozzájáruló összetevőkhöz (például felhasználói azonosító, portok, protokollok, védelmi rétegek). Ez lehetővé teszi, hogy különböző közreműködő komponenseket összehasonlítható mérőszámokkal értékeljünk (pl. felhasználói biztonságossági tudatosság, az infrastruktúra javításának lehetősége, illetve a rosszindulatú programok elleni védelem hatékonysága alapján). A DVA lehetővé teszi az információbiztonsági menedzserek számára, hogy a "mi lenne, ha" típusú lekérdezések eredményei alapján összehasonlíthassák a különböző rendelkezésre álló lehetőségeket a szervezet fenyegetettségének csökkentése érdekében, amelyek egyébként nem lennének mennyiségi szempontból összehasonlíthatók (pl. további cybersecurity alkalmazások és szolgáltatások.)

A DVA módszer működésének alapfeltétele, hogy a felhasználói biztonságtudatosságot is automatikusan lehessen mérni. Ebben a cikkben a felhasználói biztonságtudatosságra jellemző metrikával foglalkozunk, milyen jellemzők alapján mérhető.

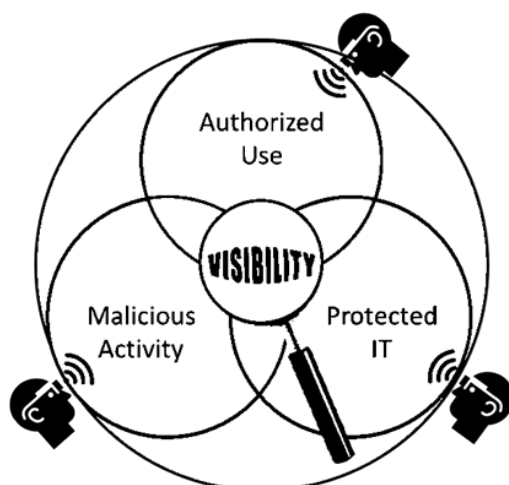
1. Fenyegetettségek modellezése

Ahhoz, hogy egy kártékony támadás sikeres legyen egy védett hálózattal szemben, a kártékonykódsikeresvégrehajtásaszükséges. Afelhasználóioldalonalegegyszerűbb minimális viselkedés nem más, mint a végpont eszköznek az internethez történő csatlakoztatása. Az informatikai biztonsági metrikák manapság a védett IT-re (pl. folyamatos sérülékenység-tesztelés), illetve a kártevők tevékenységére, tulajdonságaira (pl. védelmi rendszerek tesztelése) [6] fókuszálnak. A felhasználói magatartásra vonatkozó informatikai biztonsági metrika kevésbé fejlett [3], habár a hálózati forgalom megfigyelése lehetőséget ad a fejlesztésekre (pl. NetFlow/IPFIX). A passzív figyelés mellett az interaktív metrikát is alkalmazhatjuk [10].

A sikeres kártékony támadásokat a védett környezetben megvalósítható kártékony tevékenység és a megfelelő felhasználói magatartás metszeteként lehet reprezentálni. Ez a koncepcionális keret az NSS Lab által használt működési szabályokra épül [18], ugyanakkor praktikus és kényelmes egyszerűsítése a támadási felületek komplett kezelésének. Az alábbiakban csak a humán-interaktív végpontokra fókuszálunk (IT), a beágyazott rendszerek biztonsági architektúrájával (IoT, OT) jelenleg nem foglalkozunk. Három különálló, de erősen interaktív sérülékenységi forrást veszünk figyelembe:

- (1) kártékony tevékenység azok által, akik saját céljaikra használják ki a hálózat képességeit, hogy megsértsék a megbízható IT rendszer védelmét;
- (2) veszélyes IT felhasználói magatartás (pl. alkalmazottak, vevők, beszállítók); és
- (3) védelem nélküli sérülékenység az IT hálózati infrastruktúrában.

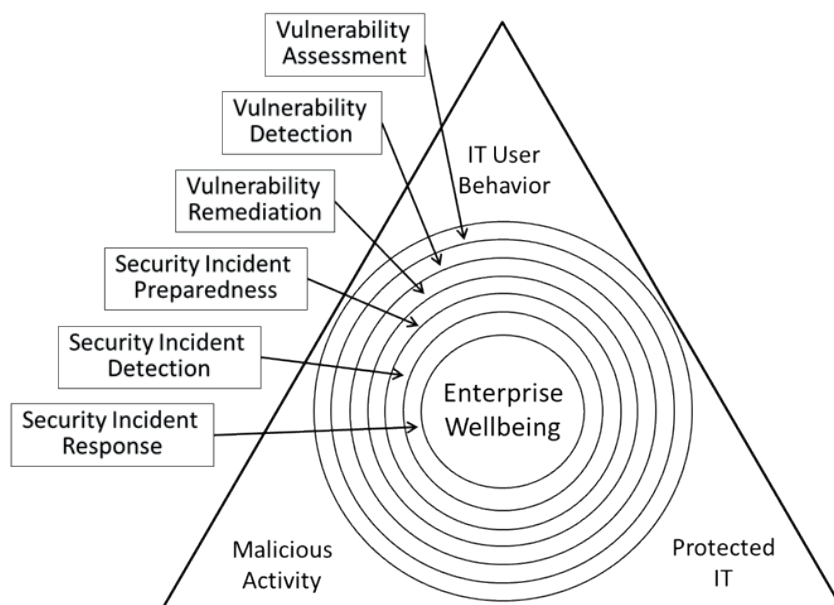
A legkritikusabb sérülékenység e három terület közös részében, metszetében található (1. ábra). E sérülékenységek megfelelő láthatóságot, ellenőrzést és megkülönböztetést követelnek a megfigyelésükhöz, megértésükhöz és az ellenük történő hatékony védekezéshez. A meglevő és esetleg felmerülő sérülékenységek láthatóságához éber



1. ábra: Az IT sérülékenység komponenseit és tényezőit három területre lehet osztani, melyek mindegyikének saját módszere és eszközei vannak a láthatóság, ellenőrzés és megkülönböztetés céljára [13]

kockázatelemzés szükséges, ami mindhárom területet figyeli (1. ábra).

Az információs folyamatok sérülékenységeinek láthatósága szükséges, de önmagában elégtelen az informatikai biztonság szempontjából. A sérülékenység értékelése a biztonság biztosításának legkülsőbb rétege. A következő rétegek: sérülékenység érzékelés, sérülékenység javítása, biztonsági incidensre való felkészülés, biztonsági incidens érzékelése, és biztonsági incidensre való reagálás (2. ábra).



2. ábra: Sérülékenység felmérése a teljes biztonság érdekében a szervezet jóléte céljából [13]

A szervezetjólétének biztosításához a sérülékenységek kezelése a sérülékenységek forrásainak gyakorlati és hatékony azonosítását követeli meg. A biztonsági incidensre való reagálás követelményét az esemény információkezelő rendszerek elégítik ki (SIEM). A sérülékenységek hatékony kezeléséhez az informatikai sérülékenység



hármass modellje szükséges. A korábbi szabályokból eredően [11, 12] a hármass modell a sérülékenységek mérését 3 forrásra osztja: i) kártékony tevékenység; ii) védelemmel rendelkező IT; és iii) nem megfelelő felhasználói magatartás. Mindegyik forrásban specifikus tényezőket azonosítunk és jellemzünk (pl. vírusküldés és kihasználás a kártékony tevékenységi hármassban). A modell alapot ad a tényezők korrelációjához és kombinálásához a sérülékenységek integrált nézetéhez.

2. A DVA számításának matematikai háttere

Az elosztott fenyegetettség felmérés (DVA) matematikai háttere valószínűségszámításra épül ([19]):

Legyen

$$\mu(t, u) = \frac{\text{number of attempts of } t \text{ are enabled by the user } u}{\text{number of attempts of } t \text{ are enabled by the average user}}$$

Ekkor

$$p_s(l) = 1 - \prod_{\text{for all } t, u \text{ and } i} (1 - p_{\text{user}}(t, u) \cdot p_{\text{device}}(t, i) \cdot p_{\text{prev}}(t, l))^{k(t, u)}$$

ahol $u \in U$, $i \in I$, $t \in T_l$, $l \in L$ és

$$k(t, u) = \frac{T}{\Delta T} \cdot \frac{T_u}{T_{\text{average}}} \cdot \mu(t, u)$$

ahol

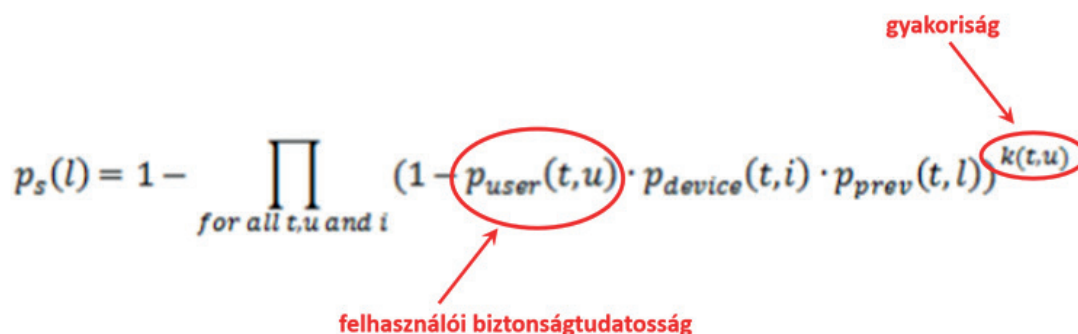
ΔT az elterjedtségekre (prevalence) vonatkozó időintervallum,

T az az időintervallum, amire az általunk számított valószínűségi mérték vonatkozik,

T_u az az időintervallum, amennyi ideig az u felhasználó használja a számítógépet,

T_{average} az az időintervallum, amennyi ideig egy átlagos felhasználó használja a számítógépet, $\mu(t, u)$ a fentiek alapján számított érték.

A fenti képletben a felhasználói viselkedés az alábbi tényezőkben jelenik meg:

$$p_s(l) = 1 - \prod_{\text{for all } t, u \text{ and } i} (1 - p_{\text{user}}(t, u) \cdot p_{\text{device}}(t, i) \cdot p_{\text{prev}}(t, l) \cdot k(t, u))$$


A DVA modell szerint a felhasználói viselkedés két tényező formájában jelenik meg, egy számítógép felhasználónak két lényeges tulajdonsága van, melyek befolyásolják a szervezet veszélyeztetettségi mértékét. Egyrészt a felhasználói biztonságtudatosságra jellemző mérték, mely azt mondja meg, hogy egy adott felhasználó esetén, ha egy bizonyos szituációba kerül milyen valószínűséggel hoz olyan döntést vagy végez olyan tevékenységet, mely aztán a veszély okozta esemény bekövetkezését eredményezi. A másik jellemző tulajdonság arra utal, hogy egy adott felhasználó milyen gyakran kerül olyan szituációba, hogy döntést kelljen hoznia. Ez utóbbi esetben rengeteg olyan eszköz, szolgáltatás áll rendelkezésre, mely képes arra, hogy például a felhasználók be- és kijelentkezési idejét, a szoftverek használatát naplózza. Ebben a cikkben a továbbiakban elsősorban a felhasználói biztonságtudatossággal, annak mérésével, illetve a felhasználói biztonságtudatossági mérték meghatározásával foglalkozunk.

3. Felhasználói biztonságtudatosság mérése

A felhasználói biztonságtudatosság esetén két lényeges kérdés merül fel: egyrészt azt vizsgáljuk meg, hogy melyek azok a mérhető jelek, melyek szoros kapcsolatban vannak a biztonságtudatossággal, illetve a mérési módszer lehetőségeit is megvizsgáljuk.

Mit mérhetünk?

A felhasználói biztonságtudatosság esetén számos olyan mérhető információ áll rendelkezésre, melyek segíthetnek a felhasználói biztonságtudatosságra jellemző mérték meghatározásában. Ilyenek például:

- Eszköz használata
- Alkalmazások használata (főként: kommunikációs alkalmazások)
- Különböző típusú fájlok megnyitása/indítása
- Védelmek befolyásolása (pl.: frissítés, felfüggesztés)
- Böngészés az interneten

Hogyan mérhetünk?

A felhasználói biztonságtudatosság mérésére manuális, kérdőív, tesztjellegű felmérés az általában megszokott módszer. Ennek a legnagyobb hátránya, hogy nem a tényleges viselkedés biztonságosságát méri, hanem azt, hogy a felhasználó



milyen biztonságtudatossági ismeretekkel rendelkeznek. Azaz a mérés hibás lesz azokra vonatkozóan, akik pontosan tudják, hogy mi a helyes, biztonságtudatos viselkedés, de nem azt teszik.

A biztonságtudatosság mérése alapvetően két módszerrel képzelhető el:

- A **passzív módszer** révén a felhasználók szokásos viselkedését vizsgáljuk és ebben keresünk olyan jeleket, amelyek olyan magatartásra utalnak, amelyek valamilyen veszély/támadás elhárítását vagy elősegítését jelentik.
- Az **aktív módszer** segítségével szándékosan előidézzük olyan szituációkat, melyekkel valamilyen veszély/támadás esetén a felhasználói döntést szimuláljuk. Ebben az esetben a felhasználók szokásos viselkedése helyett a felhasználók előidézett szituációkban történő döntését.

Mind a passzív, mind az aktív módszernek megvannak az előnyei és hátrányai. Passzív módszer esetén a felmérés semmilyen módon nem akadályozza a szervezet szokásos működését, míg az aktív módszer esetén ez nem igaz. Az aktív módszer viszont különböző típusú szituációkra vonatkozóan képes a mérést elvégezni, míg a passzív esetben csak olyan szituációk esetén történik ez meg, amelyek ténylegesen előfordulnak.

4. Felhasználói esetek

Amennyiben a felhasználók biztonságtudatos viselkedéséhez valamilyen mérőszámot szeretnénk hozzárendelni, alapvető, hogy ez a mérőszám lehetőséget adjon az összehasonlításra. Az alábbiakban ennek fényében három egyszerű esetet vizsgálunk.

I. eset

Az első esetben a különböző böngészési szokások oldaláról közelítjük meg a felhasználói biztonságtudatosságot. Adott egy felhasználó, aki minden munkanap a böngészőben csak a www.port.hu és a www.idokep.hu oldalakat nyitja meg. Egy másik felhasználó pedig a böngészőjében minden héten legalább 20 olyan weboldalt nyit meg, amit nem látogatott az elmúlt 6 hónapban. Minden egyéb vonatkozásban a két felhasználó viselkedése azonos. Ebben az esetben azt mondhatjuk, hogy a második felhasználó nyilván nagyobb veszélyt jelent, azaz jobban hozzájárul a szervezet veszélyeztetettségéhez.

II. eset

A második esetben azt mutatjuk meg, hogy akár különböző munkahelyi feladatkörök is befolyásolják a felhasználói biztonságtudatosságot. Adott egy felhasználó, aki a HR osztályon dolgozva, fogadja az álláspályázatok email-ben és munkaköri feladata, hogy megnyissa a bennük lévő PDF csatolmányokban lévő önéletrajzokat. Egy másik felhasználó pedig soha nem kapott olyan email-t, amiben PDF melléklet lett volna. Minden egyéb vonatkozásban a két felhasználó viselkedése azonos. Ebben az esetben azt mondhatjuk, hogy az első felhasználó nyilván nagyobb veszélyt jelent, azaz jobban, azaz jobban hozzájárul a szervezet veszélyeztetettségéhez. Ez tehát nem abból adódik, hogy az információbiztonsági tudása alacsonyabb lenne, csupán a munkahelyi feladatköre jelenti a nagyobb kockázatot.

III. eset

A harmadik esetben az első és a második esetet egyesítjük. Adott egy felhasználó, aki a HR osztályon dolgozva, fogadja az álláspályázatok email-ben és munkaköri feladata, hogy megnyissa a bennük lévő PDF csatolmányokban lévő önéletrajzokat, ugyanakkor a böngészőben csak a www.port.hu és a www.idokep.hu oldalakat nyitja meg. Egy másik felhasználó pedig soha nem kapott olyan email-t, amiben PDF melléklet lett volna, ugyanakkor a böngészőjében minden héten legalább 20 olyan weboldalt nyit meg, amit nem látogatott az elmúlt 6 hónapban. Minden egyéb vonatkozásban a két felhasználó viselkedése azonos. Ebben az esetben nem tudjuk egyértelműen megállapítani, hogy melyik felhasználó jelent nagyobb veszélyt a szervezet számára. Ezt akkor tudnánk megtenni, ha ismerjük azokat a veszélyforrásokat, amelyek emailben, PDF csatolmányokban terjednek, illetve azokat a veszélyforrásokat, amelyek a böngészési szokásokra építenek.

5. Összegzés

A fentiekben módszert mutattunk be a sérülékenység mérésére. Három információforrást használunk: külső informatikai fenyegetés intelligencia („biztonsági intelligencia”), szervezeti IT infrastruktúra gyengeség („behatolás tesztelés”), és a felhasználók fogékonysága, naivsága a támadásokra („felhasználói magatartás”). A módszer lehetővé teszi a mért források kombinálását egy metrikába, amit összevethető sérülékenységekre bonthatunk. A módszer számszerűsíti a relatív sérülékenység evolúcióját időben, külön mérheti az egyedi osztályok (LAN) sérülékenységét és a specifikus fenyegetéseket (pl. zsaroló vírusok, adathalászat). A módszer előrejelzi a potenciális javítási tevékenység következményeit („Mi lesz, ha?”), ezáltal segíti a biztonsággal kapcsolatos döntéshozatalt az adott helyzetben.

A programozott fenyegetések száma manapság 7-800 millió körüli, az aktív támadások köre folyamatosan változik, ráadásul a támadások kb. 90%-át egyedi fertőzések okozzák. Ilyen körülmények között az egy szervezetre vonatkozó veszélyeztetettség mérése sokkal inkább becslés, mint pontos számítás. A bemeneti adatok minél pontosabb meghatározásával, a figyelembe vett kártevők körének kiválasztásával pontosabbá tehető az analízis.

A módszer egyik legfontosabb összetevője a felhasználói viselkedés mérése. A 4. fejezetben leírtak szerint azonban a felhasználói biztonságtudatosság nem csupán egy felhasználórajellemzőmetrika, hanem ezt veszélyforrásonként kell meghatározni az egyes felhasználók vonatkozásában. Ennek érdekében a viselkedésre vonatkozó jeleket, jelzéseket kell mérni, majd ezek segítségével határozhatjuk meg a különböző támadási vektorok vonatkozásában a felhasználókra jellemző metrikákat. Minden, a felhasználók viselkedésére vonatkozó mérés természetesen felveti a GDPR vonatkozását is. A GDPR szempontjából a felhasználói viselkedésre vonatkozó mérések, egy szervezet veszélyeztetettségi előrejelzése az információbiztonság javításának az irányába tett lépésnek tekinthető és így a GDPR céljait szolgálja. Természetesen a GDPR-nak megfelelően a módszer csakis az arányosság elvét szem előtt tartva és a felhasználók megfelelő tájékoztatásával alkalmazható.



Irodalom

- [1] ARROTT, A., F. Lalonde Levesque, D. Batchelder, and J.M. Fernandez. "Citizen cyber-security health metrics for Windows computers". Proceedings of Eastern European eGov Days Conference, EEGOV, Budapest, Hungary. 2016.
- [2] BATCHELDER, D., et al. "Microsoft Security Intelligence Report." Volume 18: July-December 2014, Microsoft, 2015.
- [3] CHAPMAN, M.T., "Establishing metrics to manage the human layer." ISSA Security Education Awareness Special Interest Group, 2013.
- [4] CLEMENTI, Andreas, Peter Stelzhammer, and Fernando C. Colon Osorio. "Global and local prevalence weighting of missed attack sample impacts for endpoint security product comparative detection testing." Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on. IEEE, 2014.
- [5] COLON OSORIO, F.C., and A. Arrott. "Fabric of security - changing our theory and expectations of modern security". Proceedings of Eastern European eGov Days Conference, EEGOV, Budapest, Hungary. 2016.
- [6] EDWARDS, S.E., R. Ford, and G. Szappanos., "Effectively testing APT defenses". Virus Bulletin Conference, Prague, Czech Republic, 2015.
- [7] KLEINER, A., P. Nicholas, K. Sullivan, "Linking Cybersecurity Policy and Performance, Microsoft Trustworthy Computing", 2013,
- [8] KSHETRI, Nir. "Cybercrime and Cybersecurity in the Middle East and North African Economies." Cybercrime and Cybersecurity in the Global South. Palgrave Macmillan UK, 2013.
- [9] LALONDE LEVESQUE, F., A. Somayaji, D. Batchelder, and J.M. Fernandez. "Measuring the health of antivirus ecosystems." Malicious and Unwanted Software (MALWARE), 2015 10th International Conference on. IEEE, 2015.
- [10] LALONDE LEVESQUE, F., J. M. Fernandez, and A. Somayaji. "Risk prediction of malware victimization based on user behavior." Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on. IEEE, 2014.
- [11] LEITOLD, F and K. Hadarics. "Measuring security risk in the cloud-enabled enterprise." Malicious and Unwanted Software (MALWARE), 7th International Conference on Malicious and Unwanted Software, pp: 62-66, ISBN: 978-1-4673-4880-5. 2012.
- [12] LEITOLD, F. "Security Risk analysis using Markov Chain Model." 19th Annual EICAR Conference, Paris, France. 2010.

- [13] LEITOLD, F., A. ARROTT and K. HADARICS, "Quantifying cyber-threat vulnerability by combining threat intelligence, IT infrastructure weakness, and user susceptibility" 24th Annual EICAR Conference, Nuremberg, Germany, 2016
- [14] LEITOLD, F., A. ARROTT and K. HADARICS, "Automating visibility into user behavior vulnerabilities to malware attack" Proceedings of the 26th Virus Bulletin International Conference (VB2016), pp. 16-24, Denver, USA, 2016.
- [15] MICROSOFT. "Evolution of malware and the threat landscape - a 10-year review". 2012.
- [16] MICROSOFT. "Malicious Software Removal Tool (MSRT)". Microsoft Knowledge Base, article KB890830 revision 161.2, <https://support.microsoft.com/en-us/kb/890830>
- [17] RUBENKING N., "Why Microsoft Doesn't Need Independent Antivirus Lab Tests". PC Magazine, 28 October 2013.
- [18] SHAH P, Phatak V, Scipioni R, inventors. "Adaptive intrusion detection system." United States patent application US 10/443,568. 2003 May 22.
- [19] LEITOLD, F and K. Hadarics. "Elosztott fenyegetettség felmérés" Networkshop Konferencia, Eger, 2018.