

Tudományos könyvtárak folyóirat-állományának idézetelemzéses értékelése automatizált számítógépes eljárással

A világon először készült olyan automatikus számítógépes könyvtárértékelő rendszer, amely a könyvtár folyóiratait tematikai csoportokra osztja, majd mindegyik mellett felsorolja a témának az illető folyóiratcsoport által meghatározott magfolyóiratait rangsorba állítva, tekintet nélkül arra, hogy megtalálhatók-e a könyvtárban, vagy nem. A módszer lelke az idézetelemzés.

Bevezetés

A tudományos folyóiratok a tudományos információ kiemelkedően legfontosabb, az alapkutató-sokban abszolút hordozói. Számukat tízezrekben, árukat millió dollárokban mérik. Még a gazdag országokban sem közömbös tehát, hogy milyen a hatékonysága az előfizetéseknek, valóban a legfontosabb folyóiratok járnak-e a könyvtárba. Magyarországon a tudományos könyvtárakban a koncepciózus állománygyarapítás ritka kivételnek számít [1], ezért különösen fontos szerepe lehet egy többé-kevésbé objektív, számokkal operáló értékelő rendszernek.

Egy-egy téma, tudományterület magfolyóiratainak kiválasztására régtől fogva sok kísérlet történt. Mára az idézetelemzés vált a legelterjedtebb módszerrel [2], bár az idézés motívumai egyáltalán nem csak a szakmai relevanciára korlátozódnak [3]. Nem ismerünk azonban olyan eljárást, amely egy könyvtár állományából kiindulva értékelné folyóiratainak tematikai összetételét, mégpedig úgy, hogy e tematikát is a könyvtárba járó folyóiratok segítségével állapítja meg, s az azonosított szakterületek magfolyóiratait a vizsgált állomány figyelembevételével készített hierarchikus rangsorba állítva mutatja ki: melyek az előfizetett, illetve a hiányzó folyóiratok.

Cikkünkben egy ilyen módszert ismertetünk, annak első kipróbálási eredményeivel együtt.

Anyag és módszer

A vizsgált könyvtár folyóiratait tematikai csoportokba soroljuk be, az őket bizonyos küszöböt meg-

haladó mértékben idéző folyóiratok segítségével létrehozott és mért kapcsolataik alapján. Az egyes témák fő folyóiratai az e csoportok tagjait idéző folyóiratok lesznek. Rangsorolásuk az idézési kapcsolat erőssége és az abszolút idézettséget kifejező *impact factor* alapján történik.

Az első elemzést az MTA Szegei Biológiai Központ (SZBK) Könyvtárának állományán végeztük. Kiderült, hogy az ott gyűjtött kísérletes biológiai nagyon jól működik a módszer, s erre alapozva joggal várhatjuk, hogy általában a természettudományokban és az úgynevezett kemény társadalomtudományokban is sikerrel lehet majd alkalmazni.

A módszer lépései

1. A mintafolyóiratok kiválasztása

A könyvtár állományából kiválasztottuk az SZBK kutatási profilja szempontjából releváns, speciális tematikájú primer tudományos folyóiratokat. Ezeket nevezzük a továbbiakban *saját* folyóiratoknak. Az SZBK 150 külföldi tudományos folyóiratából 106 volt ilyen. Kimaradtak tehát a szemle-folyóiratok, az általános témájúak, és a nem szoros témába vágó (kémiai, számítástechnikai stb.) folyóiratok.

2. Az idéző folyóiratok kiszűrése

Az idézési kapcsolatok korábbi vizsgálataiban általában a bázisfolyóiratok által idézett folyóiratokat használták, mi azonban megfigyeltük, hogy az őket idéző folyóiratok segítségével jobb differenciálást és szélesebb spektrumot lehet megvalósítani.

Minden *saját* folyóiratnak kiválasztottuk a legfontosabb idéző folyóiratait a *Science Citation (SC)* legutóbbi (vizsgálatunk kezdetekor az 1992-

es) *Journal Citation Reports (JCR)* kötetéből. A kiválasztás során figyelembe kellett venni néhány körülményt a jobb eredmény érdekében:

- Az idéző folyóiratok cikkszámú széles skálán oszlik el. A nagyobb folyóiratok tehát nagyobb eséllyel kerülnek be az idézők közé. Erre nézve azonban nem végeztünk korrekciót, ugyanis ha egy folyóiratban sok releváns cikk van, akkor nem érdekes, hogy mennyi egyéb van mellette.
- Ki akartuk viszont küszöbölni a *saját* folyóiratok eltérő korából adódó összehasonlítási nehézségeket. Ezért csak a 2 és 3 éves cikkekre eső idézeteket vettük figyelembe (ezek száma a legmagasabb a koreloszlási görbe mentén).
- Ezzel az intézkedéssel fokoztuk a tematikai relevanciát is, mivel a fiatalabb cikkek esetében ez a fő idézési motívum [4].
- Valamely *saját* folyóirat *idéző* folyóiratai közé azok kerültek be, amelyek az illető *saját* folyóirat 2 és 3 éves cikkeire kapott idézeteinek legalább 1,5%-át produkálták.

Az idézetszámokat figyelmen kívül hagyva *saját* folyóirataink és a kiszűrt *idéző* folyóiratok címeit bevittük az automatikus értékelést végrehajtó programba.

3. Társfolyóirat-kapcsolat megállapítása a saját folyóiratok között

Két *saját* folyóirat akkor *társa* egymásnak, ha idéző folyóirataik összegének legalább egyötödét eléri a közös idézők száma. A 106 SZBK-s *saját* folyóiratból 13-nak nem lett társa.

4. Partnerkapcsolat megállapítása a saját folyóiratok között

Két *saját* folyóirat akkor *partner*, ha egyikük társfolyóiratainak legalább 80%-a megtalálható a másik társai között is. (Értelemszerűen a kisebb társzámú folyóiraatra vonatkozik a 80%.) A társall rendelkező folyóiratok mindegyikének lett legalább egy partnere. Erre a lépésre azért volt szükség, hogy egyrészt tovább szűküljenek és specializálódjanak a kapcsolatok, másrészt pedig, hogy e szinten teljesen a *saját* folyóiratok váljanak a kapcsolódás meghatározóivá.

5. Elsődleges témacsoportok létrehozása a saját folyóiratokból

- A partnerrel rendelkező *saját* folyóiratokat sorba állítottuk a partnerek száma szerint.
- Az első egypartneres folyóiratot csoportkezdővé minősítettük, s hozzárendeltük partnerét.
- Tovább haladva felfelé a sorrend mentén, mindazon folyóiratokból csoportkezdő lett, amelyek addig nem szerepeltek valamely korábban létrehozott csoportban partnerként. Az *elsődleges témacsoportok* tehát valamely

(részben véletlenül) csoportkezdővé vált folyóiratból és egy vagy több partnerből állnak.

- Csoporttaggá többször is válhattak a folyóiratok aszerint, hogy partnereik közül hány lett csoportkezdő.
- Ez az eljárás értelemszerűen valamennyi partnerrel rendelkező *saját* folyóiratot besorolta legalább egyszer az összesen 23 *elsődleges témacsoport* valamelyikébe.

6. Másodlagos témacsoportok létrehozása összevonással

Egyes témacsoportok nem, mások viszont nagy átfedésben állnak egymással. Hogy egyrészt csökkentjük a csoportok számát, másrészt pedig fokozzuk a szakterületek szerinti szétválást, sorra vettük az elsődleges témacsoportokat, és mindegyiket összevontuk azzal, amelyikre nézve a legnagyobb átfedés állt fenn. Az összevonáshoz az átfedésnek legalább 50%-ot kellett elérnie az egyik csoportra nézve. Így összesen 14 *másodlagos témacsoport* maradt.

7. Harmadlagos témacsoportok létrehozása új csoportkezdő folyóiratokkal

Mindegyik *másodlagos témacsoport* tagjai közül kiválasztottuk azt a kettőt, amelynek partneri köre a legnagyobb százalékos átfedésben állt valamely csoporttársának partneri körével, miközben a két folyóirat együttes partnerszámát 100%-nak vettük. E párok lettek az új csoportkezdők, partnereikkel együtt alkotva a *harmadlagos csoportokat*. Számuk értelemszerűen megegyezik a másodlagos témacsoportokéval.

8. A harmadlagos témacsoportok feltöltése

Néhány partnerrel rendelkező *saját* folyóirat kimaradt a harmadlagos csoportokból. Oda soroltuk be őket, ahol a csoport tagjainak partnerfolyóiratai közül a legtöbb szerepelt a kérdéses folyóirat partnerei között.

9. Többes előfordulások kiszűrése

Szelekciót végeztünk abból a célból, hogy minden *saját* folyóirat csak egy témacsoportban legyen tag. Ezt úgy valósítottuk meg, hogy a több *harmadlagos témacsoportban* szereplőkről eldöntöttük, hogy melyik csoport vonzza őket legjobban. A vonzás mértéke az a százalékkérték lett, amelyet az illető csoport tagjai közül a vizsgált folyóiraatra nézve partnerfolyóiratok tettek ki a csoport összes folyóirata közül.

10. A végső témacsoportok kialakítása összevonással

Összevontuk azokat a harmadlagos témacsoportokat, amelyek tagjainak partnerfolyóirat csoportja valamely másik hasonló csoporttal több mint

80%-os átfedést mutatott. Így alakult ki a 12 végső témacsoport. Tagszámuk 2 és 18 között váltakozott.

11. A témacsoportok elnevezése

A tagfolyóiratok tematikája egyértelműen eldöntötte a csoport témáját.

12. A témacsoportok magfolyóirat-rangsorai

A témacsoportok tagjai között eddig csak saját folyóiratok voltak. Vizsgálatuk fő célja azonban nem saját folyóirataink tematikus besorolása volt, hanem az, hogy a megállapított témák rangsorolt magfolyóirat-listáit hozzuk létre, tekintet nélkül arra, hogy azok megtalálható-e az SZBK-ba járók között, vagy nem.

A magfolyóiratok a témacsoportok tagjait idéző folyóiratok közül kerülnek ki. Rangsorolásuk két szempont alapján történt:

- a fő rangsorolási szempont az volt, hogy az illető folyóirat hány csoporttag idéző folyóiratai között szerepelt;
- a másik, az azonos számú csoporttagot idéző folyóiratok sorrendjét eldöntő mérőszám a folyóiratok impact factora volt.

Magfolyóirattá minősítettük az idéző folyóirat rangsorok első harminc tagját minden témacsoportban. Tudjuk, hogy az egyes diszciplínák folyóiratszámát nem azonos (akár hússzoros különbségek is előfordulnak), szem előtt kell azonban tartanunk, hogy a cikkek korántsem csak a szűken vett témába teljesen beletartozó folyóiratokban jelennek meg. Ilyen módon nem is olyan nagy a különbség a témák releváns folyóiratszámait között, körülbelül egy kettes faktor van az idéző folyóiratok két szélső értéke közt.

Eredmények és megvitatás

A 106 saját folyóiratnak összesen 835 idéző folyóirata volt. Az SZBK Könyvtár 150 előfizetett külföldi folyóirata közül 142 szerepelt köztük. A kimaradtak többsége kevés cikk és/vagy nem rendszeresen megjelenő szemle-folyóirat. A magfolyóiratok közé 102 SZBK folyóirat került be, közülük 29 nem volt a saját folyóiratokból álló témacsoportok tagja.

A szelekció tehát eredményesen működött, széthúzta a mezőnyt, teret engedve az átgondolásnak, a célul kitűzött revízióknak.

A 12 témacsoportot és magfolyóirataik előfizettségét az 1. táblázatban találjuk. Megjegyezzük, hogy az SCI JCR is készített tematikus listákat referált folyóiratokról, nem választja azonban külön a biokémiát a molekuláris biológiától, holott esetünkben még a molekuláris biológián belül is két csoportot azonosítottunk. Ennek oka a szelekciós

eljárás érzékenysége volt, nem pedig a folyóiratpark csonkaságaiból eredő műtermék-keletkezés, hiszen e csoportok a legjobb előfizettségűek közé tartoznak az SZBK-ban.

1. táblázat

Az egyes szakterületek első harminc folyóiratának előfizetési adatai

Diszciplína	Előfizetések száma helyezés szerint			
	1-10	11-20	21-30	1-30
Molekuláris biológia	10	9	8	27
Molekuláris fejlődésbiológia	10	8	6	24
Biofizika	10	8	6	24
Növénybiológia	9	8	7	24
Biokémia	9	4	5	18
Gyógyszertan	6	8	4	18
Genetika	7	5	4	16
Virologia	6	3	6	15
Immunológia	4	3	6	13
Neurobiológia	4	3	3	10
Rovarbiológia	5	3	1	9
Biotechnológia	3	1	0	4
Összesen	83	63	56	202

Látható ugyanakkor, hogy a 12 témacsoport magfolyóiratainak előfizetettsége korántsem egyenletes, a nagyon jótól a nagyon gyengéig terjed.

Az előfizetettségre rányomja a bélyegét az a körülmény is, hogy az SZBK-ban elsősorban molekuláris és sejtbiológiai szintű kísérletes alap kutatás folyik. A hiányzó folyóiratok többsége orvosi, ökológiai, agrártudományi, technikai témájú.

Az irodalom erős szóródását mi sem mutatja jobban, mint hogy egyes alacsony saját folyóirat-számú témacsoportok magfolyóirat-listájának előfizetettsége korántsem oly szegényes. Csak egy példa: a szorosan vett virológiát mindössze három folyóirat képviseli az SZBK-ban. Ezek alkották a végső témacsoportot is e területen. A csoport első harminc magfolyóiratából azonban 15-öt fizet elő az SZBK, azaz a legfontosabb virológiai érdekes-ségű cikkek többségét nem a szűk szakterület folyóiratai közlik. Elhamarkodott lett volna rossznak ítélni az SZBK virológiai folyóiratparkját azon három folyóirata alapján, amely címében hordja a vírus szót, bár a virológia témacsoportba csak ezek kerültek be.

Ellenkező irányú kivételre is számos példa van, arra ti., hogy valamely szorosan diszciplinárisnak tűnő folyóirat nem abba a csoportba tartozik, ahová címe alapján besorolni vélénk. Erre a növénybiológia témacsoportban találhatunk könnyen értelmezhető példát: a *Protoplasma*, a *Theoretical and Applied Genetics* és a *Zeitschrift für Naturforschung* idézési kapcsolataik alapján egyértelműen növénytani folyóiratok elsősorban, jellelhet ez címeikből nem állapítható meg, sőt mivel

bizonyos mértékben mégiscsak interdiszciplinárisak, állandó veszélyben vannak a folyóirat-revíziók alkalmából, hiszen senki sem tudja ilyenkor, kinek is kellene a legjobban.

A rangsorolt magfolyóirat-listákból kiderülnek a hiányzó folyóiratok is, és az esetleg lemondásra megfontolandók is. Hogy a gyengén képviselt diszciplínákat leépíteni vagy feljavítani kell-e, esetleg éppen a jó előfizettségük rovására, azt tudománypolitikai megfontolások alapján kell eldönteni.

További segítséget jelenthet a kérdéses folyóiratok előfizetéséhez vagy lemondásához az a lista, amelyben aszerint állítottuk sorba az idéző folyóiratokat, hogy hány témacsoporthoz kapcsolódnak ilyen minőségben. A megoszlást a 2. táblázatban találjuk. A számokból látható, hogy a lefutás igen erősen exponenciális, az összes idéző folyóirat 76%-a csak egyetlen témacsoporthoz kötődik (többnyire gyengén). Az előfizettség – a várakozásnak megfelelően – a multidiszciplinaritással párhuzamosan nő.

2. táblázat

Az idéző folyóiratok megoszlása a témacsoportokhoz való kapcsolódásuk száma szerint

Kapcsolódás-szám	Folyóirat	SZBK-előfizetés	Előfizettség %-a
9-12	11	11	100
5-8	49	30	61
3-4	140	46	33
1-2	635	55	9

Pontosíthatjuk a könyvtárról alkotott képet azzal is, ha megvizsgáljuk az egyes témacsoportok egymással fennálló kapcsolatait. Erre egy igen egyszerű módszert használtunk fel. A harmincas magfolyóirat-listák alapján minden csoportról eldöntöttük, hogy melyik két másik csoporthoz kötődik legerősebben, egyszerűen annak alapján, hogy hány közös magfolyóiratuk van. Ha az azonosságok miatt nem egyértelmű a döntés, akkor a kölcsönösséget hívtuk segítségül, azt ti., hogy a valamely csoporthoz (az illető csoport felől nézve) azonos mértékben kapcsolódó csoportok közül melyiknél hányadik a kapcsolódási rangsorban a kérdéses csoport. (Kettőnél több csoport közül sohasem kellett választani.) A kapcsolódás lehet erős és gyenge. Ha két csoport magfolyóiratai között tíz vagy több azonos volt egymással, akkor a kapcsolódást erősnek minősítettük.

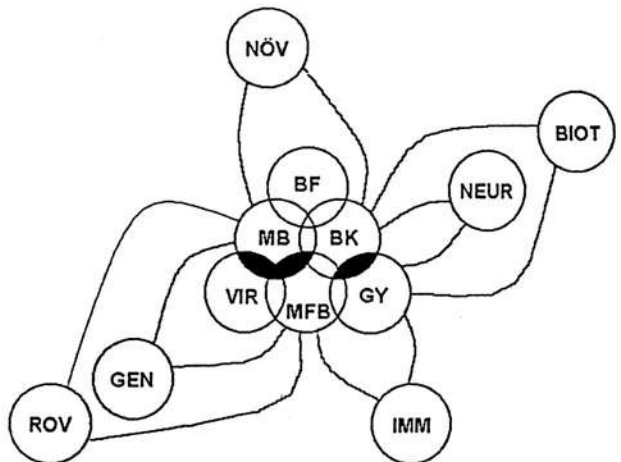
E kritériumok két félre osztották magfolyóirat-csoportjainkat. Hat olyan akadt, amelynek erős kötődése van mindkét partneréhez, és hat olyan, amelynek csak gyenge. Ez utóbbiak partnerei mindig az erősen kötődő csoportok közül kerülnek ki, egymáshoz sohasem kapcsolódnak. A leírtakat

az 1. ábrán szemléltetjük. Az egymással átfedő körök erős, a karokkal való összeköttetések gyenge kapcsolatot tartanak. A besatírozott átfedő területek kölcsönösséget jelölnek, ezekben az esetekben mindkét csoport egymás első két partnere között szerepel. A magon belül is kitüntetett szerepe van a két molekuláris biológiai csoportnak, a biokémiának és a gyógyszerertannak. A biofizika és a virológia ugyan erősen kötődik az előbb említett csoportokhoz, külső kapcsolatuk azonban nincs.

A kötődés erős vagy gyenge volta nem függ a témacsoportot alkotó saját folyóiratok számától. A növénybiológia 18 tagú témacsoportja ellenére is külön áll, a virológia viszont erősen kötődik, pedig témacsoportja mindössze három tagot számlál.

Ne feledjük, a listák és a viszonyulások az SZBK helyzetét (is) tükrözik. Jó példa erre a mikrobiológia esete. Az SZBK mikrobiológiai folyóiratait a molekuláris biológia témacsoport foglalta magában. Ennek fő oka, hogy az SZBK-ban alaputatás folyik, és az e szempontból fontos mikrobiológiai folyóiratokban komoly részt kapnak a molekuláris biológiai vonatkozások. Ha lennének az SZBK-ban orvosi, ipari, vagy biotechnológiai mikrobiológiai folyóiratok, akkor valószínűleg külön témacsoportot hoztak volna létre együtt. Az SZBK kutatási témáit megvizsgálva mindenestre egyértelműen kiderül, hogy ez a besorolódás megfelel a tematikai igényeknek.

Talán még érdekesebb a molekuláris fejlődés-biológiai csoport létrejötte. Tíz évvel ezelőtt még biztosan nem szakadt volna külön, az utóbbi évek nagy fellendülése azonban lehetővé tette a kvantitatív elkülönítést.



1. ábra A témacsoportok kapcsolatai (Rövidítések az 1. táblázat alapján.)

Az észlelt jelenségek figyelmeztetnek is: nem szabad a verbális azonosságok szintjén maradni, a diszciplínák állandó mozgásban vannak, statikus kövületként való felfogásuk tévedés. Ha a könyvtá-

rákban ez a felfogás érvényesül, pazarlást okoz a kiadásokban, s ugyanakkor kihasználatlanságot az állományban.

Magyarországon a szegénység és a szétszórtság határozza meg a tudományos könyvtári állományokat. Mivel – mint említettük – a koncepciótlanság is jellemző gyakorlatunkra, a téveszmék okozta károktól nem kell még komolyabb mértékben tartani. Ilyennel is lehet persze találkozni, elsősorban az országos tudományos könyvtárakban és a központi könyvtárakban, ahol az „innen is egy kicsit, onnan is egy kicsit” a beszerzési irányelv. Az eredmény lesújtó: a legnagyobb könyvtárak diákok melegezőhelyei, állományuk vérszesen kihasználatlan. Lehet ebben szerepe a tudomány iránti érdeklődés apályának is, ami nem a könyvtárosok, hanem a tudománnyal hivataltól-hivatásból foglalkozni köteles személyek (és vezetők) szégyene, de ahogy mondani szokás, „ez minket nem vigasztal”.

Miben segíthet a könyvtárosoknak egy ilyen elemzés?

- Megmutatja, hogy milyen tényleges szakterületek irodalmát gyűjti a könyvtár.
- Az állományból kiindulva felhívja a figyelmet a szakterületek kapcsolataira, részágazataik hangsúlyaira.
- Pontos képet ad az egyes szakterületek ellátottságáról.
- Megnevezi a hiányzó folyóiratokat szakterületenként.
- Megnevezi a felülvizsgálandó előfizetéseket.
- Kimutatja, hogy melyek a multidiszciplináris folyóiratok, és azt is, hogy ezek a könyvtár mely gyűjtött szakterületei számára fontosak.
- Segíti tudománypolitikai alapokra helyezni az állománygyarapítást.
- Megmutatja, hogy mely részterületek változtatása érint más szakterületeket is, és azt is, milyen mértékben.
- Alternatívát kínál az optimalizálásra a legkülönbözőbb irányokban (leépítés, szintentartás, növekedés), jelezve a változtatásoknak a könyvtár egészére kiható következményeit.
- A változtatások fokozataira nézve is felvilágosítást ad az optimalizálást, illetve a következményeket illetően.

Mit tehet hozzá a könyvtáros az értékelés eredményeihez?

- Tudja, hogy valamely előfizetésre javaslandó folyóirat járt-e korábban a könyvtárba.
- Tudja, hogy jár-e valamely jól szolgáltató partnerkönyvtárba.
- A lemondásra javasolhatókról is tudja mind ezeket.
- Tudja, hogy mely könyvtárakat lehet felkeresni esetleges együttműködésre kölcsönös előnyök alapján az előfizetések esetleges megsztása céljából.
- Felderíti azokat a folyóiratokat, amelyeket esetleg cserére lehet javasolni a saját állományból az új, fontosabbnak bizonyultakkal szemben.
- Felderíti az árakat.
- Felderíti a cikkszámokat.
- Felvilágosítást ad a megkérdőjelezett folyóiratok használatáról.

Az ismertetett módszer természetesen nemcsak könyvtárértékelésre, hanem bármely téma, szakterület magfolyóiratainak kiválasztására, valamint nagyfelbontású tudományterképezésre is jól alkalmazható.

Irodalom

- [1] MARTON J.: Tudósok tudása: természettudományi információellátás Magyarországon. = Az MTA Könyvtár Közleményei, 30. köt. 1992. p. 181–187.
- [2] LEYDESDORFF, L.: The generation of aggregated journal-journal citation maps on the basis of the CD ROM version of the Science Citation Index. = Scientometrics, 31. köt. 2. sz. 1994. p. 59–84.
- [3] LIU, M.: The complexities of citation practice: a review of citation studies. = Journal of Documentation, 49. köt. 3. sz. 1993. p. 370–408.
- [4] MARTON, J.: Causes of low and high citation potentials in science. = Journal of the American Society for Information Science, 31. köt. 4. sz. 1983. p. 244–246.

Beérkezett: 1995. IV. 4-én.



OMIKK FORDÍTÓ IRODA

1088 Budapest, Múzeum u. 17.
Telefon: 138-2874 Fax: 138-4924

- MŰSZAKI FORDÍTÁS, LEKTORÁLÁS
- TOLMÁCSOLÁS, SZÖVEGSZERKESZTÉS
- SZAKIRODALOM BESZERZÉS
- 30 NYELV, ORSZÁGOS HÁLÓZAT, REÁLIS ÁRAK