

NIKOLAJ LAZIĆ
Filozofski fakultet Sveučilišta u Zagrebu
nlazic@mudrac.ffzg.hr

ULOGA KLASIFIKACIJSKOGA SUSTAVA KOD PREUZIMANJA I OBRADJE PODATAKA ZA BIBLIOMETRIJSKA ISTRAŽIVANJA

Sažetak

Rezultati kvantitativnih i bibliometrijskih istraživanja, kao i odluke koje na temelju tih rezultata mogu proizaći i utjecati na određene politike, uvelike ovise o kvaliteti ulaznih podataka preuzetih iz bibliografskih i citatnih baza. Ova je tvrdnja izravno povezana s velikim brojem bibliometrijskih i scientometrijskih istraživanja koja nekritički preuzimaju podatke iz bibliografskih i citatnih baza *WoS – Web of Science* ili *Scopus*. U ovome istraživanju, koje je rađeno u okviru hrvatskoga nacionalnog projekta *Research activity, collaboration and orientation in social sciences in Croatia and other post-socialist European countries – RACOSS*, voditeljice dr. sc. Maje Jokić, analizirana je uloga klasifikacijskoga sustava kod preuzimanja i obrade podataka, odnosno donošenja zaključaka. Preuzeti bibliografski zapisi, objavljeni u 4 896 časopisa, rezultat su strategije pretraživanja znanstvene produktivnosti i citiranosti znanstvenika iz društvenih znanosti u 15 europskih postsocijalističkih zemlja (Bugarska, Hrvatska, Češka, Estonija, Mađarska, Latvija, Litva, Poljska, Rumunjska, Slovačka, Slovenija, Bosna i Hercegovina, Sjeverna Makedonija, Crna Gora i Srbija) od 1996. do 2013. godine. *Scopus* svoju klasifikaciju znanosti temelji na predmetnoj i sadržajnoj orijentaciji indeksiranih časopisa. Različitost klasifikacijskoga sustava koji koristi *Scopusovu* klasifikaciju određene zemlje može dovesti do krive interpretacije na temelju prikupljenih podataka. U radu smo napravili analizu časopisa i radova, a us-

poredna analiza pripadnosti pojedinoga časopisa području društvenih znanosti rađena je na temelju *Scopus*a i hrvatske klasifikacije društvenih znanosti, koja se temelji na *OECD Frascati fields of science*. Nakon što su predmetni stručnjaci provjerili časopise i usporedili ih s klasifikacijskim sustavom koji se koristi u Hrvatskoj, početni uzorak za pouzdane bibliometrijske analize smanjio se na udio od 44 %. Pripadnost pojedinomu području rađena je na temelju predmetne i sadržajne orijentacije indeksiranih časopisa. Da bi se dobili radovi koji pripadaju samo društvenim znanostima, potrebno je, osim filtriranja pojedinih znanstvenih polja, i ručno provjeriti pripadnost časopisa određenomu području. Tako se, osim bitno manjega broja radova i časopisa, pojavila i bitna razlika i u citiranosti.

Cilj je ovoga istraživanja upozoriti istraživače na moguće i stvarne nedostatke ulaznih podataka koji se koriste za vrjednovanje na primjeru uzorka baze *Scopus*, koja je relativno pouzdan izvor podataka za bibliometrijska istraživanja.

Ključne riječi: Obrada podataka; bibliometrijska istraživanja; klasifikacijski sustavi

Uvod

Rezultati kvantitativnih i bibliometrijskih istraživanja, kao i odluke koje na temelju tih rezultata mogu proizaći i utjecati na određene politike, uvelike ovise o kvaliteti ulaznih podataka preuzetih iz bibliografskih i citatnih baza. U tome kontekstu posljednjih se godina često spominje i problem metodologije vrjednovanja znanstvenika, ali i sveučilišta (Waltman i dr., 2012). Većina svjetskih sustava za vrjednovanje sveučilišta podatke o produktivnosti i citiranosti sveučilišta preuzima iz *WoS*-a ili iz *Scopus*a (Jokić – Petrušić, 2016). Polazi se od pretpostavke da su ti podatci pouzdani i točni. Isto tako, velik broj bibliometrijskih istraživanja koja se temelje na podacima iz spomenutih dviju baza podataka, a odnose se na vidljivost i prepoznatljivost mjerenu kroz produktivnost i citiranost autora, pojedinih institucija, sveučilišta, disciplina, zemalja i regija, često ne dovode u pitanje pouzdanost preuzetih podataka.

To se odnosi i na druge izvore podataka, npr. nacionalne bibliografske baze podataka (Rodela, 2016), ali i izvore tipa *Ulrichsweb* (Bachand – Sawallis, 2003; Grimes – Morris, 2006) ili kataloge knjižnica (Black,

2010). U literaturi se nailazi na radove vezane uz problem pouzdanosti spomenutih podataka, ali rjeđe na opsežna istraživanja o utjecaju pouzdanosti preuzetih podataka na donošenje zaključaka i odluka. Istraživanja koja su rađena pomoću klasifikacije grešaka, koja je preuzeta od Buchanan (2006) i Olensky (2014), iskorištena su za analizu pouzdanosti bibliografskih zapisa preuzetih iz *WoS*-a. Analizirano je postoje li tipske greške za neke dokumente odnosno jezike. Pregledavajući literaturu o pouzdanosti baza podataka, mogu se izdvojiti Wang – Waltman (2016) koji kažu da ni jedna studija nije sustavno istražila točnost klasifikacijskih sustava. Razvrstavanje časopisa u istraživačka područja važan je predmet bibliometrijskih studija. Dvije najvažnije baze podataka, *Web of Science* i *Scopus*, pružaju sustav klasifikacije časopisa, ali se on razlikuje od sustava klasifikacije u hrvatskoj znanstvenoj klasifikaciji.

U ovome istraživanju, koje je rađeno u okviru hrvatskoga nacionalnog projekta *Research activity, collaboration and orientation in social sciences in Croatia and other post-socialist European countries – RACOSS*, prvi indikator u analizi pouzdanosti bibliografskih zapisa preuzetih iz baze *Scopus* bila je klasifikacija časopisa. Bazu *Scopus* izabrali smo kao relevantniju za područje društvenih znanosti u odnosu na *WoS*, i to na temelju našega iskustva kao i iskustva drugih istraživača. Prema Bartol i dr. (2014) čini se da *Scopus* pruža opsežnije pokriće i više citata u društvenim i humanističkim znanostima pa bi mogao biti povoljniji za uravnoteženiju procjenu ovih područja.

Društvene znanosti nisu homogena cjelina i često se pod tim pojmom uključuju i pojedina znanstvena polja i discipline iz ostalih znanstvenih područja, od humanistike do tehnologije. Koja znanstvena polja i discipline pokriva područje društvenih znanosti (*social sciences*) – ovisi o klasifikacijskoj shemi koju određeni izvor koristi. U bibliometrijskim istraživanjima nije svejedno koja znanstvena polja i discipline pokriva područje društvenih znanosti. Ako u njemu dominiraju znanstvena polja čija su metodologija i paradigma znanstvenoga komuniciranja srodnija prirodnim ili tehničkim znanostima, za razliku od klasifikacije koja uključuje polja srodnija paradigmi znanstvenoga komuniciranja u humanističkim znanostima, usporedni rezultati mogu dati različite slike

stanja. Stoga se u ovome istraživanju analitički uspoređuje *Scopusova* klasifikacija društvenih znanosti s hrvatskom klasifikacijom koja se temelji na *OECD Frascati* klasifikaciji društvenih znanosti.

Cilj je ovoga istraživanja upozoriti istraživače na moguće i stvarne nedostatke ulaznih podataka na primjeru uzorka iz baze *Scopus*, koja je pouzdan izvor podataka za bibliometrijska istraživanja.

1. Uzorak i metodološki pristup

Kao što je već spomenuto, uzorak bibliografskih zapisa za ovo istraživanje dobiven je pretraživanjem baze *Scopus*. Strategijom pretraživanja bilo je važno dobiti sve znanstvene radove iz područja društvenih znanosti čiji su autori bili znanstvenici iz 15 postsocijalističkih europskih zemalja (Estonija, Latvija, Litva, Poljska, Češka Republika, Mađarska, Rumunjska, Bugarska, Slovenija, Hrvatska, Srbija, Bosna i Hercegovina, Crna Gora, Sjeverna Makedonija) od 1996. do 2013. godine. Stoga je pretraživanje ograničeno na polje *Affiliation country* u koje je unijom uključeno svih navedenih 15 zemalja, a zatim je ograničeno razdoblje od 1996. do 2013. i kategorija Društvene znanosti. Kod *Affiliation countryja* unija naziva zemalja omogućila je dobivanje svih radova kada je bilo koja od 15 uključenih zemalja bila registrirana barem jednom na bilo kojem radu. Budući da je važno dobiti što precizniju znanstvenu aktivnost u području *Social sciences*, daljnji je korak u pretraživanju ograničavanje skupa podataka samo na znanstvene radove: *article, article in press and review*. Ekstrakcijom samo znanstvenih radova dobiven je početni uzorak za projektnu bazu podataka od 83 059 radova. Iako, da je za projektnu bazu uzet navedeni skup radova i da je na njemu rađena bibliometrijska analiza, sigurno bi dobili bitno različite podatke o produktivnosti i citiranosti, odnosno o statusu društvenih znanosti navedene skupine europskih zemalja.

Kako bi se dobila što cjelovitija slika stanja društvenih znanosti izabranoga uzorka zemalja, odnosno specifičnosti pojedinih znanstvenih polja unutar područja društvenih znanosti, odlučeno je da će se analizirati kroz časopise u kojima su ti radovi objavljeni. Društvene znanosti

nisu homogena cjelina, a jedno od osnovnih istraživačkih pitanja ovog projekta bilo je utvrđivanje sličnosti i razlika između pojedinih polja društvenih znanosti. Već iz prvih analiza časopisa iz dobivena uzorka mogu se izdvojiti časopisi koji se uglavnom bave problematikom prirodnih ili tehničkih znanosti, odnosno područjem humanistike, a malim udjelom i društvenih znanosti. Za neke se časopise u dobivenome skupu, koji se gotovo u cijelosti bavi problematikom iz humanistike, moglo i naći opravdanje jer je u vrijeme kada smo preuzimali podatke iz *Scopus*a (listopad 2015.) postojala zadana opcija za ograničavanje pretraživanja na samo četiri glavna područja: *Health sciences*, *Life sciences*, *Physical sciences* i *Social sciences*. Kategorija *Social sciences* obuhvaćala je i *Humanities* i *Arts*. Međutim, kako se neka polja društvenih znanosti i humanistike bitno razlikuju u znanstvenome komuniciranju, koriste i različite medije za objavljivanje (*monographs*, *books*, *book chapters*, itd.). Budući da je dinamika objavljivanja različita, a ponašanje kod citiranja drugačije, samim time su i rezultati bibliometrijskih analiza različiti. Stoga je odlučeno da će se pristupiti detaljnoj provjeri svakoga uključenog časopisa. Radilo se o 4 896 časopisa čiji su sadržaji činili početni skup od spomenutih 83 059 radova.

Budući da je ovo istraživanje primarno rađeno s ciljem analiziranja stanja društvenih znanosti u Hrvatskoj s naglaskom na usporedbu s ostalim srodnim postsocijalističkim europskim zemljama, odlučeno je da će se u kategorizaciji časopisa koristiti hrvatski sustav klasifikacije društvenih znanosti (Tablica 1.).

Tablica 1.: *Scopus*, *OECD* and *Croatian Social sciences classification schema*

<i>Scopus classification</i>	<i>OECD classification</i>	<i>Croatian classification</i>
<i>Social Sciences</i>	<i>5. Social Sciences</i>	<i>5. Social sciences</i>
<i>Social Sciences(all)</i>	<i>5.1 Psychology</i>	<i>5.01. Economics</i>
– <i>Social Sciences (miscellaneous)</i>	– <i>Psychology (including human - machine relations);</i>	– <i>Economics, Econometrics; Industrial relations;</i>

<ul style="list-style-type: none"> – Archaeology – Development – Education – Geography, Planning and Development – Health (social science) – Human Factors and Ergonomics – Law – Library and Information Sciences – Linguistics and Language – Safety Research – Sociology and Political Science – Transportation – Anthropology – Communication – Cultural Studies – Demography – Gender Studies – Life-span and Life-course Studies – Political Science and International Relations – Public Administration – Urban Studies Business, Management and Accounting (all) <ul style="list-style-type: none"> – Business, Management and Accounting (miscellaneous) – Accounting – Business and International Management – Management Information Systems 	<ul style="list-style-type: none"> – Psychology, special (including therapy for learning, speech, hearing, visual and other physical and mental disabilities); 5.2 Economics and Business <ul style="list-style-type: none"> – Economics, Econometrics; Industrial relations; – Business and Management 5.3 Educational sciences <ul style="list-style-type: none"> – Education, general; including training, pedagogy, didactics; – Education, special (to gifted persons, those with learning disabilities); 5.4 Sociology <ul style="list-style-type: none"> – Sociology; Demography; Anthropology, ethnology, – Social topics (Women's and gender studies) 5.5 Law <ul style="list-style-type: none"> – Law, criminology, penology; 5.6 Political science <ul style="list-style-type: none"> – Political science; public administration; organisation theory; 	<ul style="list-style-type: none"> – Business and Management 5.02. Law 5.03. Political Science 5.04. Information and Communication science 5.05. Sociology 5.06. Psychology 5.07. Pedagogy 5.08. Special education 5.09. Logopedy¹ 5.10. Kinesiology² 5.11. Demography 5.12. Social work 5.13. Security studies 5.14. Interdisciplinary social sciences
--	--	---

¹ Logopedy smo isključili jer po metodologiji i pristupu istraživanju spada više u biomedicinske znanosti

² Kao u slučaju logopedy, i kinesiology po našem mišljenju i iskustvu više spada u biomedicinske znanosti.

<ul style="list-style-type: none"> – <i>Management of Technology and Innovation</i> – <i>Marketing</i> <i>Economics, Econometrics and Finance (all)</i> – <i>Economics, Econometrics and Finance (miscellaneous)</i> – <i>Economics and Econometrics</i> – <i>Finance</i> <i>Psychology (all)</i> – <i>Psychology (miscellaneous)</i> – <i>Applied Psychology</i> – <i>Clinical Psychology</i> – <i>Developmental and Educational Psychology</i> – <i>Experimental and Cognitive Psychology</i> – <i>Neuropsychology and Physiological Psychology</i> – <i>Social Psychology</i> 	<ul style="list-style-type: none"> <i>5.7 Social and economic geography</i> – <i>Environmental sciences (social aspects); Cultural and economic geography; Urban studies (Planning and development); Transport planning and social aspects of transport (transport engineering to be 2.1);</i> <i>5.8 Media and communications</i> – <i>Journalism; Information science (social aspects); Library science; Media and socio-cultural communication;</i> <i>5.9 Other social sciences</i> – <i>Social sciences, interdisciplinary;</i> – <i>Other social sciences</i> 	
--	--	--

Važno je napomenuti da se službena klasifikacija znanosti u Hrvatskoj razlikuje od klasifikacije koju koristi *Scopus*. Prvi korak u određivanju što pouzdanijih podataka vezanih za znanstvenu aktivnost u području društvenih znanosti u 15 europskih postsocijalističkih zemalja bilo je utvrđivanje sličnosti odnosno razlika između *Scopusove* klasifikacije društvenih znanosti i službene klasifikacije društvenih znanosti u Hrvatskoj. *Scopusova* klasifikacija i hrvatska klasifikacija društvenih znanosti bitno se razlikuju (Tablica 1.). *Scopus* u društvene znanosti uključuje npr. arheologiju, lingvistiku, jezik i antropologiju koje hrvatska klasifikacija znanosti svrstava u humanističke znanosti. *Scopusovo* polje *Health (social science)* svrstava u *Social sciences*, a u hrvatskoj klasifikaciji znanosti spada u područje biomedicine i zdravstva. Polja *Geography, Planning and Development* i *Urban studies*, koje *Scopus* svrstava u

društvene znanosti, po hrvatskoj klasifikaciji spadaju u interdisciplinarna područja ili čak u tehničko područje odnosno prirodne znanosti. Polje *Transportation Scopus* svrstava u društvene znanosti, dok se u hrvatskoj klasifikaciji primarno svrstava u tehničke znanosti. Ove razlike bitno mogu utjecati na rezultate i dati iskrivljenu sliku stanja društvenih znanosti u istraživanome uzorku, prvenstveno zbog razlika u znanstvenome komuniciranju, preciznije u citiranju i objavljivanju. S metodološkoga aspekta važno je naglasiti da bi klasificiranje svakoga pojedinačnog rada u časopisu bilo znatno preciznije, no to je iz pragmatičnih razloga prilično teško izvedivo.

Scopus kao i ostale baze podataka svoju klasifikaciju znanosti odnosno znanstvene predmetne kategorije temelji na časopisima. *Scopusova* klasifikacija časopisa naziva se *All Science Journal Classification (ASJC)*³. Prema Wangu – Waltmanu (2016), znanje o točnosti *Scopusovih* klasifikacijskih sustava vrlo je ograničeno, a točnost klasifikacijskoga sustava može bitno utjecati na bibliometrijske studije. Navedene tvrdnje i iskustvo u radu s bibliografskim i citatnim bazama potakli su detaljnija istraživanja o tome koliko klasifikacijski sustav baze podataka temeljen na časopisima može utjecati na rezultate bibliometrijskih istraživanja.

Dotadni je problem sa *Scopusovom* klasifikacijom društvenih znanosti to što su u društveno područje uključeni časopisi koji se gotovo isključivo bave problematikom prirodnih znanosti ili biomedicine. Kao primjer imamo časopise *Nature*, *Science*, *Proceedings of the National Academy of Sciences of the United States of America*, *Fuzzy Sets and Systems*. U našem uzorku iz ta četiri časopisa bilo je 1 799 ili 2,16 % radova na kojima je barem jedan od koautora bio iz najmanje jedne od 15 zemalja iz našega uzorka. Ti su radovi dobili ukupno 200 694 citata, što je udio od 36,1 % ukupna broja dobivenih citata na radove objavljene u 4 896 časopisa. Važno je naglasiti da se većina radova iz tih časopisa tematski ne bavi područjem društvenih znanosti. Prosječan broj citata po časopisu iz početnoga *Scopusova* uzorka iznosio je 113,4 citata. Ako izlučimo četiri navedena časopisa, čiji se radovi teško mogu svrstati u

³ „Scopus ASJC (All Science. Journals Classification) classification codes“, <https://artax.karlin.mff.cuni.cz/r-help/library/CITAN/html/Scopus_ASJC.html>, (31. 3. 2017.).

društvene znanosti, tada je prosječan broj citata po časopisu bitno niži i iznosi 72,5 citata. Ovaj primjer navodimo samo kao ilustraciju i potvrdu opravdanja da nastavimo detaljniju empirijsku analizu.

Prema našem mišljenju, objektivniji način dobivanja pouzdanih podataka o produktivnosti, vidljivosti, prepoznatljivosti i citiranosti rezultata istraživanja iz društvenih znanosti u 15 europskih postsocijalističkih zemalja bio je pristup provjere i reklasifikacije časopisa ($n = 4896$) koji su sadržavali naš početni uzorak preuzet iz *Scopus* – 83 059 radova. Napravljene su analiza i reklasifikacija časopisa koje su obavljali predmetni stručnjaci iz pojedinih polja društvenih znanosti prema službenoj hrvatskoj klasifikaciji znanosti: Ekonomija, Edukacijsko-rehabilitacijske znanosti, Informacijske i komunikacijske znanosti, Pravo, Politologija, Psihologija, Sociologija i Socijalne djelatnosti. Očekivano, dio časopisa iz našega skupa nije bilo moguće svrstati u samo jedno znanstveno polje.

Od 4 896 časopisa iz početnoga uzorka predmetni su stručnjaci izlučili 2 744 (56 %) časopisa koje su u opsegu definirali da se primarno bave problematikom društvenih znanosti prema navedenoj hrvatskoj klasifikaciji društvenih znanosti. Preostalih 2 152 ili 44 % časopisa koje je *Scopus* svrstao u *Social sciences*, a prema hrvatskoj klasifikaciji znanosti ne pripadaju u to područje, sadržavali su 47 406 radova ili 57,1 % radova. Ono što bi u ovome postupku dalje trebalo napraviti je dodatnom reklasifikacijom taj popis časopisa uskladiti prema hrvatskoj, odnosno *OECD*-evoj klasifikaciji znanosti. Na taj bismo način prikazali u kojoj se mjeri slika stanja produktivnosti i citiranosti radova autora iz 15 europskih postsocijalističkih zemalja, koje je *Scopus* svrstao u *Social sciences*, razlikuje u odnosu na sliku stanja nakon reklasifikacije časopisa prema hrvatskoj klasifikaciji toga znanstvenog područja.

2. Rezultati

Da smo prema opisanoj strategiji pretraživanja preuzeli dobivene radove iz *Scopus* i napravili analizu broja radova i citata, dobili bismo podatak da je prosječan broj radova po časopisu od 1996. do 2013. bio

17. Prosječna citiranost radova bila bi 6,68 citata po radu, dok je raspon broja citata po radu bio od 5 771 (Dietl i dr., 2000) do 0. Teško je očekivati u bilo kojem polju društvenih znanosti, uključujući i psihologiju, da rad dobije preko 5 000 citata. Isto tako, teško je očekivati i to da časopisi *Science* i *Nature* uključuju relativno velik broj radova iz postsocijalističkih zemalja, a naročito iz društvenih znanosti. Stoga je za dobivanje stvarnije slike stanja društvenih znanosti u našem uzroku nužno bilo provesti reklasifikaciju časopisa.

Skup je časopisa ($N = 2\,744$) nakon uvida u njihove *aims and scopes* te na temelju prosudbe predmetnih stručnjaka (*economics, education, information and communication sciences, law, political sciences, psychology, sociology and social work*), potvrđen kao skup časopisa koji se bave problematikom *social sciences* i sadržavao je tek 42,9 % ($N = 35\,653$) radova u odnosu na skup koji je *Scopus* klasificirao kao područje *Social sciences*. Prosječan broj radova po časopisu iznosi 13, a prosječan broj citata po radu bio je 5,1, dok je raspon broja citata po radu od 1572 do 0. Iako prosječan broj citata po radu u ovome podskupu na prvi pogled nije bitno manji u odnosu na početni *Scopusov* uzorak, iz Tablice 2 jasno je vidljivo koja znanstvena polja doprinose u osnovi iskrivljenu prosjeku. Ako se izuzmu časopisi iz psihologije, koja je po metodologiji kao i znanstvenome komuniciranju srodnija prirodnim znanostima, tada je za društvene znanosti prosječan broj citata po radu 3,31. Detaljniji uvid u stanje produktivnosti i citiranosti u pojedinim poljima društvenih znanosti iz našega uzorka dostupan je u Tablici 2.

Tablica 2. Distribucija broja časopisa, radova i citata u skupu časopisa koje su predmetni stručnjaci klasificirali u područje društvenih znanosti

<i>Subject field</i>	<i>N journals</i>	<i>N papers</i> (1996 – 2013)	<i>N citations</i> (1996 – 2015)	<i>citation/paper</i>
<i>Economics</i>	695	11 991	48 360	4,03
<i>Education</i>	295	2 935	9 785	3,33
<i>Information and communication sciences</i>	163	2 239	19 077	8,52
<i>Law</i>	89	885	1 088	1,23

<i>Political sciences</i>	158	1 224	5 098	0,41
<i>Psychology</i>	289	4 045	48 409	108,78
<i>Sociology</i>	56	1 472	3 446	2,34
<i>Social work</i>	27	524	1 295	2,49
<i>Multidisciplinary social sciences</i>	559	5 662	24 495	4,32
<i>Multidisciplinary social science & humanities</i>	155	2 222	3 968	1,78
<i>Multidisciplinary social science & other science fields (medicine, natural sciences, technology etc.)</i>	258	4 508	32 458	7,2

Drugomu skupu časopisa (N = 2174), izlučenu iz početnoga skupa od 4 896 časopisa, koje je *Scopus* definirao kao časopise koji pripadaju području *Social sciences*, predmetni su stručnjaci označili upitnom pripadnosti istomu području. U tome skupu časopisa nalazi se više od polovice radova (57,1 %) iz našega uzorka.

Zaključak

Svrha je ovoga istraživanja upozoriti na oprez kod preuzimanja bibliografskih podataka iz baze *Scopus*, ali i ostalih izvora, za bibliometrijska istraživanja. Kako se većina bibliometrijskih istraživanja radi i za potrebe znanstvene politike i vrjednovanja znanstvenika, tada zaključci doneseni na temelju tih istraživanja mogu imati određene posljedice. Problem klasifikacije časopisa, odnosno radova, prilično je složeno pitanje, a pogotovo kada se vrjednovanje vrši prema formalnim kriterijima koji onda previde probleme s klasifikacijom koja se koristi.

Literatura

- BACHAND, ROBERT G. – SAWALLIS, PAMELA P. (2003) „Accuracy in the Identification of Scholarly and Peer-Reviewed Journals and the Peer-Review Process Across Disciplines“, *The Serials Librarian*, god. XLV, br. 2, str. 39 – 59.

-
- BARTOL, TOMAZ – BUDIMIR, GORDANA – JUZNIC, PRIMOZ – STOPAR, KARMEN (2016) „Mapping and classification of agriculture in Web of Science: other subject categories and research fields may benefit“, *Scientometrics*, god. CIX, br. 2, str. 979 – 996.
 - BARTOL, TOMAZ – BUDIMIR, GORDANA – DEKLEVA-SMREKAR, DORIS – PUSNIK, MIRO – JUZNIC, PRIMOZ (2014) „Assessment of research fields in Scopus and Web of Science in the view of national research evaluation in Slovenia“, *Scientometrics*, god. XCVIII, br. 2, str. 1491 – 1504.
 - BLACK, STEVE (2010) „Failure Rates and Publication Status: Periodicals Reviewed in Library Journal (1980-2005) and Database Accuracy“, *Serials Review*, god. XXXVI, br. 4, str. 206 – 213.
 - BORNMANN, LUTZ – LEYDESDORFF, LOET (2013) „Macro-Indicators of Citation Impacts of Six Prolific Countries: InCites Data and the Statistical Significance of Trends“, *PLoS ONE*, god. VIII., br. 2.
 - BUCHANAN, ROBERT A. (2006) „Accuracy of cited references: The role of citation databases“, *College and Research Libraries*, god. LXVII, br. 4, str. 292 – 303.
 - DIETL, T. – OHNO, H. – MATSUKURA, F. – CIBERT, J. – FERRAND, D. (2000) „Zener model description of ferromagnetism in zinc-blende magnetic semiconductors“, *Science*, god. CCLXXXVII, br. 5455, str. 1019 – 1022.
 - FRANCESCHINI, FIORENZO – MAISANO, DOMENICO – MASTROGIACOMO, LUCA (2016a) „Empirical analysis and classification of database errors in Scopus and Web of Science“, *Journal of Informetrics*, god. X, br. 4, str. 933 – 953.
 - FRANCESCHINI, FIORENZO – MAISANO, DOMENICO – MASTROGIACOMO, LUCA (2016b) „Do Scopus and WoS correct 'old' omitted citations?“, *Scientometrics*, god. CVII, br. 2, str. 321 – 335.
 - GAVEL, YLVA – ISELID, LARS (2008) „Web of Science and Scopus: A journal title overlap study“, *Online Information Review*, god. XXXII, br. 1, str. 8 – 21.

- GORRAIZ, JUAN – MELERO-FUENTES, DAVID – GUMPENBERGER, CHRISTIAN – VALDERRAMA-ZURIÁN, JUAN CARLOS (2016) „Availability of digital object identifiers (DOIs) in Web of Science and Scopus“, *Journal of Informetrics*, god. X, br. 1, str. 98 – 109.
- GRIMES, MARYBETH – MORRIS, SARA E. (2006) „Is accuracy everything? A study of two serials directories“, *Reference & User Services Quarterly*, god. XLVI, br. 2, str. 45 – 49.
- JOKIĆ, MAJA – PETRUŠIĆ, IRENA (2016) „Neki od uzroka slabe zastupljenosti hrvatskih sveučilišta na svjetskim rang ljestvicama sveučilišta“, *Medijska istraživanja*, god. XXII, br. 1, str. 5 – 40.
- KAWASHIMA, HIROTAKA – TOMIZAWA, HIROYUKI (2015) „Accuracy evaluation of scopus author ID based on the largest funding database in Japan“, *Scientometrics*, god. CIII, br. 3, str.1061 – 1071.
- LASDA BERGMAN, ELAINE M. (2012) „Finding citations to social work literature: The relative benefits of using Web of Science, Scopus, or Google Scholar“, *The Journal of Academic Librarianship*, god. XXXVIII, br. 6, str. 370 – 379.
- MOED, HENK F. – BAR-ILAN, JUDIT – HALEVI, GALI (2016) „A new methodology for comparing Google Scholar and Scopus“, *Journal of Informetrics*, god. X, br. 2, str. 533 – 551.
- OLENSKY, MARLIES (2014) *Data Accuracy in bibliometric data sources and its impact on citation matching*, doktorska disertacija, Humboldt-Universität zu Berlin, Berlin.
- RODELA, ROMINA (2016) „On the use of databases about research performance: comments on Karlovčec and Mladenčić (2015) and others using the SICRIS database“, *Scientometrics*, god. CIX, br. 3, str. 2151 – 2157.
- VALDERRAMA-ZURIÁN, JUAN CARLOS – AGUILAR-MOYA, REMEDIOS – MELERO-FUENTES, DAVID – ALEIXANDRE-BENAVENT, RAFAEL (2015) „A systematic analysis of duplicate records in Scopus“, *Journal of Informetrics*, god. IX, br. 3, str. 570 – 576.

- WALTMAN, LUDO – CALERO-MEDINA, CLARA – KOSTEN, JOOST – NOYONS, ED C. M. – TIJSSEN, ROBERT J. W. – VAN ECK, NEES JAN – VAN LEEUWEN, THED N. – VAN RAAN, ANTHONY F. J., VIS-SER, MARTIJN S., WOUTERS, PAUL (2012) „The Leiden ranking 2011/2012: Data collection, indicators, and interpretation“, *Journal of the American Society for Information Science and Technology*, god. LXIII, br. 12, str. 2419 – 2432.
- WANG, QI – WALTMAN, LUDO (2016) „Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus“, *Journal of Informetrics*, god. X, br. 2, str. 347 – 364.

Original scientific paper
Received on March 23, 2019.

NIKOLAJ LAZIĆ

University of Zagreb, Faculty of Humanities and Social Sciences

ROLE OF THE CLASSIFICATION SYSTEM IN DOWNLOADING AND PROCESSING DATA FOR BIBLIOMETRIC RESEARCH

Abstract

Results of quantitative and bibliometric research, as well as decisions that might be made on the basis of those results and influence on certain policies, greatly depend on the quality of input data downloaded from bibliometric and citation databases. This assertion is directly connected to a high number of bibliometric and scientometric research that uncritically download data from bibliographic and citation databases, *WoS* – *Web of Science* or *Scopus*. This research, which has been conducted in the framework of the Croatian national project *Research activity, collaboration and orientation in social sciences in Croatia and other post-socialist European countries* – *RACOSS*, lead by Maja Jokic, PhD, analyzes the role of the classification system in downloading and processing data, i.e. making conclusions. Bibliographic records, downloaded in the period from 1996 to 2013, published in 4896 journals, represent the result of strategy on searching the scientific productivity and citation index of social scientists from 15 European post-socialist countries (Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia, Bosnia and Herzegovina, North Macedonia, Montenegro and Serbia). Classification of science by *Scopus* is based on the subject and content orientation of indexed journals.

Diversity of classification system that uses the *Scopus* classification of a certain country can cause wrong interpretation on the basis of collected data. In this paper we made the analysis of journals and papers, and contrastive analysis of a certain journal affiliation to the field of social sciences was made on the basis of *Scopus* and Croatian classification

of social sciences based on *OECD Frascati fields of science*. After the experts in this area had checked the journals and compared them to the classification system used in Croatia, the initial sample for reliable bibliometric analyses decreased to the portion of 44%. Affiliation to a certain field was made on the basis of the subject and content orientation of the indexed journals. Besides filtration of scientific fields, it is necessary to check the affiliation of a journal to the certain field by hand in order to get papers that belong only to social sciences. In this way there appeared a significant difference in the citation index additionally to lower number of papers and journals.

The aim of this research is to warn researchers to possible and real deficiencies of input data used for evaluation on the example of the *Scopus* database, which is relatively reliable source of data for bibliometric research.

Keywords: data processing; bibliometric research; classification systems