

Pismo uredništvu/Letter to the editor

# ANALIZA CITIRANOSTI IN NJENE PASTI

Štefan Adamič

Inštitut za biomedicinsko informatiko, Medicinska fakulteta, 1000 Ljubljana

Odmevnost znanstvenih publikacij, ki jo ugotavljamo z njihovo citiranostjo, se vse bolj uveljavlja kot najpomembnejši kazalnik kakovosti dosežkov temeljnih in uporabnih raziskav. Podatki o citiranju so zbrani v računalniških zbirkah ameriškega Inštituta za znanstveno informiranje (ISI), ki so bile zasnovane pred 50 leti, pozneje pa so jih intenzivno razvijali in dopolnjevali.

V slovenski medicini se je ta metoda kmalu uveljavila. K temu je precej pripomoglo tudi to, da ima CMK tiskano verzijo zbirke Science Citation Index (SCI) od začetka njenega izhajanja. Raba zbirke je še narasla, ko se je pojavila njena računalniška verzija na kompaktnih diskih. Ko pa je pred desetimi leti Inštitut za biomedicinsko informatiko (IBMI) po naročilu Sveta za znanost v medicini vse zapise v zbirki Biomedicina Slovenica (BS) opremil s podatki o njihovi citiranosti, je citiranost postala eden najpomembnejših elementov pri ocenjevanju raziskovalne uspešnosti v slovenski medicini. Uporablja se ne samo pri ocenjevanju raziskovalnih programov in projektov, ampak tudi pri napredovanju v akademski karieri na Medicinski fakulteti, pri podeljevanju nazivov svetnik in višji svetnik na klinikah ter inštitutih in ob drugih podobnih prilikah. Za opremljanje zapisov v Biomedicini Slovenici s podatki o citiranosti publikacij so na IBMI razvili polavtomatsko metodo za analizo citiranosti. Podatke so v začetku zajemali iz zbirke SCI na kompaktnih diskih. Ti podatki so dostopni le po prvem avtorju publikacije, zato so tudi posamezne publikacije v BS opremljene s podatki o citiranosti glede na prvega avtorja. Poleg podatka o vseh citiranjih je v BS naveden tudi podatek o citiranjih brez avtocitiranj (čista citiranja) in delež čistih citiranj, ki pripada posameznemu avtorju (indeks citiranosti).

Ko je Inštitut informacijskih znanosti (IZUM) ponudil uporabnikom v Sloveniji servis Web of Science (WoS), je IBMI začel priprave, da temu servisu prilagodi svoj postopek za polavtomatsko analizo citiranosti. Z letom 2003 je tako že lahko začel uporabljati WoS namesto SCI na kompaktnih diskih. V okviru WoS je namesto zbirke SCI dostopna njena razširjena verzija SCI Expanded, ki vključuje dobrih dva tisoč revij več kot sama SCI. Poleg tega WoS nudi zbirne podatke o citiranosti posameznega avtorja ne glede na to, ali je prvi ali ne in za poljubna obdobja. Ker po naši metodi analiziramo citiranost posamezne publikacije, citiranost avtorja pa dobimo kot vsoto citiranj njegovih publikacij, smo v opremljanju zbirke Biomedicina Slovenica s podatki o citiranosti, dobljenih iz WoS, ohranili načelo zbiranja podatkov po prvem avtorju publikacije.

WoS naj bi torej omogočal, da raziskovalec lahko pod svojim priimkom in začetnico imena dobi podatek o citiranju vseh svojih publikacij. Pa res vseh? Nekatere neskladnosti med podatki, dobljenimi po prvem avtorju publikacije, in dobljenimi za posameznega avtorja ne glede na to, kdo je prvi avtor članka, so nas spodbudile, da to preverimo.

Na vzorcu 300 znanstvenih dokumentov, ki so citirani, če ustrezen podatek iščemo po prvem avtorju, smo isti podatek iskali pri ostalih avtorjih. Ugotovili smo, da se oba podatka kar v precej primerih ne ujemata. Zato smo v nadaljevanju skušali ugotoviti, kaj je vzrok tem razlikam.

Prvi in najpogostejši vzrok je v različnih zapisih podatkov o citirani publikaciji. V servisu WoS imamo namreč dve vrsti zapisov. Pri eni vrsti imamo samo nekaj osnovnih podatkov: ime in začetnice imen prvega avtorja, skrajšan naslov revije (brez naslova članka) ali samostojne publikacije, leto izdaje, volumen (pri revijah) in začetna stran članka. To vrsto zapisov bomo imenovali nepopolni zapis. V drugi vrsti zapisov pa imamo poleg omenjenih podatkov na voljo še naslov članka ali samostojne publikacije, priimke z začetnicami imen vseh avtorjev in njihovimi naslovi, število citiranj v članku, število njegovih citiranj, izvleček in še nekatere podatke. Tak zapis, ki je izpisan v modrem in podčrtan, bomo imenovali popolni zapis.

Pri nepopolnem zapisu je torej naveden samo prvi avtor, zato citiranja te publikacije pod priimki soavtorjev ne bomo našli. Katere vrste publikacije pa se pojavljajo v skupini, kjer je zapis publikacije nepopoln in citiranje navedeno samo pri prvem avtorju, ter kakšen je pomen tega za druge avtorje?

V skupini publikacij z nepopolnim zapisom, kjer je citiranje navedeno samo pri prvem avtorju, so predvsem vse samostojne publikacije: znanstvene knjige in zborniki ter prispevki v njih. Čeprav je tega gradiva sorazmerno malo, lahko v posameznih primerih to močno vpliva na podatek o celotni citiranosti prizadetega soavtorja. To najbolje ilustrira naslednji primer, ki smo ga pri našem delu odkrili. Dva naša avtorja sta pred dobrimi tremi desetletji objavila knjigo, ki je bila od takrat citirana več kot sedemstokrat. Ta podatek pa je najti samo pri prvem avtorju. Če drugi avtor išče podatke o citiranosti pod svojim priimkom in imenom, tega citiranja ne bo odkril.

Poleg samostojnih publikacij so v skupini publikacij z nepopolnimi zapisi tudi članki in drugi prispevki v revijah. Teh je neprimerno več in lahko močno popačijo sliko citiranosti zlasti pri nekaterih avtorjih. V vseh revijah so to starejši članki, objavljeni pred uveljavitvijo računalnika. Ker so od tega pretekla že kar dobra tri desetletja, je za večino današnjih raziskovalcev to nepomembno, saj nimajo publikacij s tega obdobja. V skupini z nepopolnim zapisom je najti tudi nekaj primerov, kjer so v citirajočih člankih podatki o citiranem članku napačni ali pomanjkljivi. Takih primerov ni dosti, pa še takrat gre po navadi samo za eno citiranje.

Poleg teh pa najdemo tudi druge članke z nepopolnimi zapisi. Delež člankov s takimi zapisi se precej razlikuje od revije do revije. V revijah s faktorjem vpliva nad 2, kot npr. Journal of Physiology (London), jih praktično ni. V revijah s faktorjem vpliva okrog 1, kot je npr. Journal of Physiology (Pariz), je nepopolnih zapisov okrog 25% vseh citiranih člankov. Od treh slovenskih revij, ki so vključene v zbirko SCI Expanded, spadajo v skupino nepopolnih zapisov vsi citirani članki iz revij Strojniški vestnik in Informacije MIDEM. V reviji Acta Chimica Slovenica, ki je tudi vključena v omenjeno zbirko, pa ima nepopolne zapise okrog 38% vseh citiranih člankov. Pogledali smo tudi, kako je z medicinskimi revijami, ki izhajajo v Sloveniji. Med njimi smo članke s popolnimi zapisi našli le pri Zdravniškem vestniku, vendar je tudi tu okrog 60% člankov z nepopolnimi zapisi. Revije Radiology & Oncology, Acta der-

matovererologica APA, Zobozdravstveni vestnik, Medicinski razgledi in Zdravstveno varstvo imajo vse citirane članke v skupini z nepopolnim zapisom.

Na primeru Zdravniškega vestnika, za katerega imamo podrobnejše podatke o posameznih prispevkih, smo skušali ugotoviti, kaj bi lahko vplivalo na to, ali je posamezen članek razvrščen v skupino nepopolnih ali v skupino popolnih zapisov. Najprej smo preverili, ali so v skupini člankov s popolnim zapisom le znanstveni članki. Ugotovili smo, da to ne drži, saj so v tej skupini tudi drugi prispevki, med njimi celo sklepi skupščine Slovenskega zdravniškega društva. Po drugi strani pa so v skupini člankov z nepopolnim zapisom tudi številni članki, ki jih je uredništvo revije opredelilo kot znanstvene. Kaže torej, da kategorija članka ni bistveno vplivala na njegovo razvrstitev.

Nadalje smo skušali ugotoviti, ali se je razvrščanje citiranih člankov, objavljenih v Zdravniškem vestniku, v skupino popolnih zapisov kaj spreminjalo v času. Citirane članke, objavljene v Zdravniškem vestniku, smo uredili po letu njihove objave. Izid je podan v naslednji tabeli.

Leto objave	Število člankov z nepopolnim zapisom	Število člankov s popolnim zapisom	Skupaj
-1969	29	0	29
1970-1974	18	9	27
1975-1979	16	11	27
1980-1984	9	50	59
1985-1989	1	19	20
1990-1994	25	13	38
1995-1999	29	0	29
2000-2004	15	0	15
Skupaj	142	102	244

Razmerje med članki v skupini z nepopolnimi in v skupini s popolnimi zapisi ni bilo ves čas enako. Nekako do konca osemdesetih let je delež popolnih zapisov naraščal, nato pa je začel upadati. Zadnjih deset let ni bilo v skupini s popolnimi zapisi nobenega citiranega članka več, ampak so bili vsi v skupini z nepopolnimi zapisi.

Težave pa niso samo z nepopolnimi zapisi. Ugotovili smo namreč, da v nekaterih primerih izpadejo posamezni avtorji v člankih, ki so sicer v skupini s popolnim zapisom. Če iščemo

podatke po priimku in začetnici imena takega avtorja, podatke o citiranosti ne dobimo, dobimo pa ga pod priimki in imeni drugih avtorjev. To je najti predvsem pri avtorjih (pretežno avtoricah) z dvojnimi priimki.

Da bi ocenili, kako napaka zaradi izpada posameznega avtorja pri citiranju vpliva na rezultat analize citiranosti tega avtorja, smo podrobneje preverili podatke sedmih avtoric z dvojnimi priimki, katerih dela so sorazmerno dobro citirana. Za vsako avtorico smo iz WoS izpisali podatke o citiranih člankih, pri čemer smo upoštevali vse možne verzije zapisa njihovih priimkov in imen. Poleg standardnega zapisa obeh priimkov skupaj smo nekaj podatkov dobili tudi pod prvim priimkom in začetnico imena, nekaj pod drugim priimkom in začetnico imena z začetnico prvega priimka, nekaj pod zamenjanim zaporedjem obeh priimkov, nekaj pa celo pod drugim priimkom in začetnico imena. Iz zbirke Biomedicina Slovenica smo nato za vsako avtorico poiskali podatke o citiranosti, ki so bili poiskani po prvem avtorju. V vseh primerih, kjer je bila citiranost ugotovljena po prvem avtorju, po drugem pa ne, smo preverili, ali je bil res izpuščen samo ta avtor, drugi avtorji tega članka pa ne.

Ugotovili smo, da se je delež citiranih člankov, ki niso bili zajeti, če smo podatke iskali po avtorici z dvojnimi priimki, gibal med 26 do 52% vseh citiranih člankov posamezne avtorice, povprečje pa je znašalo 36%. Pri tem nismo upoštevali člankov, ki so izpadli zaradi nepopolnih zapisov. Takih člankov je bilo pri vseh sedmih avtoricah v povprečju 5%. Kaj je vzrok za tako pogosto izpadanje avtoric z dvojnimi priimki, nismo mogli ugotoviti. Vsekakor odgovora ni najti med navodili za delo, ki so dostopni na WoS.

Rezultati našega preverjanja kažejo, da lahko kar znaten del citiranih člankov ostane neodkrit, če podatke o citiranosti v servisu WoS iščemo po posameznem in ne po prvem avtorju publikacije. To velja tako v primerih, ko gre za zelo citirane knjige ali prispevke v njih, kot tudi za članke v revijah, zlasti domačih. Ko gre za avtorice z dvojnimi priimki, pa je zadeva še resnejša. Po našem preverjanju te avtorice lahko izgubijo od četrte do polovice vseh citiranih člankov, če podatke iščejo samo pod svojimi osebnimi podatki.

Vsem opisanim težavam se izognemo, če članke analiziramo po prvem avtorju, kot je to urejeno v zbirki Biomedicina Slovenica.