

A KSH Könyvtár társadalomtudományi hivatkozásgyűjtő adatbázisa, a Mathias

Az elmúlt évtizedekben a tudományos teljesítmény mérésének egyik legelismertebb mutatója a (független) hivatkozások számba vétele és ezek osztályozása lett. Nemzetközi szinten, különösen a természettudományokban a hivatkozások kereshetőségét elektronikus adatbázisok segítségével igyekeznek megoldani. Ezek közül a legismertebb a Web of Science (WoS) és a Scopus, amely egyes, az angol nyelvű publikációkat megkövetelő szakterületeken hazai vonatkozásban is kiválóan alkalmazható. Az inkább magyar nyelvű közleményeket előállító tudományterületek munkatársai ma még azzal a problémával szembesülnek, hogy jóformán csak „lapozgató” módszerrel tudják a rájuk vonatkozó hivatkozásokat visszakeresni. A tudományos teljesítmény értékeléséhez ugyanakkor a társadalomtudományok területén is elengedhetetlenül hozzátartozik a hivatkozásgyűjtés. Cikkünkben hazánk első működő, szolgáltatásra kész hivatkozásgyűjtő adatbázisának létrehozásáról, a fejlesztés során felmerült problémákról számolunk be.¹

Magyarországi helyzetkép

Magyarországon egyelőre nem segítik a kutatók munkáját és értékelését a fentebb említettekhez hasonló igényű, saját fejlesztésű adatbázisok. A felsőoktatási intézmények könyvtáraiban egyenesen a *Magyar Tudományos Művek Tárába (MTMT)* rögzítik a hivatkozásokat. Abban már eltér a gyakorlat, hogy csak az oktatók által felvitt hivatkozásokat hagyják-e jóvá, esetleg maguk vállalják-e az adatok rögzítését is. (Utóbbit teszi például a *Debreceni Egyetem Egyetemi és Nemzeti Könyvtár [DEENK] Társadalomtudományi Könyvtár.*) A korábbi évekből tudunk persze a helyi kutatókat kiszolgáló adatbázis-építő kezdeményezésekről, de ezek általában tartózkodtak a retrospektív gyűjtéstől, és inkább a kurrens szakirodalom figyelésére korlátozódtak. (Kivételt jelentett az *ELTE Állam- és Jogtudományi Karának* vállalkozása az ezredfordulón.)

Hazánkban két nagyobb volumenű adatbázis-építő kísérlet ismeretes. A *Magyar Tudományos Akadémia (MTA) Nyelvtudományi Intézet Nyelvtechnológiai Kutatócsoportjában* a hivatkozások automatizált kinyerésével próbálkoznak – *Várad Tamás* egri vándorgyűlési előadása és bemutatója alapján azt mondhatjuk, figyelemreméltó sikerrel. (Az eredményeiket tesztelő *Pécsi Tudományegyetem Egyetemi Könyvtár* informatikai főigazgató-helyettese, *Markó Tamás* két vándorgyűlési előadásban ismertette a tapasztalatokat.) *Pálincás József*, a Magyar

Tudományos Akadémia elnöke 2012. augusztus végén egy 2014-ig tartó projekt elindítását jelentette be.

A humán- és társadalomtudományi hivatkozáskezelést igen kevésbé érték el a digitális világ áldásai. Pozitív példaként említhető a *Közgazdasági Szemle* és a *Tudományos és Műszaki Tájékoztatás (TMT)*: mindkét folyóirat megjelenik papíron és elektronikus formában is, és az online változat lehetőséget ad az irodalomjegyzékben való keresésre is. Az *MTMT* alkalmas ugyan a hivatkozások rögzítésére és kezelésére, a kutatóknak, illetve adminisztrátorainak azonban maguknak kell a hivatkozásokat felkutatniuk, hogy az *MTMT*-ben rögzítsék őket.

Egységes koncepció hiányában a további lemaradás elkerülése érdekében teret kell engedni egy-egy tudományterületen önálló adatbázisok kiépítésének – természetesen az együttműködés lehetőségét nem kizárva más intézményekkel, kezdeményezésekkel. Az összefogásnak ezen a területen különösen nagy fontosságot tulajdonítunk; erről a későbbiekben még lesz szó.

A Mathias létrehozása és fontosabb jellemzői

A *Központi Statisztikai Hivatal (KSH) Könyvtár* az elmúlt években egyre növekvő számú megkere-

sést kapott hivatkozások keresésére. Mivel a megrendelők zömmel magyar nyelvű és vonatkozású publikációkkal rendelkeztek, a citátumok megtalálása a különféle adatbázisok alkalmazása mellett elkerülhetetlenül manuális módszereket is megkívánt. A könyvtár egyedülállóan gazdag gyűjteménye lehetővé tette az akár évtizedes retrospektív hivatkozásokeresést. Főlegesen és kiküszöbölhetőnek tűnő terhet jelentett ugyanakkor bizonyos kiadványok (köztük a magyar statisztika vezető folyóirata, a *Statisztikai Szemle*) gyakran ismétlődő végiglapozása. Az igény és a problémák felismerése vezetett oda, hogy 2012 augusztusában a könyvtár elkezdte kiépíteni a *Magyar társadalomtudományi hivatkozások adatbázisát*, a *Mathias*, amely a magyar nyelvű társadalomtudományi folyóiratok hivatkozásait rögzíti és teszi visszakereshetővé.

A munka első fázisában a KSH kurrens, illetve nemrég megszűnt folyóiratainak feldolgozására került sor. Ezekben belül is a XX. század második felében megjelent magyar vonatkozású művekre történő hivatkozásokat rögzíti a könyvtár. Magyar vonatkozás alatt azt értjük, hogy a hivatkozott szerzők közül legalább egy magyar kötődésű. (Így került például a figyelt szerzők közé a *Kiss László* néven született *Leslie Kish* (1910–2000), aki tizenhat éves korától haláláig Amerikában élt.) Hivatkozásnak tekintünk továbbá minden műre történő utalást, jelenjen az meg lábjegyzetben, képaláírásban vagy bibliográfiában. Nem rögzíti a könyvtár az 1950 előtt megjelent művekre vonatkozó hivatkozásokat. A magunk döntően statisztikatörténeti érdeklődését a *Mathias* tervezésekor igyekeztünk háttérbe szorítani. Mindezek a megkötések az adatbázis gyakorlati alkalmazása miatt történtek.

A *Mathias* azonban nemcsak a ma aktív kutatók számára kínál lehetőséget, hanem tudománytörténeti szempontból is új forrást nyújt. Mivel a tavalyi, a KSH fennállásának 145. évfordulójára rendezett konferencia előadójától is kaptunk ilyen irányú megkereséseket, és a jövőben esedékes 150. évforduló kapcsán is számíthatunk több kiemelkedő statisztikus életművének és tudományos-szakmai hatásának vizsgálatára, a *Mathias*ba beépítettük a magyar statisztika legjelentősebb képviselőire vonatkozó hivatkozásokat is. „Melléktermékként” a XX. század első felében vagy korábban jelentős teljesítményt nyújtó tizenegy szakember (*Buday László* [1873–1925], *Jekelfalussy József* [1849–1901], *Keleti Károly* [1833–1892], *Kovács Alajos* [1877–1963], *Kőrösy József* [1844–1906], *Laky Dezső* [1887–1962], *Ráth Zoltán*

[1863–1902], *Szél Tivadar* [1893–1964], *Thirring Gusztáv* [1861–1941], *Thirring Lajos* [1899–1982] és *Vargha Gyula* [1853–1929]) tudományos utóélete is nyomon követhető segítségével. Ezzel a példával is szeretnénk érzékeltetni, hogy a KSH Könyvtár adatbázisa alapvetően az intézményi tapasztalatok alapján a könyvtárba beérkező igényekre igyekszik válaszolni.

A könyvtár munkatársai 2013 szeptemberéig a fenti szempontok alapján feldolgozták a *Statisztikai Szemle* 1980 és 2013, a *Demográfia* 2000 és 2012, a *Területi Statisztika* 1998 és 2013, illetve a *Gazdaság és Statisztika* 1992 és 2006 közötti évfolyamait. Ezzel a mai napig összesen tizenötezer rekord vált kereshetővé. A *Statisztikai Szemle* 1923 óta – a háborús időszakot leszámítva – havi rendszerességgel megjelenő lektorált folyóirat, a magyar statisztikatudomány legfontosabb, az *MTA IX. Osztálya* által A-kategóriásnak nyilvánított fóruma. Az itt megjelent tanulmányok lefedik a társadalomtudományok, azon belül is kiemelten a statisztika, a közgazdaság-tudomány és a demográfia egyes ágait, illetve betekintést nyújtanak az egyéb tudásterületek statisztikai alkalmazásaiba. Az 53. évfolyamában járó *Területi Statisztika* nemcsak a szűkebb értelemben vett statisztika, hanem a regionális tudomány valamennyi területéről közöl cikkeket. Az 1957-ben alapított *Demográfia* a *KSH Népeségtudományi Kutatóintézetének* tudományos folyóirata. A *Gazdaság és Statisztika* pedig 2006-os megszűnéséig közel hat évtizeden át az ipar- és gazdaságstatisztikusok elsőszámú publikációs fóruma volt.

A feldolgozandó dokumentumok körének kijelölése után az adatbázis szerkezetének megtervezése jelentette a következő feladatot. Már a kezdet kezdetén eldöntöttük, hogy ragaszkodunk a „keep it simple and stupid” elvéhez, azaz a lehető legegyszerűbb megoldást választjuk, és mind az adatfeltöltéshez, mind a lekérdezéshez táblázatkezelő programot fogunk használni. Tettük ezt többek között azért, mert a *Mathias* építésének első évében, a tesztüzem idején nem számoltunk néhány tízezerrel több rekorddal. Ekkora mennyiség kényelmesen kezelhető akár *Excel* is. Az egyszerűen exportálható és konvertálható rekordok létrehozása mellett szót az is, hogy ez később jócskán megkönnyítheti az együttműködést más adatbázisgazdákkal. Nem követhetjük ugyanis a „mindenki másként szabványos” elvét, nem szaporíthatjuk főlegesen a formátumok számát, ha nem csupán szétszórt intézményi, hanem országos közös hivatkozási adatbázisban gondolkodunk.

Egy-egy rekord értelemszerűen két részre oszlik: egyfelől a hivatkozó, másfelől a hivatkozott mű leírására. A hivatkozott művek sokszínűsége viszonylag gyorsan felülírta az eredeti terveket: jelenleg összesen huszonhárom mező áll rendelkezésre egy-egy tétel rögzítéséhez. Az első elképzeléshez képest új mezőt kaptak az eltérő szerzői minőségek (szerkesztők-összeállítók és szerzők), illetve külön cellákban szerepelnek a konferencia-előadásokra történő utalások. Bár sokáig vívódunk, külön mezőkbe soroljuk-e a társszerzőket, végül – meglehetősen vitatható módon – csupán írásjelekkel választottuk el őket egymástól. A névanyag egységesítése során ugyanazokat a problémákat tapasztaljuk, amelyekkel más adatbázisokban, könyvtári katalógusokban is nap mint nap szembesülhetünk.

A szerzőség megállapításakor nehézséget okoznak a KSH 1989 előtt megjelent összeállításai. Ezeknél szerzőségi közlésként többnyire csupán a hivatal, esetleg egy-egy főosztály megnevezése szerepel. Az előszóból és további rejtett adatokból azonban olykor kideríthető az összeállító, szerkesztő személye. Ebből a megfontolásból a KSH kiadványaira történő hivatkozások minden esetben szerepelnek a *Mathias*ban. A későbbiekben szükség lesz ezeknek a rekordoknak a folyamatos ellenőrzésére, javítására, bővítésére, könyvtári katalógussal való összevetésére. A KSH-kiadványok szerepeltetése a gyakorlati alkalmazást is szolgálja: a megrendelő ugyanis a publikációs jegyzékében szerepeltetheti azokat az adattárakat, amelyek létrehozásában részt vett, azonban csak az előszóban említik meg közreműködését.

Természetesen számos egyéb problémával is találkozunk az adatrögzítés során. Több folyóiratnál is megfigyelhető gyakorlat, hogy akár egyazon számon, de egyetlen cikken belül is kétféleképpen hivatkozik. Egyes szerzők az adattárakat eleve nem tekintik hivatkozásra érdemesnek. Gyakran előfordul, hogy egy táblázat vagy egy ábra alatt feltüntetik az összeállítás címét, de pontos bibliográfiai leírást nem mellékelnek hozzá, és az irodalomjegyzékbe sem veszik fel. A kiadvány címe helyett több esetben a sorozatcím és a kötettség szerepel. Néhányszor a hivatkozott periodikum nem szokványos megjelenési adataival gyűlt meg a bajunk: a *Medvetánc* 1986. 4. – 1987. 1. összevont számát például a legnagyobb jóindulattal sem tudtuk kielégítő módon megjeleníteni az adatbázisban.

Ugyanígy fejtörést okozott számunkra már a *Mathias*ba felvitt második rekord is. Az éppen feldolgozott cikk bibliográfiájában egyebek mellett *Bibó István* válogatott tanulmányainak a Magvetőnél 1986-ban napvilágot látott harmadik kötete is szerepelt, akkurátusan közölve a mű megjelenési adatait és terjedelmét. A szövegben azonban csak a 7–8. oldalról idézett a szerző, vagyis nem a gyűjteményes kötetre, hanem *Az európai társadalomfejlődés értelme* című írásra kellett volna hivatkoznia.

A *Mathias* egyelőre csak a megkereséseket kezelő könyvtárosok számára elérhető. A hivatkozásgyűjtést a KSH és intézeteinek munkatársai számára ingyenesen, külső megrendelőknek térítésért végezzük.

Terveink szerint a következő hónapokban a már említett folyóiratok kurrens és régebbi számainak feldolgozása mellett megkezdjük a statisztikai-demográfiai tárgyú kutatási jelentések, monográfiák, Festschriften, tanulmány- és konferenciakötetek hivatkozásanyagának rögzítését is.

Összefoglalás

Cikkünkben igyekeztünk felvázolni az első magyar hivatkozásgyűjtő adatbázis, a *Mathias* létrejöttének körülményeit és működésének alapelveit. Mivel ezen kívül egyetlen hasonló adatbázis tartalmáról, „képességeiről” sincsenek pontos ismereteink, ma még nem látszik világosan, milyen együttműködések alakulhatnak ki a jövőben. A KSH Könyvtár mindenestre egyetlen megkeresés elől sem zárkózik el, amely a tartalomszolgáltatói szféra erőforrásainak lehető leghatékonyabb felhasználását célozza.

Megjegyzés

- ¹ A cikk a szerzők 2013. március 27-én, a Networkshop 2013 *Tartalomszolgáltatók: könyvtárak, levéltárak, múzeumok* szekciójában azonos címmel elhangzott előadásának szerkesztett és átdolgozott változata. Az előadás megtekinthető a Videotoriumban a http://videotorium.hu/hu/recordings/details/6176.A_KSH_Konyvtar_tarsadalomtudomanyi_hivatkozasgyujto_adatbazisa_a_Mathias linken. Köszönjük CSEH Gabriellának a cikk megírásához nyújtott segítségét; DUDÁS Anikónak a *Mathias* iránt tanúsított rendkívül szíves érdeklődését, a társadalomtudományi hivatkozás-kereséssel kapcsolatos észrevételeit és szakirodalmi ajánlatait; valamint PERGÉNÉ SZABÓ Enikőnek a *DEENK Társadalomtudományi Könyvtár* hivatkozásgyűjtő gyakorlatáról szóló szíves közlését.

Irodalom

- [1] DUDÁS Anikó: Hivatkozások a bölcsészet- és társadalomtudományokban: MTMT – tudományos követelményrendszerek – bibliográfiai szempontok. = Networkshop, 2012. április 11. Veszprém.
http://videotorium.hu/hu/recordings/details/4099,Hivatkozasok_a_bolcseszet-es_tarsadalomtudomanyokban_
- [2] DUDÁS Anikó: Hivatkozásokra vezérlő kalauz. = Networkshop, 2013. március 27. Sopron.
http://videotorium.hu/hu/recordings/details/6224,Hivatkozasokra_vezerlo_kalauz
- [3] HUNYADI László – VISI LAKATOS Mária: A Statisztikai Szemle kilencven éve. = Statisztikai Szemle, 90. köt. 11–12. sz. 2012. p. 1023–1035.
- [4] LENCSE S Ákos – RÓZSA Dávid: A statisztika és a digitális átállás. = Magyar Könyvtárosok Egyesülete 44. Vándorgyűlés, 2012. július 12–14. Győr.
https://videotorium.hu/hu/recordings/details/5210,A_statisztika_es_a_digitalis_atallas
- [5] LIESZKOV SZKY József Pál: A Területi Statisztika citációs listája, 2002–2006. = Területi Statisztika, 10. köt. 4. sz. 2007. p. 380–384.
<http://www.ksh.hu/docs/hun/xtmp/terstat/2007/terstat0704.pdf>
- [6] MARKÓ Tamás: Adatbázisba kerülnek a magyar szakfolyóiratok cikkeinek hivatkozásai is? = Magyar Könyvtárosok Egyesülete 44. Vándorgyűlés, 2012. július 12–14. Győr.
https://videotorium.hu/hu/recordings/details/5199,A_datbazisba_kerulnek_a_magyar_szakfolyoiratok_cikkeinek_hivatkozasai_is_
- [7] MARKÓ Tamás: Magyar humán- és bölcsészettudományi folyóiratok hivatkozásainak adatbázisa – egy kísérleti projekt tesztelésének tanulságai. = Magyar Könyvtárosok Egyesülete 45. Vándorgyűlés, 2013. július 18–20. Eger.
- [8] MURÁNYI Péter: A magyar cikkanyag a nemzetközi adatbázisokban. = TMT, 52. köt. 7–8. sz. 2005. p. 332–346.
http://tmt.omikk.bme.hu/show_news.html?id=4009&issue_id=464
- [9] SZABÓ István: A Statisztikai Szemle statisztikája. = Statisztikai Szemle, 86. köt. 12. sz. 2008. p. 1158–1162.
http://www.ksh.hu/statszemle_archive/2008/2008_12/2008_12_1158.pdf
- [10] VÁRADI Tamás: MATRICA – Magyar humán- és bölcsészettudományi folyóiratok hivatkozási adatbázisának építéséről. = Magyar Könyvtárosok Egyesülete 45. Vándorgyűlés, 2013. július 18–20. Eger.
- [11] VÁRADI Tamás – PINTÉR Tibor – MITTELHOLCZ Iván – PEREDY Márta: Bibliográfiai hivatkozások automatikus kinyerése. = TANÁCS Attila – VINCZE Veronika, szerk.: VII. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, 2010. p. 56–68. ISBN 978 963 306 075 9
<http://www.inf.u-szeged.hu/mszny2010/mszny2010.pdf>
- [12] VIDA Ildikó: Kísérleti hivatkozási index összeállítása egy jogi könyvtárban. = Könyvtári Figyelő, 49. köt. 3. sz. 2003. p. 555–559.
<http://epa.oszk.hu/00100/00143/00046/vida.html>

Beérkezett: 2013. IX. 4-én.

