

HOZZÁSZÓLÁS

Tolnai Márton *Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében* című cikkéhez¹

Kollár István

Michelberger Pál

egyetemi tanár,
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Villamosmérnöki és Informatikai Kar
kollar@mit.bme.hu

az MTA rendes tagja, egyetemi tanár,
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Közlekedésmérnöki Kar

A *Magyar Tudomány* augusztusi számában érdekes cikk jelent meg az MTA osztályaihoz kapcsolódó magyar tudósok egyes tudomány-metriai adataival (Tolnai, 2008). A cikk még biztosan sok vitát fog kiváltani, különösen, hogy a benne leírt adatok háttere, forrása és feldolgozási módja sokak számára nem is ismert. Ráadásul az átlagos olvasó amúgy is nehezen tájékozódik az adatbázisok, mérőszámok és adatok között. Ez a hozzászólás ebben kíván kiegészítő információt adni „laikusok”, vagyis tudományometriával hivatásszerűen nem foglalkozók számára: egyáltalán mit jelentenek az adatok, és főleg *mit nem jelentenek*. Aki esetleg máris bosszankodik, ezek alapján talán kevésbé lesz erre indíthatva.

Előrebocsátjuk, hogy a publikációs tevékenység átláthatóvá tételével mélysegesen egyetértünk. Tudományos életünk nagy fegyverténye lenne, ha a PhD-fokozatot vagy doktori címet szerzők előzetes publikációs

tevékenységét egyformán, tárgyilagosan szemlélni lehetne, és az is sokat segítené, ha a fiatalok látnák az idősebbek listáit. Ezért a Köztudományi Publikációs Adattár (KPA) és általában a mycite² léte önmagában is nagy dolog, és azzal, hogy klónjai (MTA Kísérleti Orvostudományi Kutatóintézet – KOKI, Budapesti Műszaki és Gazdaságtudományi Egyetem – BME, MTA Szegedi Biológiai Központ – MTA SZBK, Mezőgazdasági Biotechnológiai Kutatóközpont – MBK) lényegében ezzel azonosak és többé-kevésbé átjárhatók is,³ az első lépést máris megtettünk.

² A KPA és klónjai a „mycite” program továbbfejlesztései. Ez egy mysql alapú adatbázis és egy php alapú webes kezelőprogram együttese. A továbbiakban összefoglalóan mycite programként hivatkozunk rájuk. Ezek: a KPA, és a következő intézmények adattára: MTA KOKI, BME, MTA SZBK, MBK, valamint hasonló az SE adattára is, de sajnos kissé eltérő.

³ Szerzők szintjén az egyik adattár be tudja emelni keresőjébe a másikban szereplő szerzőt, sőt, összesítő adatok is lekérhetők, és a fő publikációkategoriók is megegyeznek. Publikációk szintjén (például társszerzők hozzárendelése) ez sajnos még nem működik.

¹ A cikk a lap 2008. augusztusi számának 976–988. oldalain jelent meg.

Egyelőre sajnos csak álom, hogy a nagyobb magyar egyetemek és kutatóintézetek mind ugyanígy összeköthető adatbázis(oka)t használjanak, pedig ez lenne a közérdek. Még az MTA Tudományos Publikációs Adattár (az MTA kutatóintézeteinek adattára) is különálló, eltérő rendszer, melyet nem lehet az MTA KPA-val együtt keresni vagy listázni.⁴

A cikk legszembetűnőbb része kétségtelenül az osztályonkénti táblázat (nyolc osztályra). A táblázatok láttán legtöbbször valószínűleg azonnal az ismerős neveket keresi, és önkéntelenül is értékeli a sorrendekben elfoglalt helyeket. Véleményünk szerint ebből fog származni a legtöbb vita, ugyanis az írásban megjelenő sorrendek („rangsor” felirattal) szinte mindenkit eleve felpiszkálnak. Különösen zavaró lehet a „bekerülhetnének a KPA fenti listájába” megjegyzés és névsor a lábjegyzetekben azok számára, akik nem kerültek be ide sem. Nem lenne baj, ha ezek a táblázatok csak tudományos versenyre, többleteljesítményre sarkallnának, de baj, ha a publikációk áttekintésének elutasítására ösztönöznek. Tudományos életünk átláthatósága, a publikációk szerepe biztosan javul, ha egyre jobban át tudjuk tekinteni, és egymáshoz tudjuk viszonyítani publikációs tevékenységünket. Könnyen lehet azonban, hogy ezek a táblázatok gyakori elutasításra fognak találni. Miért?

- köztudott, hogy a tudományometriai számokat nem szabad szakterületek között összehasonlítani, az osztályokon belül sem.

⁴ A cikkben említett KPA-TPA-összekötötés is sajnos csak azt jelenti, hogy adatokat jól/rosszul át lehet vinni a TPA-ból a KPA-ba, de a két rendszer együttesen nem kereshető, és *on-line* össze sem köthető: külön-külön kell feltölteni őket, vagyis vagy dupla munka a karbantartás, vagy pedig az egyik el-elmarad. Ráadásul a TPA-ban eleve kevesebb a mező, tehát a KPA-ba átvitt információk még hiányosak is. A TPA idéző része nem is publikus (lásd <http://www.mtatpa.hu>)

Az osztályon belüli megkülönböztetés nélküli sorrend azt a képzetet kelti, mintha a megjelent nyolc osztály egy-egy homogén, közvetlenül összehasonlítható csoportot takarna,

- az adatok eredetileg más célból (tehát nem nyilvános rangsor céljából, hanem egyedi publikációs/idézetségi listák megjelenítésére) kerültek be az adattárba,
- az érintettek nem tudtak előre arról, hogy a nevük ilyen rangsorokban megjelenik,
- a kimaradók nem tudtak előre arról, hogy nevük itt megjelenhetne, de nem jelenik meg, mert a KPA-ban vagy klónjai egyikeben sincsenek rögzítve az adataik,
- a táblázatok alapjául szolgáló oldalak (Statistikai adatok [mycite]: URL) nem teljesek: nem mutatják, mely adatok melyik klónból származnak, és hogy mikor (feltehetőleg most is a június 30-i állapot látszik, ahelyett, hogy a máig bevitt adatokat mutatná), és nem lehet az egyéni listákat a weben szokásos módon kattintással elérni pontosabb információszerezés céljából,⁵
- osztályonként megjelennek „szöveges sajtó közlések” (ezek ellenőrizhetetlenek); fel-tűnnek – ha nem is nagyon nagy számban – rossz kategóriákba besorolt adatok is (kutatási jelentések könyvként, konferenciák könyvfejezetként, nem közzétett kutatási jelentések publikációként, nem nyomtatott előadások cikként),
- a felhasznált számok hitelessége mérsékelt, nem előzte meg sem a KPA, sem a többi

⁵ A nevek rákattintva most az MTA honlapján látható egyéni weboldalak jelennek meg, ami nem baj, de innen a publikációs jegyzék sokszor nem érhető el. Megoldás lenne, ha például a publikációk számára kattintva megjelenne a teljes publikációs jegyzék, és ebből legalább a fontosabb cikkek teljes szövege is.

mycite-ok átfogó tartalmi ellenőrzése túlmenően a szokásos ellenőrzéseken („a szerző a felelős saját adataiért” – ez igaz, de nem zárja ki a szerzői elfogultságot a bevitelnél, és az adatok egy része az adatellenőrzések ellenére máig így is maradt),

- a publikációs listák a szerzők által megadva, vagy az általuk megadott listák alapján készültek – ezek nagyjából teljesek, de a besorolások emiatt kissé esetlegesek,
- a hivatkozások sokszor csak az adatbázisokból (*WoS, MedLine*) kerültek be, az ezeken kívül esők sokszor nem is szerepelnek (például: a *Scopus* – az OKM ígérete ellenére – még mindig nem férhető hozzá általánosan),
- az idézetek egy része önidézet, aminek nincs értéke, az itt közölt Hirsch-indexek kiszámítása mégis figyelembe veszi őket,
- a kiszámított egyéni mérőszámok, különösen a Hirsch-indexek érvényessége a fentiek miatt kétséges, pedig a mycite önmagában jó program, csak az adatfeltöltés hiányos, illetve még nem elég pontos.

A következőkben a fenti szempontok egy részét elemezzük.

Össze nem mérhető adatok összehasonlítása

Trivialitás, hogy a tudományometriai számokat nem szabad szakterületek között összehasonlítani. A Tudományos Publikációs Adattár weboldalon (Statisztikai adatok [mycite]: URL) ezt olvassuk: „A sorrendek különböző célú szakmai értékeléseket segíthetnek, de a szakterületek és a sorrendek nem összevethetőek. A mutatók használatakor tekintettel kell lenni az egyes tudomány- és szakterületek sajátosságaira, publikálási szokásaira.” A cikk 3. oldalán hasonló figyelmeztetés olvasható. Ez mind nagyon igaz, de a cikkben ténylegesen kinyomtatott és hangsúlyosan „rangsor-

nak” hívott sorrendeken sajnos nem segít. Az önálló szakterület ugyanis *nem akadémiai osztályokat jelent*, szűkebb szakmacsoportokat is csak nagyon óvatosan. A Műszaki Osztályon biztos igazságtalan ilyen számok alapján összehasonlítani egy anyagtechnológust egy áramlástanossal, vagy egy elektronikai technológust egy szabályozásméleti szakemberrel, vagy egy biológiai/villamos határterületen dolgozó szakembert egy villamosgépes tudóssal. Ezen valamit segíthetne a szakterületek megadása a táblázatokban, de ez sajnos tipikusan formális és semmitmondó („Műszaki tudomány”). Az osztályonkénti sorrendek megadása valószínűleg többet borzol az idegeken, mint amennyire tevékenységre sarkall az egészséges hiúságon keresztül. . .

Mindezek tetejében ezek a számok egyrészt az életmű gyarapodásával (életkor) nőnek, másrészt a kutatási időszakról is függenek: korábban a publikációs és keresési lehetőségek, valamint a „cikkyártás hatékonysága” és sebessége, valamint a számítógépes lehetőségek a töredékét tették ki a mainak. Ezért korrektil nem mérhető össze egy fiatalabb és egy idősebb kutató, még akkor sem, ha azonos életkorra vetített adatokat nézünk.

Rendkívül sokat számít a konkrét kutatási terület helyzete és publikációinak szerepe a világban. Például a textilipar mára döntő részben Kínába került. A kínai szakemberek főleg kínaiul publikálnak és olvasnak. Hogyan értékeliük a megmaradt európai és amerikai textiles szakemberek publikációit, csekély hivatkozottságukkal? Leírjuk-e őket, mint keveset produkálókat? Másokkal biztosan nem lehet összevetni a mérőszámaikat. A hatékony témaváltás jó néhány évbe kerül, egy ideig kevesebb közleménnyel és idézettel. Figyelembe tudja-e venni akármilyen mérés a kénytelen témaváltásokat?

Legyünk reálisak. A számok fontosak és valóban mérnek értékeket. *Azonban nem általános összehasonlításban*, hanem az illetőt saját magához, illetve szűkebb szakterületéhez viszonyítva. A sok publikáció a legtöbb esetben az átlagnál nagyobb aktivitást jelent, a folyóiratcikkek hiánya pedig valószínűleg értékrendbeli zavarokra utal. Az idézettség azt jelenti, hogy mások feltehetőleg olvassák, amit a szerző ír. A publikációk és az idézetek számában azonban nem mindegy, hogy általában mennyi a társszerző és milyen sorrendben (három hasonló társszerző együtt mintegy háromszor annyi cikket ír, mint egyedül...), melyek a szakterület publikációs szokásai, milyen hosszú a szakterület átlagos hivatkozási listája, mennyi a közvetlenül vele dolgozó vagy névlegesen hozzá tartozó doktórandszok száma stb.

Mindezek miatt kijelenthetjük, hogy a mérőszámok önmagukban **nem alkalmasak értékmérésre!** Egy-egy szakembert csak szakember kollégái tudják helyesen értékelni, az adatbázis(ok)ból nyert információk alapján. Erre való a tudományos bizottságok habitusvizsgálata. Ők azok, akik a rengeteg módosító tényezőt és egyéni körülményt tényleg mérlegelni tudják. A számok tehát mindössze segédeszközök, ahogy az adatbázisok is azok. És az adatbázisok közül a megfelelően kitöltött mycite nyújthatja a legteljesebb segédinformációt az egyes szerzőkről. Korrektil azonban csak emberek képesek dönteni, főleg, ha a számítógép a kezük alá dolgozik az adatok megfelelő tállalásával.

Rangsorok közzététele

Az adatvédelem egyik legfontosabb elve, hogy az adatokat csak a megadott célnak megfelelő módon szabad felhasználni, vagyis az információs célból közzétett publikációs listákat

minősítésre, sorrendezésre és ezek megjelenítésére a szerzők külön engedélye nélkül nem szabad felhasználni. Természetesen nem azt akarjuk mondani, hogy ezek személyes adatok, a „publikáció” neve is azt mutatja, hogy közzétett adatokról van szó, de nagyon sok szerző nem az összehasonlíthatóság igényével készítette el a listáját, és különösen a hivatkozási listák nem teljesek.

Megjegyezzük, hogy ezek a nehézségek a táblázatok közzétételéből származnak. A mellékletben részletesebben is szerepel, melyik másik adatbázis használatából milyen más nehézségek származhattak volna. A mycite – megfelelő feltöltéssel – az elérhető legteljesebb információt tartalmazná, és így jobban használható eredményeket adna.

Valószínű, hogy ilyen rangsorokat nem is lenne szabad nyomtatásban rögzíteni. A (Statisztikai adatok [mycite]: URL) alatti weboldal azért jobb ennél, mert nem rögzít sorrendet, hanem szempontok szerint átrendezhető listákat mutat. Sajnos azonban jelenleg (aug. 31) ez a weboldal nem dinamikus, vagyis hiába javítja ki valaki az adatait valamelyik mycite-ban, az ismeretlen dátumú frissítéstől függ, hogy ezek a táblázatban mikor jelennek meg, és a listában semmilyen információ nem látható arra nézve, hogy valakinek az adatai mikor frissültek utoljára (az utolsó talán június végén lehetett), vagy hogy hol kezeli őket. Úgy néz ki, mintha az adatok frissek lennének. De nem feltétlenül azok. . .

A „tudósok teljesítményét” egyébként jobb lenne a „tudósok munkásságának” hívni: a teljesítmény inkább az alkotó évek számával osztott mutató lehetne, legalábbis ha a cikk a publikációs tevékenységet mutatná be és nem az idézettséget. Azonban a mérőszámok közül kettő az idézettségre vonatkozik, a harmadik pedig arra, hogy *hol, melyik*

folyóiratban publikált az illető (bizonyos idézeteken keresztül értékelve), nem arra, hogy mennyit, vagy milyen minőségben. Ráadásul csak az *ismert adatok alapján*, nem általában.

Van még egy nehézség, ami az adattárak gondos tervezésével sem teljesen kerülhető ki. Az ember gondolkodó lény. Ha tudja, hogy tevékenységét méri, különösen, ha ez értékelését, támogatását, lehetőségeit stb. érinti, akkor aktivitását önkéntelenül is a jobb mérési eredmény irányában módosítja. Ha a darabszámokat listázzuk, megszorodnak a könnyen publikálható írások (konferenciák, közölt kutatási jelentések [Technical Report]). Ha a könyveket jobban értékeljük, megjelennek a könyv-szerű összeállítások. Ha folyamatosan csak a folyóiratcikkek összdarabszámát nézzük, akkor megszorodnak a hazai folyóiratcikkek. Ha az önidézetek számítanak, akkor a cikkekben több lesz az önhivatkozás. Nem mindenki jár így el, de ez még rosszabb: az jár jól, aki alkalmazkodik a körülményekhez. Aki elviszéri, függetlenül gondolkodik, az rosszul jár. Részben ez a magyarozata annak, hogy sokan csak az impakt faktoros folyóiratcikkek akarják értékelni: ezt nehezebb itthonról befolyásolni. Bár ha a összegzett IF-et mutatjuk, a hangsúly észrevétlenül el fog csúszni a nagyobb impakt faktorú folyóiratokban való publikálás felé (például műszakiak az orvosi határterületek irányába)... Nagyon vigyázni kell tehát azzal, mit teszünk közzé és hogyan.

A szerzők által bevitt adatok

A bevitt adatoknál a szerzők (adminisztrátorok) hozzáállása nagyon eltérő. Ugyan mindannyian tudjuk, mi a publikáció, mégis sok publikálatlan (például OTKA) kutatási részjelentés jelenik meg a listákban, jó esetben az „Egyéb” kategóriában. Van, aki gondosan

megjelöli, melyek a nem tudományos cikkei, de van, aki válogatás nélkül mindent megad. Van, aki szóban elhangzott előadásait is beírja (jó esetben oldalszám nélkül, rosszabb esetben a kivonat oldalszámával), van, aki az utolsó cikkig aprólékosan pontos. Van, aki a lexikon minden címszavát külön tételként beírja, van, aki egyetlen publikációnak tekintti ezeket. Van, aki egy konferencia szerkesztőbizottsági tagsága esetén megadja magát szerkesztőnek (és a kiadványt könyvként), van, aki még a konferenciáit sem adja meg. Sokan vagyunk, és sokfélék. Ezért a publikációk darabszámát nem szerencsés egyszerűen megadni. Az az érv, hogy az adatokért a szerző felel, tetszetős, de ha listákat készítenek, akkor a számok és a rangsor nem függhetnek a szerzők szokásaitól és hozzáállásától. Helytelen a lelkiismeretes és pontos szerzőket hátrányosabb színben feltüntetni, csak mert lelkiismeretük érzékenyebb, mint másoké.

Hasonló a helyzet a „szerzők által megadott” számokkal. Ez a kifejezés azt jelenti, hogy ezek a szerzők nem vitték fel adataikat adattárba, ehelyett a számaik ellenőrizhetetlen Word file-okból származnak, valószínűleg a doktori értekezés habitusvizsgálatához beadott számok alapján. Az ilyen számokba és listákba még a gondos szerzőknél is sok hiba becsúszik. Sokszor még ismétléseket is lehet látni az ilyen listákban. Ezért nem szerencsés a többiek által az adattárban többé-kevésbé ellenőrizhető módon megadott adatok közé kézzel megadott, ellenőrizhetetlen adatokat keverni. A jövőben jobb ezt elkerülni. Aki meg akar jelenni a listában, az vagy vigye fel az adattárba az adatait, vagy az MTA megbízottja⁶ vigye fel a Word file-ok tartalmát.

⁶ Például: Kutatásszervezési Intézet (KSZI) vagy a Doktori Tanács

A hivatkozások száma

A szerzők mycite-ban megadott hivatkozásaik száma az „igazi” hivatkozási számokhoz képest szerzőnként nagyon ingadozó, ahogy mindegyik adatbázisban az. A mycite előnye az, hogy *teljessé tehető, ha a szerző akarja*. Most sok esetben még nem az.

Aggályos továbbá, hogy a táblázatban a hivatkozások számai sajnos az önhivatkozásokot is beleszámítva vannak megadva (minden ilyen értelmű megjegyzés nélkül), pedig ezeket semelyik osztály sem értékeli a habitusvizsgálatban. A mycite egyik előnye pont az lenne, hogy benne az önhivatkozások egyértelműen meg vannak jelölve.⁷

Nyilvánvaló előnyben vannak azok, akik a doktori értekezés beadása miatt nemrég végigkeres(tet)ték a könyvtárat. Az ő listájuk közel teljes lehet (legalábbis, ha bevitették). Azoknak is jó a helyzetük, akiknek a szakmai szokások alapján a WoS-ban vagy a Medline-ben benn van a legtöbb hivatkozása. De a többiek helyzete ilyen szempontból rossz.

Mérszámok

A fentiek miatt a megjelentetett számokat (összes idézetek száma, Hirsch-index, összegzett impakt faktor) nagyon óvatosan kell kezelnünk. Van, aki pontosan megadta a publikációit, és van, aki alig-alig.

Az összes idézetek száma

Az idézetek száma elvben a cikkek szakmai közvéleményre gyakorolt hatását méri. Legalábbis ez az első gondolatunk. Csakhogy...

- a táblázatokban közölt számokba az önidézeteket is beleszámítja,

⁷ Szigorúan véve a szerző szempontjából még az önhivatkozásokat is osztályozni lehetne saját és társszerzői önhivatkozásokra.

- a megadott, de nem publikációértékű idézeteket (szóbeli hivatkozás, tudományos diákköri munka stb.) minden további nélkül beleszámítja,
- nem lehet látni, melyek az ellenőrizhető, adatbázisokból (WoS, Scopus, IEEE Cross-Ref) vett idézetek, és melyek a közvetlenül nem ellenőrizhető idézetek,
- nem lehet látni, kik vitték be kézzel is idézeteket, és kik csak adatbázisokból,
- a soktárs szerzős publikációk idézetei ugyanúgy számítanak, mint az egyszemélyesöké (lásd 5.4).

A további gondokat és megoldási javaslatokat lásd a következő pontban (Hirsch-index).

Hirsch-index

Az ún. Hirsch-indexet a fizikus Jorge E. Hirsch találta fel az idézettség összehasonlítása céljából. Az index azt adja meg, hogy a szerzőnek hány olyan publikációja van, aminek legalább annyi idézője van, mint ez a szám. Ha tehát az index 12, akkor a szerzőnek van 12 olyan publikációja, melyekre egyenként legalább 12 idézet ismert. Ez az index jobb, mint a semmi, de a következő bajai vannak:

- csak aránylag nagy publikáció- és idézőszámokra mér elfogadhatóan (fizikus találta ki, akinél ez nem baj...), kis értékeknél nagyon érzékeny a véletlenre, és hamar „telítődik” (vagyis az első néhány értéken könnyű túljutni, aztán hirtelen nehezedik a továbblépés, tehát pl. a műszaki területen nem ott differenciál, ahol kellene),
- mivel az eredeti definíció az SCI-n alapult, az önhivatkozásokat is beszámította, ami elvileg nagyon vitatható – éppen ezért képes a mycite a keresést leszűkíteni a független hivatkozásokra, ami szintén érvényes, sőt vonzóbb index-definíciót jelent,

- nem vesz tudomást a kiugró hivatkozási értékekről – emiatt használják több helyen a g-indexet, mely azt adja meg, hogy a szerzőnek hány olyan publikációja van, aminek az átlagos hivatkozási száma legalább annyi, mint az index. Ha tehát a g-index 15, akkor a szerzőnek van 15 olyan publikációja, melyekre összesen legalább $15 \times 15 = 225$ idézet ismert (de ez lehet egyetlen publikáció 225 idézete is, ha van 15 publikációja),
- nem vesz tudomást a hivatkozott publikációk társszerzőinek számáról (lásd 5.4).

A Hirsch-index minden olyan adatbázisból kiszámítható, mely idézeteket tartalmaz. Ámde csak az adott adatbázisban fellelhető információt használja fel. Nem is tehet mást. Íme hét, bárki által használható lehetőség a Hirsch-index kiszámítására:

- *Web of Science* (önhivatkozásokkal)
- *Scopus* (önhivatkozásokkal, illetve azok nélkül)
- mycite (önhivatkozásokkal illetve azok nélkül)
- Quadsearch (Quadsearch honlap, URL) (Google Scholar-ból, önhivatkozásokkal, weboldal)
- Publish or Perish (Publish or Perish honlap, URL) (Google Scholar-ból, önhivatkozásokkal. PC-s program)⁸

Van, akinek az így kapott Hirsch-indexei közül öt más és más... sőt, az index elvben bármikor változhat is, bár ilyenkor általában legalább nő, és nem csökken.

A Tolnai-cikk (Tolnai, 2008) a mycite alapján számol, az önhivatkozásokat is beleszámítva. Azt mondhatnánk: helyesen jár el, hiszen mindenkit azonos mércével mér. Csakhogy...

⁸ Kicsit mást ad, mint a Quadsearch...

- a szerzők egy része az önhivatkozásait nem is vitte be, mert az gondolta (helyesen), hogy ezeknek nincs különösebb értékük,
- más szerzők az „Egyéb” csoportban feltüntettek nem publikáció jellegű hivatkozásokat is (diplomaterv, TDK-dolgozat, szóbeli említés), s ezzel adataik „javultak”,
- kevesen vették a fáradságot, hogy a hivatkozásait a nehezebben feldolgozható forrásokból is összegyűjtsék (például Google Scholar).

Vagyis az történt, hogy ebben az nyert, aki (egyébként cél és indok nélkül) bevitte az önhivatkozásait is, és most ráadásul még azt is üzenjük a szerzőknek ezzel a cikkel, hogy vigyék be ezeket, mert azzal jobbak lesznek az adataik – sőt minél több önhivatkozást tegyenek a cikkeikbe, mert így az adataik még tovább fognak javulni. Ezt biztosan nem akarjuk.

Sokat segítene, ha az önhivatkozások bármilyen figyelembe vételét megszüntetnénk, hiszen a mycite-ban pontosan tudjuk, melyek a független hivatkozások. Ráadásul a megadott index csak a mycite által látott Hirsch-index..., akinek nincsenek ide gondosan bevive a hivatkozásai, vagy csak részben, annak az indexe picit, és mindez nincsen feltűnően megemlítve. Ugyanez a baj az KPA-ban (Statisztikai adatok [mycite]: URL) megadott értékekkel is.

Összegezett impakt faktor (cumulative impact factor)

A folyóiratok impakt faktora a *Web of Science*-ben figyelt folyóiratokon belül van definiálva: adott évben a figyelt folyóiratokban az adott (szintén figyelt) folyóirat előző két évfolyamában megjelent cikkekre történő átlagos hivatkozásszám. Az $IF=0,816$ tehát azt jelenti, hogy az adott évben az adott folyóirat előző két évi

cikkeire a fenti körben átlagosan kevesebb, mint cikkenként egy hivatkozás esett.

Az impakt faktor elsősorban folyóiratok tudományos életre gyakorolt hatásának „minősítésére” szolgál. Ami az IF felhasználását illeti, ugyanúgy vigyázni kell vele, mint az összes tudományometriai számmal (Joint Committee, 2008; Schubert, URL), különösen mert

- az IF értéke nagyon függ a szakterület (és a folyóirat) szokásaitól: a társszerzők tipikus száma, a hivatkozások mennyisége, az önhivatkozások száma (beleszámítanak...), a cikkek hossza, az összefoglaló cikkek száma stb.,
- az ISI döntésétől, és nem a szakterület közvéleményétől függ az egyes folyóiratok figyelése/negligálása,
- a folyóiratválogatás döntően alapkutatócentrikus.

Az utóbbi időben egyre inkább terjed az „összegezett impakt faktor” figyelése. A táblázat is ezt tartalmazza: a publikált cikkek folyóiratai megfelelő évben érvényes IF-jeinek az összege. Ezt kíváncsiságból meg lehet nézni, de csak nagy óvatossággal szabad akármire is felhasználni. Oly mértékben szakterületfüggő, hogy egymással összevetve az alma és a körte esetével állunk szemben. Persze ha valakinek nincsen IF-es cikke, az könnyen lehet, hogy baj. De valóban százszor kevesebbet ér-e egy 0,021-es impakt faktorú folyóiratban megjelent cikk, mint a 2,100-esben? És a 0,000 impakt faktorú folyóiratokról vajon mit gondolunk?

Az összes publikációk száma

Az Tolnai-cikk (Tolnai, 2008) táblázatai nem tartalmazzák az összes publikációk számát, melyet azonban minden adatbázis megad. Ez a szerzők munkásságának értékelésében szin-

tén fontos lenne, de „sorrendi” mutatóként valóban jobb ezt sem használni. Egészen más például egy 600 oldalas könyv, egy 150 oldalas fejezet, egy konferenciakikk, és egy napilap-glossza. Összeadva ez mind darab-darab, sőt a mycite beleszámítja az „Egyéb” alatt esetleg megadott nem publikációkat is. A darabszám tehát túl globális mérőszám. Ha valakinek ötven publikációja van, akkor ez várhatóan kevesebb, mint a 250, de konkrét esetekben még ez sem biztos.

Nagyon ingadozó a társszerzők száma is. Nyilvánvaló, hogy a „magányos farkasok” eredménye egy publikációban egy főre számítva sokkal több, mint a tipikusan hat-nyolc fős publikációk szerzőié, hát még a tizenöt-húsz fősek szerzőié. Az MTA Műszaki Tudományok Osztálya habitusvizsgálatában ezért automatikusan osztunk a szerzők számával. Lehet azon vitatkozni, hogy ez nem a másik véglet-e. Mindenesetre a mycite képes arra, hogy tárolja a szerzők százalékos hozzájárulását (ezt használják fel a BME VIK doktori tanácsai), és értekezés beadásakor a publikációkhoz beírt százalékok nyilvánosak is. Ez elegendő ellenőrzésnek látszik: a méltatlanul leértékelt társszerző tiltakozik, ha akar.

Kitűzhető célok

Adatbevitel • Nagyon jó lenne, ha a tudomány valamennyi területén előbbre lépnénk a publikációk áttekintésével és közzétételével. Ez nagy munka, és elosztott rendszerben lehetne hatékonyan végrehajtani: minden intézmény (vagy intézménycsoport) futtathatna egy-egy helyi programot, és országosan kereshető lehetne minden adat. Hozzá kell tenni, hogy ez Magyarország versenyhátrányából is faraghatna egy kicsit: Európa általában nem áll még olyan jól a publikációs adatokkal, mint a mycite lehetőségei... , de tudjuk, hogy

minden adatbázis csak annyit ér, amennyit a benne tárolt adatok. Most kellene tehát az adatokat feltöltenünk.

Karbantartás • Elvben a fő bevitel után a karbantartás nem nagy dolog: időről-időre ki kell gyűjteni a *WoS*-ból, a *Scopus*-ból és a *Google Scholar*-ból az adatokat, és be kell vinni. Ez központilag reménytelenül nagy feladat, viszont helyileg, megfelelően kiképzett adminisztrátorokkal elképzelhető. Vagyis: az intézményi döntéshozókat meg kell győzni, hogy ez milyen fontos (pl. mert az intézményi és pályázati pénzek várhatóan össze fognak függeni a kimutatható adatokkal [lásd Kollár, 2008a]), és rendszeres tanfolyamokat kell tartani a jelentkező adatkezelőknek (könyvtárosok, adminisztrátorok, érdeklődő szerzők).

A jövő igényei

A publikációs listák kezelése, áttekinthetősége jó, de nem szabad, hogy bürokratikus öncéllá váljon. Nem egy „vezetői információs rendszer” a fő cél (bár ők is használni tudják), hanem egy mindenki számára egyformán áttekinthető, informatív, jól felhasználható, országos rendszer. Ezért

- fontos, hogy az adattartalom lehetőleg ellenőrizhető és javítható legyen, vagyis ahol csak lehet, a mycite-ban szereplő publikációs/idézetségi adatokat össze kell kapcsolni adatbázisokkal (mi melyikben található meg – ez növeli a hitelességet is),
- az adatfelvételi arányt javítani kell (myciteklónok telepítése, Word file-ok intelligens bevitel, adminisztrátorok kiképzése stb.),
- a szerzőket érdekeltébbé kell tenni a bevitelben (az ellenőrizhető adatok előnyben részesítése, doktori/akadémiai cím odaítélése adatbázisban lévő adatok figyelembevételével, kutatási és ösztöndíjpályázatoknál a mycite-ból való feltöltés az OTKA

példájára (például Bolyai-ösztöndíj), intézményi támogatások publikációkkal kimutatható tevékenységhez kötése stb.),

- a publikációs adatok részeként legalább a fontos cikkeket is teljes formában elérhetővé kell tenni (tartalomszolgáltatás – erre a mycite-ban már most is van lehetőség),
- a mycite-ban lévő adatokat időről-időre fel kell tölteni adatbázisokba (például: *Scopus* vagy *Google Scholar*), legalábbis azokat, amelyek még nincsenek ott meg, illetve a listákat angol nyelven kereshetővé tenni,
- (az előző két pontban leírtak hatékonyan és tisztességes úton növelik kutatóink nemzetközi láthatóságát, hírnevét, olvasókat, felhasználókat és így idézőket toboroznak, ez pedig mindenkinek jó)
- a szerzők számára még több szolgáltatást kell nyújtani (információs oldalak, statisztikák, mérőszámok),
- intézmények, tanszékek, csoportok számára állandóan naprakész összefoglaló listákat és táblázatokat kell nyújtani,
- a mycite programot a szélesebb elterjesztéshez professzionális szolgáltatássá kell fejleszteni (ehhez máris közel van, de ha az MTA rászánna magát, még egy fejlesztési ütem sokat lendíthetne elterjedésén).⁹

Ha az adatfeltöltést támogatni szeretnénk, érdemes lenne az MTA KPA „gazdahelyén” (MTA KSZI) az érintettek, illetve adminisztrátorai számára olyan tanácsadó/adatfeldolgozó szolgáltatást biztosítani, hogy a listákat helyileg be tudják vinni. Tréningeket lehetne szervezni, hogy adminisztrátorok/könyvtárosok az adatbevitt gyakorlatják. A KSZI jelenlegi könyvtárosi kapacitása a karbantar-

⁹ Más dolog a jelenlegi fejlesztés kevés felhasználóval és kézi kipróbálásokkal, és más dolog lenne sok felhasználó és sok helyen futó program, széleskörű műszaki támogatással.

tásra sem igazán elegendő. Azután mindazok, akiknek hiányosak az adataik, kaphatnának barátságos e-mailes felszólítást a bevitelre. Akiknek aránylag régi (mondjuk fél évnél régebbi) a tulajdonukban lévő adatok utolsó változtatása, vagy a korábbi trendtől nagyon eltér publikációik/idézők száma, időről időre kaphatnának erről e-mailen értesítést. Ehhez

mindössze az kell, hogy a KPA-ban (mycite-ban) tárolt e-mailek helyesek legyenek, és a program képes legyen ilyen szolgáltatásra.

Kulcsszavak: *tudománymetria, scientometria, impakt faktor, Hirsch-index, idézettség, hivatkozás, KPA, köztestületi publikációs adattár, publikációs lista, adattár*

Melléklet

A tudománymetriai adatok néhány lehetséges forrása – háttérinformációk a cikkhez –

A tudományos publikációk adatait már régóta katalógusok/adatbázisok tartalmazzák. Ezek először nyomtatásban készültek, majd kb. a 90-es évektől számítógépen, először CD-n, majd a weben is elérhető formában.

Web of Science

Különleges helyet foglalt el köztük az amerikai ISI által összeállított *Science Citation Index* (a weben ma elérhető formáját *Web of Science*-nek hívják). Ennek különlegessége, hogy az *idézetekeket figyeli*, 1960-tól jelenik meg, de 1945-ig visszamenőleg tartalmaz adatokat. Azóta kb. 6500–8700 folyóiratot figyelnek úgy, hogy a cikkek irodalomjegyzékeit rendszeresen, standard formában beírják az adatbázisukba. Bár ez a forrásszám korlátozott, eléggé sok ahhoz, hogy hosszú ideig az *SCI* legyen a hivatkozások szisztematikus keresési forrása.

A *WoS* nehézsége, hogy csak adott körön belül mozog (bár a figyelt folyóiratok összes hivatkozását feldolgozza, tehát például a nem figyelt konferenciacikkekre történőket is, és 2008-tól megjelentek bizonyos konferencia-

anyagok is), és az önhivatkozásokat¹ alapértelmezésben ugyanúgy számolja,² mint a független hivatkozásokat. Ennek részben az az oka, hogy az azonos nevű szerzőket nem képes megkülönböztetni: külön információforrás nélkül ugyanis nem lehet megállapítani, hogy a „külső cikkekre” történő hivatkozások megegyező szerzőre történnek-e.

Scopus

A *Scopus* a *WoS* versenytársaként jelent meg 2005 környékén. Abból a felismerésből indult ki, hogy ma már sok folyóirat és konferenciaközlemény hozzáférhető a weben, ezért lehetséges az elektronikus feldolgozásuk. Ezért szélesebb körből és kevésbé válogatva merít, mint a *WoS* (mintegy 23 ezer periodikus kiadványból), ami jó, de kicsit esetleges. Ugyanakkor a szerzőket nevek és munkahelyek szerint csoportosítja, és a szerzők maguk meg tudják

¹ Azonos szerző van a hivatkozó és a hivatkozott cikk szerzői között.

² A független hivatkozó cikkeket külön ki lehet szűrni, de az nem használható ugyanúgy, és nem is a hivatkozásokat, hanem a hivatkozó cikkeket szűri...

adni, hogy mely név/munkahely-párok fedik őket. Ebben azonban a *Scopus* kezelői nagyon elővigyázatosak: csak a publikációs listákat látva hajlandók a csoportosításra, ami a hitelességüket nagymértékben emeli. Tapasztalataink szerint az alaptudományok területén (matematika, fizika, kémia, biológia stb.) az azonos szerzők hivatkozási száma hasonló a *WoS*-ban és a *Scopus*-ban (bár az önhivatkozások az utóbbiban jól kiszűrhetők), míg az alkalmazott tudományokban, pl. műszaki területen a *Scopus* több hivatkozást tartalmaz.

Google Scholar

Van egy különleges rendszer: a Google Scholar. Ez bevallottan a weben „szüretel”, web crawlerek (automatikus lépkedők/olvasók) segítségével, és ebből állít össze adatbázist. Ez nagyon jól hangzik, de előnye hátrány is: válogatás nélkül minden használhatónak látszó weboldalt feldolgoz. Ezért hibás, elgépel, meg nem jelent cikkeket, hivatkozásokat is tartalmaz, s mérésre csak korlátozottan alkalmas. Az azonos nevű szerzőket nem kezeli jól.

Szakmai adatbázisok

Vannak szűkebb szakmai adatbázisok is: *MedLine* (orvosi/biológiai terület), *Chemical Abstracts/SciFinder Scholar* (kémia), *Zentralblatt für Mathematik/ZMATH* (matematika), *Mathematical Reviews/MathSciNet* (matematika), *ArXiv* (fizika), *Inspec* (villamosmérnöki), *IEEE Xplore* (villamosmérnöki) stb. Ezek merítése a szűkebb terület miatt gondosabb és átfogóbb, de csak az adott részterületre korlátozódnak.

A Köztisztületi Publikációs Adattár (KPA) és a többi mycite program

A fentiek miatt született meg a *Köztisztületi Publikációs Adattár*. Ebbe a legfontosabb pub-

likációk és idézetek az adatbázisokra támaszkodva automatikusan bevihetők, és kiegészíthetők a szerző által megadottakkal: a szerző saját publikációs listája alapján sok adat átvehető, és a hiányzó adatok kézzel pótolhatók. Vagyis gondosabb szerzők esetében a „teljes” adatsor összeállítható. Lehetséges az is, hogy Word listákat félautomatikus beolvasunk, és ami esetleg téves, utólag javítsuk, sőt, a Google Scholar és konferencia CD-k adatai is átvehető, legalábbis ha a szerző ezek kikérésére/átnézésére hajlandó, vagy könyvtárosokat/adminisztrátorokat ezzel megbíz.

A KPA szűkebb területre koncentrál (köztestület), tehát a fiatal tudóspalánták (PhD-ra készülők) nem tudják ide felvinni az adataikat. Ez az egyéni fejlődés szempontjából aggályos: legkésőbb a PhD-értekezés benyújtásakor kellene bevinni a publikációs listákat, és a védés után csak karbantartani. Ezt az MTA KSZI által kezelt KPA nem engedi meg, ezért az intézményi klónok (például BME PA) kezelése elengedhetetlen.³

A BME PA jelenleg mintegy 28 ezer publikációs tételt és mintegy 30 ezer idézőt tartalmaz. A cikk 977. oldalán megadott adat (mindössze a BME-ből származó egyetlen személy 368 cikkének adatait használták fel) csak valamilyen tévedésen alapulhat, mert az MTA Műszaki Tudományok Osztálya akadémikusainak, doktor képviselőinek és bizottsági elnökeinek mintegy 50 %-a a BME PA-ban kezeli az adatait (lásd Az MTA Műszaki Tudományok Osztálya tagjainak publikációi,

³ Itt külön is köszönjük Kocsis Józsefnek a mycite rendszer használhatóságát a BME számára, és Földvári Zoltánnak az MTA Műszaki Tudományok Osztálya doktori és a BME Villamosmérnöki és Informatikai Kar PhD-pontozásos értékeléseinek programozását. Ez utóbbi tette lehetővé, hogy a BME VIK doktori tanácsai kötelezően előírták a BME publikációs adattár használatát PhD-értekezés benyújtásakor.

2008), és a megírásakor felhasznált adatokat a cikk is csak innen vehette.

Intézményi Publikációs Adattárak

A publikációs adatok nyilvántartási igénye párhuzamosan több intézményben is megjelent. Azt gondoljuk, a BME helyesen döntött, amikor – óvatos lépésekkel – a KPA klónját elkezdte alkalmazni. Tudomásunk van más helyekről is, ahol valamilyen szintű adatbázis feltöltése elkezdődött, vagy tervezik (például: Miskolci Egyetem, Semmelweis Egyetem, Eötvös Lóránd Tudományegyetem, de biztosan vannak mások is). Ha a tudományos élet és az ország érdekeit nézzük, akkor ez az a pont, ahol azonos irányba kellene mennünk. A mycite- alapú programok ma már elég sokat tudnak ahhoz, hogy minden igényt kielégítsenek, és az esetleges intézményi mycite-ok önállóan is futtathatók, tehát az intézményi függetlenség is megőrizhető. Össze kellene fognunk, az esetleg még felmerülő igényeket megfogalmazni és programoztatni: például az egyik adattárba bevitt publikációkat a másik adattárban kezelt szerzőhöz

IRODALOM

- A legtöbbet idézett publikációk (mycite). URL: <http://www.mtakpa.hu/kpa/multi/molist.php>
 Az MTA Műszaki Tud. Osztálya tagjainak publikációi (kigyűjtés, folyamatosan frissítve). <http://www.mit.bme.hu/services/pubinfo/MTA-MTO.xls>
 Hirsch index. http://en.wikipedia.org/wiki/Hirsch_number
 Joint Committee on Quantitative Assessment of Research (Robert Adler, John Ewing, Peter Taylor) (2008): *Citation Statistics*. IMU (Nemzetközi Matematikai Unió) -tanulmány az idézettségi statisztikákról. URL: <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>
 Kollár István (2008a): Tudományos publikálás hatékonyan. <http://www.mit.bme.hu/services/pubinfo/szakirodkezeles.pdf>

is hozzá kellene tudni rendelni. A programok futtatásához – mint minden ilyen programhoz – azonos háttér adatok kezelésére lesz szükség (folyóiratok neve, ISBN száma, kiadási helye, lektoráltsága, impakt faktora, városok és országok magyarul/angolul/eredeti nyelven, államok nevei, amelyek országban vannak, kiadók és székhelyük, konferencia-szervezők, konferenciák, szerkesztett könyvek, publikációk és idézők kategóriái és alkategóriái stb.), ami központilag, szolgáltatás-szerűen felkínálható lenne. Ezekkel mindenki jól járna. Felesleges mindezt külön-külön megtenni. Vagy mégis ilyen gazdagok vagyunk?

A Hirsch-index meghatározása

Az eljárás elvileg egyszerű: a publikációkat sorba rendezzük az idézők száma szerint, és megkeressük, melyik az utolsó, melynek legalább annyi idézete van, mint a sorszáma. Szerencsére az adatbázisok mindegyike nyújt valamilyen automatikus kiszámítási lehetőséget. Ezek „receptjét” a weben külön megjelenttük (Kollár, 2008), hogy így az olvasók a saját indexüket könnyen ellenőrizni tudják.

- Kollár István (2008b): A Hirsch-index meghatározása (receptek). <http://www.mit.bme.hu/services/pubinfo/MT-tudomanymetria-kiem.pdf>
 Publikációs információs weboldal a BME-n. <http://www.mit.bme.hu/services/pubinfo/>
 Publish or Perish honlap. <http://www.harzing.com/pop.htm>
 Quadsearch honlap. <http://quadsearch.csd.auth.gr/index.php?lan=1&cs=2>
 Schubert András: Impakt Faktor. URL: http://www.kfki.hu/library/imp/impakt_faktor.htm
 Statisztikai adatok (mycite). URL: <http://www.mtakpa.hu/kpa/multi/statisztika.php>
 Tolnai Márton (2008): Tudósaink mérhető teljesítménye az MTA Köztisztületi Publikációs Adattár adatainak tükrében. Magyar Tudomány, 8, 976–988. URL: <http://www.matud.iif.hu/o8aug/tolnai.html>