

# Bibliometrijska analiza pomoću alata BibExcel i SciMAT

---

Škvorc, Božidar

Undergraduate thesis / Završni rad

2019

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:131:613742>

*Rights / Prava:* [In copyright](#) / [Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-28**



Sveučilište u Zagrebu  
Filozofski fakultet  
University of Zagreb  
Faculty of Humanities  
and Social Sciences

*Repository / Repozitorij:*

[ODRAZ - open repository of the University of Zagreb  
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
*ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI*  
Ak. god. 2018./2019.

Božidar Škvorc

**Bibliometrijska analiza pomoću alata BibExcel i SciMAT**

Završni rad

Mentor: dr. sc. Tomislav Ivanjko, docent

Zagreb, lipanj 2019.

## Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

---

(potpis)

# SADRŽAJ

Uvod .....	4
1. Povijesni razvoj bibliometrije .....	6
1.1. Pojava pojma bibliometrije.....	6
1.2. Scientometrija.....	7
1.3. Informetrija .....	8
1.4. Webometrija .....	8
1.5. Kritički osvrt na postojeće stanje-Informetrija kao zajednička disciplina .....	9
2. Bibliometrijski zakoni .....	10
2.1. Lotkin zakon o produktivnosti autora .....	11
2.2. Bradfordov zakon distribucije .....	14
2.3. Zipfov zakon učestalosti pojave riječi u tekstu.....	16
3. Faktor odjeka časopisa – primarno sredstvo vrednovanja časopisa .....	18
3.1. Journal citation report.....	18
3.2. Faktor odjeka – Impact factor.....	19
3.2.1. Faktor odjeka i citati .....	20
3.2.2. Faktor odjeka i samocitati.....	21
3.2.3. Indeks brzine citiranja i poluvrijeme citiranja .....	21
3.2.4. Bibliografske baze podataka kao sekundarni izvori informacija .....	22
4. Bibliometrijski alati:.....	24
4.1. BibExcel .....	26
4.1.1. Priprema i uvoz podataka.....	27
4.1.2. Kreiranje OUT datoteke .....	28
4.1.3. Osnovne analitičke funkcije BibExcel-a.....	30
4.1.4. Kreiranje Pajek datoteke u BibExcel-u .....	32
4.2. SciMAT – Science Mapping Analysis Software Tool.....	34
4.2.1. Glavne karakteristike SciMAT-a.....	35
4.2.2. Baza znanja .....	35
4.2.2.1. Upravitelj baze znanja .....	36
4.2.3. Čarobnjak za konfiguraciju analize znanstvenog mapiranja.....	39

4.2.4. Modul za vizualizaciju .....	40
4.3. Usporedna analiza bibliometrijskih alata BibExcel i SciMAT .....	42
Zaključak .....	43
Literatura .....	45
Sažetak.....	47
Summary .....	48

## UVOD

Ovaj rad nastao je u želji da se objasne neke osnovne nejasnoće oko definicija informetrije i bibliometrije kao discipline koja spada u uži dio informetrije, te da se u praktičnom djelu rada podastre usporedna analiza dvaju bibliometrijskih alata, BibExcel-a i SciMAT-a. Naime, Hood i Wilson u svome članku „*The literature of bibliometrics, scientometrics and informetrics*“ navode da se informetrija kao poddisciplina informacijskih znanosti bavi procesima pretraživanja informacija i da je to područje znatno šire u odnosu na područje bibliometrije kojega ćemo detaljnije obraditi u ovome radu (Hood i Wilson, 2001). Također je važno napomenuti da, iako se cijeli prvi, teorijski dio, rada temelji na temi bibliometrije kao discipline unutar informetrije, nije moguće, zbog isprepletenosti pojmova i samih disciplina, potpuno odijeliti pojam bibliometrije od pojmova informetrije, scientometrije i webometrije. Nadalje, neki od autora radova, a u prvome redu mislim na Williama W. Hooda, Concepción S. Willson, Vesnu Olujić Vuković te donekle na Maju Jokić, raspravljaju i donose zanimljive argumente po kojima bi se discipline informetrije, scientometrije i bibliometrije u dogledno vrijeme trebale ujediniti u informetriju kao jedinstvenu disciplinu. U daljnjem, teorijskom dijelu rada, objasniti ćemo povijesni razvoj, u prvom redu bibliometrije, a kasnije i informetrije, scientometrije i webometrije. Nakon povijesnog razvoja navedenih disciplina koncentrirati ćemo se na „*pravilnosti u objavljivanju i korištenju informacija*“, kako zakonitosti u bibliometriji definira Franjo Pehar u svome radu „*Od statističke bibliografije do bibliometrije*“ (Pehar, 2011). Te pravilnosti su naime tri glavna zakona korištena u bibliometriji po kojima se može s pristojnom vjerojatnošću odrediti znanstvena produktivnost autora nekoga rada i zastupljenost znanstvenih radova u znanstvenim časopisima. Pri kraju prvoga, teorijskog dijela rada, koncentrirati ćemo se na problem vrednovanja časopisa, gdje ćemo dublje ući u analizu „*faktora odjeka*“, (eng. impact factor), kako ga u svojoj knjizi prevodi sa engleskog jezika Maja Jokić, kao jednog od najpopularnijih pokazatelja statističke vrijednosti časopisa (Jokić, 2005). Važno je napomenuti da se ovdje faktor odjeka ne može kvalitetno objasniti bez da se detaljnije ne razrade pojmovi citat, samocitat i citatne baze podataka, tako da ćemo na kraju teorijskog dijela rada detaljnije ući u problematiku citiranja članaka u časopisima i obradu časopisa u citatnim bazama podataka koje obrađuju najutjecajnije svjetske izvore literature. Drugi, praktični dio rada ćemo započeti kratkim uvodom o bibliografskim

alatima BibExcel-u i SciMAT-u, a u nastavku ćemo se detaljnije koncentrirati najprije na važnije značajke alata BibExcel, a zatim i alata SciMAT. S obzirom da ovaj dio rada nije napravljen na temelju prave i originalne bibliometrijske analize, svaku pojedinu značajku oba navedena alata potkrijepiti ćemo slikama iz već postojeće literature vezane za maloprije spomenute bibliometrijske alate. Za sam kraj pokušati ćemo što jednostavnije i konciznije objasniti prednosti i nedostatke navedenih bibliometrijskih alata BibExcel-a i SciMAT-a.

# 1. POVIJESNI RAZVOJ BIBLIOMETRIJE

Da bismo uopće mogli početi govoriti o bibliometriji kao zasebnoj disciplini unutar informetrije potrebno je vratiti se malo dublje u prošlost i objasniti same početke popisivanja dokumenata. Bibliografija kao umijeće popisivanja dokumenata poznata je još u starom vijeku, pa sve do Antike i Rima. Tako se među prvima koji je pokušao popisati autore nekih djela pojavljuje Sv. Jeronim. On u djelu „*De viris illustribus*“ u četvrtom stoljeću objavljuje skup od 135 književnih biografija, većinom istaknutih crkvenih otaca čija su imena poredana abecednim redom (Pehar, 2011). Dakako da je pravi razvoj bibliografije moguće pratiti tek od pojave tiskarskog stroja u 15. stoljeću kada su se stvorile sve pretpostavke za širu primjenu tiska i distribuciju knjige. „*Do kraja 17. stoljeća bibliografija je uspjela izrasti u visokorazvijeno umijeće. U razdoblju između 17. i 19. stoljeća glavni predmet spoznaje bibliografije bila je knjiga, a sama bibliografija percipirana je kao znanost o knjizi*“. (Pehar, 2011). U 19. stoljeću vidljivo je snažnije približavanje bibliografije ka novonastalim znanstvenim disciplinama kao što su historiografija, knjižničarstvo, dokumentalistika, informacijske znanosti, bibliometrija i ostale metrijske discipline. Prvo sustavno korištenje statističkih podataka o znanosti analiziranih tijekom dužeg vremenskog razdoblja povezujemo sa djelovanjem američkog psihologa Jamesa McKeena Catella (1860-1944) koji je u časopisu Science, čiji je bio dugogodišnji urednik, pokušao podastrijeti podatke o autorima članaka u navedenom časopisu (Pehar, 2011). U njegovu fokusu bila su dva statistička pristupa. Mjerenje kvantitete (produktivnosti) autora i kvalitete (postignuća) autora. Produktivnost je odredio kao broj znanstvenika određene nacije, države, grada. Postignuće je definirao kao „mjeru doprinosa napretku znanosti“, no za razliku od današnje prakse mjerenja produktivnosti i postignuća znanstvenika na temelju broja objavljenih publikacija ili primljenih citata, koju ćemo u nastavku rada pobliže objasniti, Catell je svoju metodu temeljio na mišljenjima recenzenata.

## 1.1. POJAVA POJMA BIBLIOMETRIJE

Sam pojam bibliometrije najčešće nalazimo u literaturi koja se bavi primjenom matematičkih i statističkih metoda u kvantificiranju pisanih komunikacija, no dobar dio autora koristi i pojmove scientometrija, a u novije vrijeme informetrija. A kao posljedica pojave znanstvenih sadržaja u elektroničkom obliku, koji su dostupni isključivo preko interneta, od druge polovice 20. stoljeća javlja se i pojam



webometrije. Kao što ćemo u nastavku rada vidjeti, svi su ovi pojmovi u današnje vrijeme vrlo povezani i ne bi se trebala raditi razlika između njih, s obzirom da se ne razlikuju u istraživačkim problemima niti u metodama pristupa, već su razlike vidljive tek na terminološkoj razini (Olujčić-Vuković, 2007). Ono u čemu se svi citirani znanstvenici slažu jest da se pojava pojma bibliometrije (grč. *biblion* = knjiga, *metrein* = mjeriti; eng. *bibliometrics*, njem. *Bibliometrie*) pojavljuje sa člankom Alana Pritcharda iz 1969. godine (Jokić, 2005; Hood i Wilson, 2001; Pehar, 2011). Pritchard je uveo pojam bibliometrije nezadovoljan zastarjelim pojmom „statistička bibliografija“ kojega je do tada koristio E. W. Hulme. Maja Jokić u svojoj knjizi citira Pritcharda u njegovoj definiciji bibliometrije gdje on tvrdi da je bibliometrija „*primjena matematičkih i statističkih metoda na knjige i ostale medije u znanstvenom komuniciranju.*“ (Jokić, 2005). Franjo Pehar u svom znanstvenom radu daje nam jednostavniju definiciju bibliometrije citirajući Williama Graya Pottera gdje kaže „*Bibliometrija je, pojednostavljeno, istraživanje i mjerenje publikacijskih obrazaca svih oblika pisane komunikacije i njihovih autora.*“ (Pehar, 2011).

## 1.2. SCIENTOMETRIJA

Istodobno s pojavom bibliometrije u Istočnoj Europi pojavljuje se nova disciplina pod nazivom naukometrija, a koja je kasnije preimenovana u scientometriju. Pojam su prvi uveli ruski autori Dovrov i Korennoi (Jokić, 2005), premda Maja Jokić citira i Garfielda iz 1998. godine koji tvrdi da je Price utemeljitelj scientometrije, odnosno „*znanosti o znanosti*“. Scientometrija se bavi kvantitativnim metodama koje se odnose na analizu znanosti, a bibliometrija bi se po Wilsonu trebala baviti literaturom o znanosti i znanstvenicima. Premda se po nazivu čini da se ove dvije discipline, a mislimo na bibliometriju i scientometriju, jasno razlikuju, u literaturi se vrlo često koriste kao sinonimi. U prihvaćanju pojma scientometrije posebno je pridonio časopis „*Scientometrics*“ koji je počeo izlaziti 1978. godine. Doduše, tu bi trebalo obavezno napomenuti da upravo u ovom časopisu veći broj radova koristi pojam bibliometrija (Jokić, 2005). Franjo Pehar u svom znanstvenom radu napominje da je Eugene Garfield utvrdio da je po Pritchardovoj definiciji bibliometrije većina scientometrijskih istraživanja u biti bibliografske naravi (Pehar, 2011). Prema Garfieldu „*Ukoliko se istražuju znanstvene publikacije, tada je riječ o bibliometriji znanosti*“ (Hood i Wilson, 2001). No, bez obzira na prethodno opisane razlike u definicijama ovih dviju

disciplina, te razlike u zadnjih nekoliko desetljeća iščezavaju, a danas se navedena dva pojma najčešće koriste kao sinonimi.

### **1.3. INFORMetriJA**

Informetriju je prvi predložio Nacke 1979. godine definirajući je kao poddisciplinu unutar informacijskih znanosti koja se bavi mjerenjem fenomena informacija i primjenom matematičkih metoda na probleme informacijskih znanosti. (Pehar, 2011; Jokić, 2005). Informetrija se katkada koristi kao sinonim za bibliometriju, međutim većina autora se slaže da je informetrija disciplina koja pokriva znatno šire područje od same bibliometrije. Franjo Pehar ističe misao Nackea koji tvrdi da su informetrija i scientometrija dva sestrinska područja unutar informacijskih znanosti. William W. Hood i Concepción S. Wilson u svom članku ističu da je na prvoj međunarodnoj konferenciji o informetriji (*The first International Conference of bibliometrics and Theoretical Aspects of Information Retrieval*) Nacke predložio da se pojmovi bibliometrija i scientometrija zamijene pojmom informetrija, čiji bi predmet bili tiskani i elektronički dokumenti, no ta ideja ipak nije zaživjela. U prethodno iznesenim argumentima vidljivo je da su discipline bibliometrija, scientometrija i informetrija donekle srodne, ali ipak zasebne discipline. Razlike se u njima očituju većinom na terminološkoj razini.

### **1.4. WEBOMETRIJA**

Razvojem World Wide Weba u posljednjih nekoliko desetljeća interes znanstvenika u navedenim metrijskim disciplinama okreće se prema kvantitativnom mjerenju web stranica i na taj je način stvoren novi termin koji bi mogao opisati primjenu informetrijskih metoda na mrežne stranice, a taj termin je webometrija ili kako ga Wilson u svom članku još naziva „*cybermetrija*“ (Hood i Wilson, 2001). No, kao i do sada, možemo očekivati da će se oba termina neko vrijeme koristiti istovremeno i da će znanstvenici odabrati termin koji najtočnije opisuje njihovo istraživanje. Ono što i ovdje možemo primijetiti jest preklapanje pojmova i metoda istraživanja weba. Naime, u selekcioniranju kvalitetnih informacija sa interneta koriste se osnovne bibliometrijske metode, a to se osobito odnosi na citatne analize, koje su dostupne hipervezama (veza između izvornog dokumenta sa srodnim dokumentima i onima koji ga citiraju) kao novoj tehnološkoj mogućnosti (Jokić, 2005.).

## 1.5. KRITIČKI OSVRT NA POSTOJEĆE STANJE-INFORMETRIJA KAO ZAJEDNIČKA DISCIPLINA

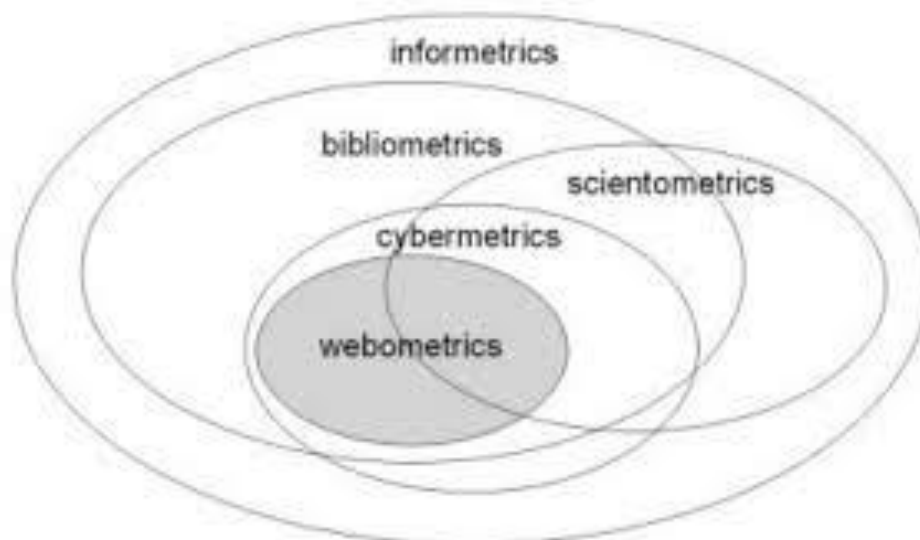
Mnogi se znanstvenici danas ne slažu s do sada navedenom podjelom disciplina i pokušavaju nam dokazati da se do sada navedene metrijske discipline ne bi trebale odvojeno promatrati, nego da bi trebale biti ujedinjene u jednu zajedničku disciplinu, a to bi trebala biti informetrija. Vesna Olujić-Vuković tvrdi da su razlike između bibliometrije, informetrije i scientometrije vidljive samo na terminološkoj razini i da je to vrlo lako provjerljivo već i površnim pregledom literature koja se odnosi na ovu problematiku. U brojnim udžbenicima, priručnicima i znanstvenim člancima, tvrdi Olujić Vuković, različiti nazivi za istu tematiku posljedica su u najvećoj mjeri stava samih autora radova (Olujić-Vuković, 2007). Ovu tvrdnju pokazuje i grafikon na slici 1 preuzet iz članka William W. Hood, Concepción S. Wilson: *“The literature of bibliometrics, scientometrics, and informetrics”* koji nam zorno pokazuje koliko se na terminološkoj razini razlikuju, odnosno miješaju različiti nazivi za istu tematiku.

Table 1  
Number of documents with each of the different terms related to the metric fields in Information Science

Term	Frequency
BIBLIOMETRICS	5097
BIBLIOMETRIC	2653
SCIENTOMETRICS	1326
SCIENTOMETRIC	552
INFORMETRICS	418
TECHNOMETRICS	274
INFORMETRIC	197
BIBLIOMETRY	73
BIBLIOMETRICALLY	40
STATISTICAL BIBLIOGRAPHY	38
BIBLIOMETRICAL	24
TECHNOMETRIC	20
BIBLIOMETRICIANS	17
SCIENTOMETRY	17
LIBRAMETRY	16
SCIENTOMETRICAL	11
SCIENTOMETRICALLY	10
SCIENTOMETRICIANS	9
BIBLIOMETRICIAN	7
LIBRAMETRICS	7
SCIENTOMETRICIAN	6
INFORMETRICIANS	5
INFORMETRY	5
LIBRAMETRIC	5
TECHNOMETRICALLY	1

Slika 1, Miješanje pojmova u bibliometriji, scientometriji i informetriji (prema Hood i Wilson, 2001)

Maja Jokić nam u svojoj knjizi daje nešto drugačije objašnjenje. Naime, ona smatra da se, iako se čini da su definicije metrijskih disciplina prilično jasne, i dalje u literaturi pojavljuju preklapanja i neujednačenosti u terminologiji i to zbog činjenice da se problemom publikacija u današnje vrijeme bave različiti stručnjaci, u prvom redu knjižničari i informacijski specijalisti, znanstvenici, financijeri raznih istraživanja te na kraju sami bibliometričari, scientometričari i informetričari kao zasebna zanimanja (Jokić, 2005). Stoga, obje autorice smatraju „*da je uspostava jedinstvene discipline koja bi se bavila ispitivanjem kvantitativnih aspekata procesa stvaranja, prijenosa i korištenja znanstvenih informacija nužna i da je u tom smislu informetrija najbolje rješenje*“ (Jokić, 2005). Da bismo si vizualno mogli predočiti isprepletenost ovih navedenih disciplina najbolje je pogledati sliku 2 u nastavku rada.



Slika 2, Preklapanje metrijskih disciplina (prema Bjerneborn i Ingwersen, 2004)

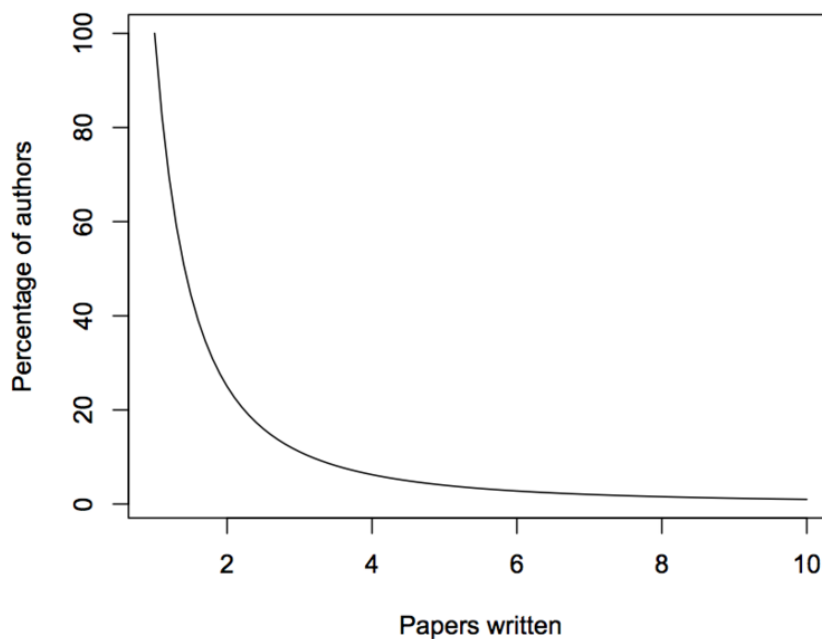
## 2. BIBLIOMETRIJSKI ZAKONI

Osnovni bibliometrijski zakoni mogli bi se podijeliti u tri cjeline, s obzirom da se svaki od njih odnosi na pojedino područje bibliometrijskih istraživanja. Prvi od njih, Lotkin zakon o produktivnosti autora, prvenstveno je vezan za autore i autorstvo radova. Drugi, Bradfordov zakon distribucije, vezan je za časopise kao osnovni medij u znanstvenom komuniciranju, a treći, Zipfov zakon učestalosti pojave riječi u tekstu, također se odnosi na radove u časopisima, no baziran je na kvantitativnim mjerjenjima pojava riječi u tekstovima. Stoga ćemo ovaj dio rada koncipirati tako da

pojedinačno dotaknemo svaki od zakona, a ujedno pokušamo odgovoriti i na problematiku koja je vezana uz pojedini zakon.

## 2.1. LOTKIN ZAKON O PRODUKTIVNOSTI AUTORA

Alfred Lotka, matematičar najviše poznat po svom radu u biologiji, razradio je zakon koji je dobio naziv po njegovom imenu 1929. godine. Po njemu, produktivnost nekog autora definira se kao broj radova objavljenih u određenom razdoblju i izračunava se prema formuli  $Y=C/X^n$  gdje Y označava postotak autora, X označava broj radova, a C i N su konstante. Lotka je tako izračunao da oko 60 % autora objavi po jedan rad unutar određenog područja dok dva rada objavi otprilike jedna četvrtina broja autora koji su objavili po jedan rad (Jokić, 2005). Slike 3 i 4 zorno pokazuju Lotkin zakon o produktivnosti autora



Slika 3, Lotkin zakon o produktivnosti autora (prema Tuđman, 2004)

LOTKIN ZAKON O PRODUKTIVNOSTI AUTORA	
Autori	Članci
100	1
25	2
11	3
6	4
4	5

Slika 4, Omjer autora i članaka prema Lotkinom zakonu o produktivnosti autora (prema Tuđman, 2004)

Kada govorimo o znanstvenoj produktivnosti autora moramo uzeti u obzir nekoliko faktora. Naime, Maja Jokić navodi da je mjerenje znanstvene produktivnosti, kada se radi o jednom ili o dva autora, prilično jednostavno s obzirom da se u tom slučaju koriste neki od uobičajenih pokazatelja, kao što je broj radova u časopisima zastupljenim u ISI-jevim bazama podataka (Institute for Scientific Information). Veći se problem javlja pri vrednovanju doprinosa autora u višeautorskim radovima. U ovim slučajevima mogu se koristiti dva principa.

**a)** Rad se vrednuje kao jedan rad za svakog autora.

Tu, međutim, dolazimo do problema sustava vrednovanja kod izbora znanstvenika u zvanje, odnosno napredovanja. Naravno da nije svejedno je li među autorima dva do tri autora ili se radi o grupi od stotinjak autora kao što je čest slučaj u radovima iz područja fizike.

**b)** Rad se temelji na mjerenju doprinosa pojedinog autora.

Po ovom principu prvi autor je onaj koji je proveo istraživanje, na posljednjem mjestu je supervizor, a između su autori koji su radili na pojedinim segmentima rada.

Wilson i Hood, a također i Maja Jokić navode Moeda kao jednog od autora članka u kojem je naveo jednu od sveobuhvatnijih metodologija za lakše razumijevanje znanstvene produkcije (Moed, 2005 prema Jokić, 2005). On pri vrednovanju znanstvenog rada razlikuje sljedeće pokazatelje:

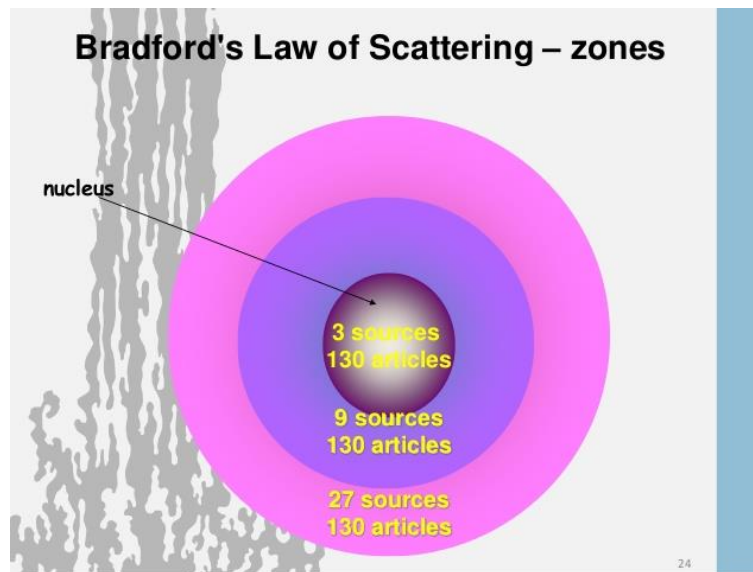
- Koliko je netko od autora djela sudjelovao u istraživanju i pisanju rada. On je suradnike razvrstao u četiri kategorije: voditelje, postdoktorande, mlade znanstvenike i tehničko osoblje. Najviše su kao autori rada vrednovani voditelji, a najmanje tehničko osoblje.
- Doprinos pojedinog autora na radu računao je kroz omjer 3:2:1 za prvog, odnosno drugog i trećeg autora.
- U nekim se radovima u fusnotama navodi doprinos svakog pojedinog autora. Neki znanstvenici smatraju da je autor koji je označen kao kontakt osoba najvažniji autor na radu.
- U biti postoje dvije strategije u publiciranju. Prva je da se objavljuje što veći broj radova, što prakticiraju znanstvenici koji se tek pojavljuju na znanstvenoj sceni, a drugi je da se objavljuje manji broj radova, ali da budu objavljeni u vodećim znanstvenim časopisima. Ovoj drugoj strategiji pribjegavaju već renomirani i priznati znanstvenici.
- Doktorandi trebaju imati objavljen određen broj radova, obično 5, kako bi postigli doktorski stupanj. Moed nije postavio uvjet trebaju li radovi biti objavljeni u časopisima sa visokim faktorom odjeka ili ne.
- Mjerenje udjela doprinosa pojedine grupe autora na radu mjeri se po određenoj shemi. Što je više grupa ili autora, to je manji broj bodova. Na taj način radovi s više autora bivaju kažnjeni za suradnju jer dobivaju manji broj bodova. Ovdje bi se također dalo raspravljati je li pravedno kažnjavati grupu autora radova samo zato što su podijeljeni u grupe, a nisu autorstvo ostvarili sami.

Naravno da uz Moeda postoje i drugi autori članaka koji daju svoje viđenje problema vezanih uz znanstvenu produkciju, no za ovu priliku odlučili smo se izdvojiti samo Moedovo mišljenje kao zanimljivo za ovaj rad.

## 2.2. BRADFORDOV ZAKON DISTRIBUCIJE

Bradfordov zakon distribucije direktno je vezan za časopise i broj članaka u njima. Samuel C. Bradford je 1934. godine došao do zaključka da se najveći broj važnih članaka o nekoj temi nalazi u malom broju časopisa koje on naziva jezgrom (Jokić, 2005). David Bawden i Lyn Robinson u svojoj knjizi „*Introduction to Information Science*“ navode da to znači da je trećina korisnih članaka raspršena u velikom broju izvora (Bawden i Robinson, 2012). Primjenom Bradfordova zakona distribucije možemo dobiti distribuciju časopisa po takozvanim zonama. Naime, prvu, A zonu činiti će najmanji broj časopisa sa velikim brojem relevantnih članaka. Zonu B činiti će veći broj časopisa sa manjim brojem relevantnih članaka, dok će C zonu činiti najveći broj časopisa, ali s vrlo malim brojem relevantnih članaka za temu koju istražujemo. Bradfordov zakon možemo definirati formulom  $1:n:n^2$  gdje broj 1 označava broj časopisa u jezgri,  $n$  označava broj časopisa u drugoj zoni, a  $n^2$  označava broj časopisa u trećoj zoni. To bi u primjeru na slici 5 značilo da u prvoj zoni (jezgri) imamo 3 časopisa u kojima postoji 130 relevantnih članaka, u drugoj zoni za isti broj relevantnih članaka je potreban  $n^2$  broj časopisa, odnosno 9, a u trećoj zoni je za isti broj relevantnih članaka potreban  $n^3$  broj časopisa, odnosno 27. Nakon takvog načina pretraživanja članaka u časopisima većina istraživača će brzo zaključiti da nema smisla tražiti dalje jer će svi relevantni članci već biti nađeni u samoj jezgri časopisa. Ilustracija Bradfordovog zakona vidljiva je na slici 5 u nastavku ovoga rada.





Slika 5, Bradfordov zakon distribucije

Preuzeto sa:

<https://www.slideshare.net/ranjsumit/bibliometrics-scitometrics>

Bawden i Robinson u svom radu ističu da Lotkin i Bradfordov zakon također imaju veliku sličnost sa Paretovim zakonom koji glasi da se 80 % postignutog rezultata postiže u 20 % od ukupnog vremena tijekom projekta (Bawden i Robinson, 2012). Za postizanje preostalih 20 % potrebno je najviše rada. Vilfredo Pareto je pri istraživanju o raspodjeli nacionalnog bogatstva otkrio da u Italiji oko 20% obitelji posjeduju oko 80 % kapitala i taj zakon je po njegovom imenu dobio ime Pareto zakon. Iz primjera Paretovog zakona potpuno je očita sličnost sa dva gore navedena bibliometrijska zakona. Na ovu temu nadovezuje se i Maja Jokić u svojoj knjizi povezujući sa Lotkinim i Bradfordovim zakonom takozvani „Matejev efekt“, koji je u sociologiju uveo Robert King Merton. Ovaj fenomen je prilično poznat u sustavu nagrađivanja znanstvenika i u sustavu citiranja i u načelu glasi da poznati znanstvenici, upravo zato što su poznati, dobivaju priznanja kroz nagrade i citiranje (Jokić, 2005). Sam Matejev efekt proizašao je iz Biblije, iz evanđelja po Mateju i glasi „(Mk 4,10-12; Lk 8,9-10) (11) On im odgovori: "Zato što je vama dano znati otajstva kraljevstva nebeskoga, a njima nije dano. (12) Doista, onomu tko ima dat će se i obilovat će, a onomu tko nema oduzet će se i ono što ima.“

Jokić nam u svojoj knjizi napominje i nekoliko važnih parametara koji su vrlo važni za svaki časopis ukoliko želi kao takav ostati na izdavačkoj sceni. Ovdje

bismo htjeli izdvojiti dva parametra za koja smatramo da su bitna za istaknuti u ovome radu.

**a) Naslov časopisa:**

Naslov časopisa vrlo je važan u pogledu prepoznatljivosti samoga časopisa za potencijalne čitatelje i autore. Neprijeporno je da promjena naziva časopisa može uzrokovati poteškoće u kategorizaciji i održavanju zbirki časopisa. Uredništvima časopisa u interesu je da njihovi časopisi budu u što većoj mjeri dostupni većem broju korisnika.

**b) Izdavači časopisa:**

Sami izdavači časopisa također su vrlo važni kada govorimo o citatnim analizama i bibliometrijskim metodama pretraživanja baza podataka. Naime, nikako nije svejedno jesu li izdavači časopisa komercijalni nakladnici, poznate akademske institucije ili vladine organizacije, ili pak neko strukovno udruženje odnosno manja znanstvena institucija. Ako se iza časopisa krije poznata institucija ili nacionalna akademija gotovo se automatski može reći da autoritet takvih institucija garantira kvalitetu časopisa. Ovdje bismo citirali Maju Jokić, koja u ovih nekoliko rečenica vrlo ilustrativno prikazuje način razmišljanja prilikom pretraživanja relevantnih informacija u časopisima iza kojih stoji neka institucija sa autoritetom. *„Ako se uzme u obzir činjenica da su u takvim „paketima“ časopisi s cjelovitim tekstovima dostupni i retrospektivno do najmanje 5 godina, da su u visokom postotku zastupljeni u ISI-jevima bazama podataka, znanstvenici zbog laganog izravnog pristupa na svom računaru i uštede vremena, uglavnom prate časopise koji su im na raspolaganju. Dakle, koriste ih i za objavljivanje i za citiranje, pri čemu drugi časopisi nemaju takvu šansu.“* (Jokić, 2005).

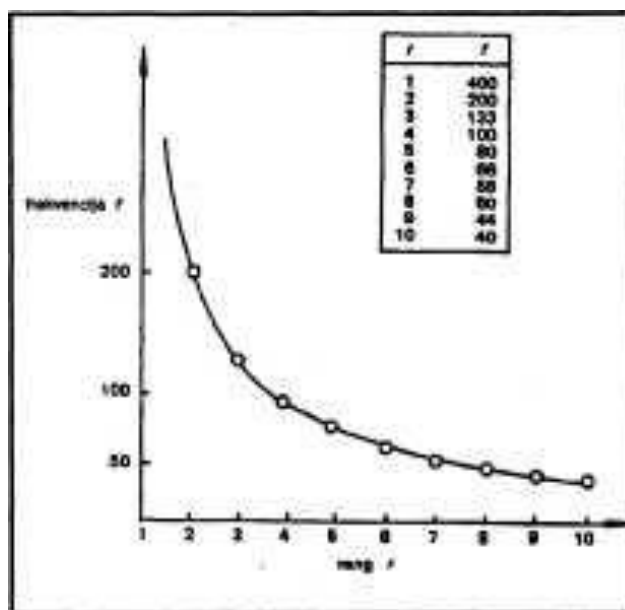
### **2.3. ZIPFOV ZAKON UČESTALOSTI POJAVE RIJEČI U TEKSTU**

Treći zakon koji čini trolist glavnih zakonitosti bibliometrije jest Zipfov zakon učestalosti pojave riječi u tekstu. Američki lingvist George Kingsley Zipf istraživao je distribuciju riječi u tekstovima i 1935. godine ustanovio sličnu pojavu kao u Bradfordovom i Lotkinom zakonu. Naime, prilikom istraživanja distribucije riječi u tekstu ustanovio je da se mali broj riječi u tekstovima pojavljuje vrlo često. Zanimljivo je da je Zipf priznao da mu je tu ideju zapravo istaknuo njegov neimenovani prijatelj (Bawden i Robinson, 2012). On je ustanovio da će, ako se riječi u određenom tekstu

poredaju u padajućem slijedu po frekvenciji pojavljivanja, položaj riječi odnosno njeno mjesto na listi biti određeno njenom frekvencijom pojavljivanja. „Zipfov zakon tvrdi da, ako su riječi poredane po padajućem poretku, da će tada redni broj riječi pomnožen sa svojom frekvencijom pojavljivanja biti konstanta. Matematički, za riječ sa frekvencijom pojavljivanja  $Y$  i rednim brojem  $R$  vrijedi  $Y=K/R$ , gdje  $K$  predstavlja konstantu, ovisno o kontekstu.“ (Bawden i Robinson, 2012). Zipfov zakon možemo zorno pokazati tablicom i parabolom prikazanom na slikama 6 i 7. Zipfov se zakon najčešće primjenjuje u knjižnicama u analizama i indeksiranju riječi.

ZIPFOVA DISTRIBUCIJA			
riječ	redoslijed	frekvencija	red. × frekv.
the	1	245	245
of	2	136	274
terms	3	98	294
to	4	81	324
a	5	65	325
and	6	61	366
in	7	55	385
were	8	52	416

Slika 6, Distribucija riječi po Zipfovom zakonu (prema Tuđman, 2004)



Slika 7, Parabola frekvencije pojave riječi u tekstu po Zipfovom zakonu (prema Tuđman, 2004)

### 3. FAKTOR ODJEKA ČASOPISA – PRIMARNO SREDSTVO VREDNOVANJA ČASOPISA

U nastavku teorijskog dijela rada i za njegov kraj, a zbog vrlo velike opsežnosti materijala o zadanoj temi, željeli bismo se fokusirati na nezaobilazno gradivo kada se govori o bibliometriji, a to je faktor odjeka i indeks brzine citiranja. No prije toga, kao uvod u ovu temu, smatramo da bi trebalo reći par riječi o Journal Citation reportu<sup>1</sup> s obzirom da je neizbježno povezan sa ovom temom.

#### 3.1. JOURNAL CITATION REPORT

Poznata je činjenica da članci u časopisima donose popis literature na koji se njegovi autori pozivaju. Koji će izvori biti citirani ovisi o autorima članaka i o temi samih članaka. Na ovoj se tvrdnji temelji cijela teorija o faktoru odjeka i koncept citatnih baza podataka. JCR je kao posebnu statistički bazu podataka kreirao Eugene Garfield 1975. godine, a nakon što su se razvile računalne tehnike i citatne baze, gdje u prvom redu mislimo na SCI (Science Citation Index<sup>2</sup>) i SSCI (Social Science

<sup>1</sup> U nastavku rada koristit ćemo kraticu JCR

<sup>2</sup> U nastavku rada koristit ćemo kraticu SCI

Citation Index<sup>3</sup>). „JCR je kvantitativno pomagalo za rangiranje, vrednovanje, kategoriziranje, i komparaciju časopisa. Sadrži statističke podatke o ukupnom broju citata koje je dobio pojedini časopis, broj članaka koji su objavljeni u tom časopisu i prosječan broj citata po članku. Za svaki uključeni časopis donosi popis časopisa koji ga citiraju kao i popis i učestalost časopisa koje on citira. Na temelju navedenih podataka ISI (Institute for Scientific Information<sup>4</sup>, na hrvatskom jeziku zvan Institut za znanstvene informacije) izračunava faktor odjeka (impact factor<sup>5</sup> – IF) časopisa, indeks brzine citiranja (immediacy index) i poluvrijeme citiranja časopisa odnosno poluvrijeme citata u časopisu (cited and citing half time). Baza podataka se sastoji od dva dijela: JCR Science edition, koja obrađuje oko 5700 časopisa i JCR Social Science edition, koja obrađuje oko 1700 časopisa, što ukupno iznosi oko 7000 časopisa razvrstanih u dvjestotinjak predmetnih područja.“ (Jokić, 2005). Danas JCR izlazi jednom godišnje i najčešće se koristi kao pomagalo za vrednovanje radova pojedinih znanstvenika. Iako JCR nudi pregršt statističkih podataka o časopisu, najveći broj korisnika ove baze podataka koristi se isključivo faktorom odjeka časopisa.

### **3.2. FAKTOR ODJEKA – IMPACT FACTOR**

Đilda Pečarić u svojoj doktorskoj disertaciji ističe da, prema C. Borgmanu i C. Furneru, bibliometrija nudi moćan skup metoda i mjera za istraživanje struktura znanstvene komunikacije. Nadalje navodi da su osnovne jedinice na kojima se provode bibliometrijske analize zbirke publikacija kao što smo već i prije napomenuli (Pečarić, 2010). S pojavom ISI-ja koncem 1960-ih godina počinje se objavljivati SCI koji daje pregled objavljenih znanstvenih radova i citiranih bibliografskih jedinica u tim radovima. U osnovi IF označava omjer broja citata dobiven na objavljene članke i broja objavljenih članaka u određenom razdoblju. IF časopisa nalazimo u godišnjem izdanju JCR-a i računa se tako da se broj citata dobiven u tekućoj godini na radovima objavljenim u protekle dvije godine podijeli s brojem radova objavljenim u tom istom razdoblju.

---

<sup>3</sup> U nastavku rada koristit ćemo kraticu SSCI

<sup>4</sup> U nastavku rada koristit ćemo kraticu ISI

<sup>5</sup> U nastavku rada koristit ćemo kraticu IF

Zorno prikazano to izgleda ovako:

*„ $C = A/B =$  faktor odjeka (IF) časopisa u 2003. godini*

*A = broj citata dobiven u 2003. godini na članke objavljene u 2001. i 2002. godini*

*B = broj radova objavljen u 2001. i 2002. godini“ (Jokić, 2005)*

Ovdje treba napomenuti da je u posljednjim desetljećima JCR postao idealan alat za analizu i usporedbu znanstvenih časopisa i da, štoviše, u mnogim zemljama kao dokaz uspješnosti vlastite znanstvene produktivnosti znanstvenici moraju navoditi citiranost, odnosno IF vlastitih radova, da bi mogli biti promaknuti u više znanstveno zvanje (Pečarić, 2010). Uz to, profesorica Pečarić navodi da je termin „faktor odjeka“ kao prijevod sa engleskog jezika skovala Maja Jokić u svom priručniku pa stoga i koristimo ovaj prevedeni termin kroz cijeli rad.

Govoreći o IF-u ne možemo se ne dotaknuti problematike koju on nosi sa sobom. Naime, ne slažu se svi znanstvenici da je formula za izračun samoga IF-a idealna ovakva kakva postoji na papiru. Neke od problema vezane za izračun IF-a izložiti ćemo u nastavku ovoga rada:

### **3.2.1. FAKTOR ODJEKA I CITATI**

Nesporazumi koji se često javljaju kod interpretiranja IF-a odnose se na ulogu ukupnog broja citata u godini. Naime, kod izračuna IF-a za neki časopis, po ukupnom broju citata u godini taj isti časopis može biti visoko rangiran na ljestvici dok njegov IF ne mora biti visok. Ta se diskrepancija javlja zbog toga što ukupan broj citata u godini u kojoj se IF izračunava može biti visok, ali broj citata na radove objavljene u posljednje dvije godine ne mora nužno biti visok. Primjerice, neki časopis može u 2003. godini imati velik broj citiranih članaka, što bi po očekivanju značilo i visok IF faktor. No, zbog toga što se IF izračunava tako da se broj citata ne uzima iz aktualne godine, a to je u ovom slučaju 2003., već iz prethodne dvije godine, sam IF časopisa nije u skladu sa očekivanjima visok, nego je znatno niži. Ovaj problem apostrofiraju mnogi znanstvenici koji zahtijevaju da se za određene discipline i područja razdoblje citiranja članaka u časopisima produži na tri, četiri ili čak na pet godina umjesto dosadašnje dvije godine i to zbog toga što u različitim disciplinama i područjima postoje razlike u učestalosti i navikama u citiranju (Jokić, 2005). Naime, neka područja i discipline su takvog karaktera da je potrebno preko dvije godina da bi se moglo procijeniti koliko je relevantan broj citiranih članaka u prethodnim godinama.

Za neke discipline potrebno je uzeti i do pet prethodnih godina da bi se mogao kvalitetno izračunati IF, dok je za neke sasvim dostatan takozvani standardni IF koji kod izračuna koristi prethodne dvije godine.

### **3.2.2. FAKTOR ODJEKA I SAMOCITATI**

Drugi vid problema, također vezan uz IF, javlja se u pogledu samocitiranja članaka. „O samo citiranju časopisa govorimo kada članak u časopisu A citira prethodne radove iz istog tog časopisa. Samocitati časopisa prosječno iznose oko 13% u odnosu na ukupan broj citata koje časopis primi... Garfield (1998) navodi da mali časopisi, a naročito kada su novi, kao glavni izvor citata imaju samocitate.“ (Jokić, 2005). Ovi podatci su sasvim razumljivi s obzirom da novi časopisi streme k tomu da imaju što bolji IF, a najlakši način da se IF učini velikim jest samocitiranje. No tu Maja Jokić dotiče još jedan veliki problem, a to je, mogli bismo reći lažiranje faktora odjeka i to objašnjava na ovaj način: „Osim što je samocitiranost časopisa dijelom prirodan proces, u nekim slučajevima postoji i negativna strana. Poznato je da uredništva nekih časopisa, u težnji da njihov časopis bude što bolje pozicioniran na JCR-ovoj ranglisti nastoje povećati faktor odjeka sugeriranjem autorima rukopisa da citiraju radove iz tog časopisa.“ Dakako da i za to postoji način izračuna IF-a, a to je izračun revidiranog IF-a koji ne uzima u obzir samocitate, odnosno u izračunu IF-a uzima se u obzir ukupan broj citata minus broj samocitata.

### **3.2.3. INDEKS BRZINE CITIRANJA I POLUVRIJEME CITIRANJA**

Kao što smo napomenuli na početku trećeg poglavlja ovoga rada, preostalo nam je samo ukratko objasniti pojmove brzine citiranja i poluvremena citiranja da bi ovaj teorijski dio rada mogli zaokružiti u jednu smislenu cjelinu. Stoga ćemo krenuti sa indeksom brzine citiranja.

- JCR nam za svaki časopis nudi i mogućnost mjerenja indeksa brzine citiranja kojom je prosječni članak citiran. Indeks brzine citirana ili immediacy index definira se kao omjer broja citata koje je časopis dobio u istoj godini u kojoj su članci objavljeni i broja članaka objavljenih u istoj toj godini. Formulom se to može iskazati ovako:

IB = indeks brzine citiranja u kalendarskoj godini

$IB = C/S$

C = broj dobivenih citata u istoj godini

S = broj članaka u istoj godini

Dakako da periodičnost izlaženja časopisa u ovom slučaju igra vrlo važnu ulogu. Časopisi koji izlaze na tjednoj bazi imaju indeks brzine citiranja veći nego oni časopisi koji izlaze na kvartalnoj ili polugodišnjoj razini. Ako časopis izlazi dva puta godišnje, njegov drugi broj će izaći pred kraj godine, čime radovi u tom časopisu imaju manju šansu za citiranje u godini objavljivanja nego članci iz časopisa koji izlaze na primjer na mjesečnoj bazi ili čak češće (Jokić, 2005). Na temelju ovih podataka o brzini citiranja JCR rangira časopise i na časopise čiji su članci u prosjeku najcitiraniji u tekućoj godini. Upravo ovi razlozi, a ovdje mislimo na frekvenciju izlaženja časopisa, vjerojatno su jedan od razloga manje korištenosti ovog pokazatelja u odnosu na IF.

- Još jedan pokazatelj koji nam nudi JCR za koji smatramo da je relevantan da bi se prikazao u ovome radu jest poluvrijeme citiranja (cited half-life). Ovaj pokazatelj jest zapravo pokazatelj zastarijevanja literature. On predstavlja broj godina računatih unatrag u odnosu na tekuću godinu, koje čine 50 % od ukupnog broja citata što ih je taj časopis dobio u tekućoj godini. Na primjeru to izgleda ovako. Ako je poluvrijeme citiranja nekog časopisa 5 godina za godinu 2010., to bi značilo da je polovina citata koje je časopis dobio u 2010. godini bila publicirana u zadnjih 5 godina. Preostala polovica citata tog časopisa u 2010. godini bila bi starija od navedenih 5 godina. Vrijeme zastarijevanja je literature je naravno brže u području tehnologije i bioloških znanosti, dok je znatno sporije kod teorijskih znanosti, matematike i fizike.

#### **3.2.4. BIBLIOGRAFSKE BAZE PODATAKA KAO SEKUNDARNI IZVORI INFORMACIJA**

Nakon što smo prilično detaljno objasnili teorijske značajke bibliografije, za sam se kraj teorijskog dijela rada moramo dotaknuti i baza podataka u kojima na kraju završe svi ovi članci i časopisi koji su prošli navedena bibliografska mjerenja. To je za krajnje korisnike, odnosno znanstvenike i sve zainteresirane za tematiku, zapravo najvažnije područje interesa kod pretraživanja materijala na zadanu temu.



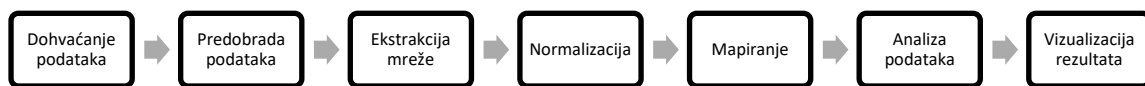
„Baze podataka su jedan od osnovnih pomagala kojima je svrha da olakšavaju dostupnost i pristup velikom broju primarnih izvora informacija, u što ubrajamo časopise odnosno članke iz časopisa, zbornike radove sa konferencija, knjige, patente, tehničke izvještaje i sl.“ (Jokić, 2005). Baze podataka su nastale nakon što se kroz vrijeme povećao broj znanstvenih časopisa i ostalih publikacija, a i zbog potrebe da se korisnicima omogući lakši način praćenja i pristupa navedenim informacijama, s obzirom da su se prije pojava baza podataka sve ove informacije čuvale u tiskanom obliku, što je znatno otežavalo pretragu literature. Prava snaga i funkcija baza podataka došla je do izražaja tek pojavom računalnih tehnologija koje omogućavaju relativno jednostavno i brzo pretraživanje informacija.

Jedne od najpoznatijih baza podataka su ISI-jeve bibliografske baze podataka (The Thomson Corporation). Eugene Garfield osnovao je 1960. godine Institut za znanstvene informacije (ISI), no od 2004. godine ISI više nije samostalna kompanija, već je prodan u vlasništvo privatne tvrtke iz Kanade pod nazivom The Thomson Corporation. Glavna djelatnost ISI-ja je proizvodnja bibliografskih odnosno citatnih baza podataka koje obrađuju najutjecajnije svjetske izvore literature (Jokić, 2005). U samoj bazi obrađeno je oko 8700 časopisa iz cijeloga svijeta. Svaki članak obrađen u citatnim bazama podataka u svom bibliografskom zapisu ima autora, naslov članka, podatak o časopisu ili nekom drugom izvoru, ključne riječi, sažetak na engleskom jeziku, adresu autora i nakladnika i ono što ih doista razlikuje od ostalih baza podataka, popis bibliografskih referenci. No, istraživanja su pokazala da je jedna od čestih primjedbi prevelika zastupljenost časopisa s američkog govornog područja. Stoga se sve češće javlja potreba da se uspostavi europski centar za vrednovanje znanosti koji će moći izrađivati vlastita vrednovanja literature u znanosti. Dodatni motiv koji potkrepljuje ovu ideju jest i prelazak ISI-ja u vlasništvo privatne tvrtke koja prilično snažno može utjecati na rezultate pretraživanja njihovih baza podataka. Kao što smo već jednom spomenuli, ali nije na odmet da se naglasi ponovno, najpoznatije ISI-jeve baze podataka su Science Citation Index (SCI), Social Science Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI). Također već navedeno, ISI-jev poseban proizvod jest i statistička baza podataka Journal Citation Report (JCR) koji smo detaljnije dotaknuli u odjeljku 3.1 ovoga rada.

## 4. BIBLIOMETRIJSKI ALATI:

Za sam početak praktičnog dijela ovoga rada moramo napomenuti da je znanstveno ili bibliometrijsko mapiranje važna istraživačka tema iz područja bibliometrije koja pokušava pronaći predodžbe intelektualnih veza unutar dinamički promjenjivog sustava znanstvenog znanja. Drugim riječima, znanstvenom mapiranju cilj je prikazivanje strukturalnih i dinamičkih gledišta znanstvenog istraživanja. Općeniti tijek rada u analizi znanstvenog mapiranja sadrži nekoliko koraka koji se moraju poštovati koristimo li bilo koji od bibliometrijskih alata pa tako i BibExcel ili SciMAT koji ćemo u nastavku praktičnog dijela rada pobliže objasniti. Ti koraci su redom:

- a. Dohvaćanje podataka
- b. Predobrada podataka
- c. Ekstrakcija mreže
- d. Normalizacija
- e. Mapiranje
- f. Analiza podataka
- g. Vizualizacija rezultata



Slika 8, Tijek rada u analizi znanstvenog mapiranja (prema Cobo, M.J. i dr.,2012)

- a. Za dohvaćanje podataka u današnje vrijeme postoji nekoliko mrežnih bibliografskih pa tako i bibliometrijskih baza podataka, a najvažnije su ISI Web of Science (ISIWoS), Scopus, Google Scholar i MEDLINE američke Nacionalne knjižnice za medicinu.
- b. Podatci dohvaćeni iz navedenih baza podataka uobičajeno sadrže pogreške, stoga se mora primijeniti postupak predobrade podataka, a to je jedan od najvažnijih procesa za dobivanje dobrih rezultata u analizi znanstvenog mapiranja. U predobradi podataka mogu se primijeniti različiti procesi kao što su otkrivanje dvostrukih i pogrešno napisanih stavki te redukcija podataka kako bi se mogli odabrati najreprezentativniji podatci za analizu. (Cobo, López-Herrera, Herrera-Viedma i Herrera, 2012)

- c.** Nakon predobrade podataka slijedi proces ekstrakcije podataka. Gledajući dohvaćene i predobrađene podatke može se primijetiti nekoliko vrsta odnosa među podacima, a to su supojavnost (eng. co-occurrence), uparenost (eng. coupling) ili direktna veza (eng. direct linkage) između dokumenata. Odnos supojavnosti je uspostavljen između dvije jedinice (autori, pojmovi, reference) kada se jedinice pojavljuju zajedno u nizu dokumenata tj. kada se supojavljaju u korpusu. Odnos uparenosti uspostavljen je između dva dokumenta kada imaju zajednički skup jedinica (autora, pojma ili referenci). Konačno, izravna veza između dokumenata uspostavlja vezu između dokumenata i referenci, osobito citatnog odnosa. (Cobo, López-Herrera, Herrera-Viedma i Herrera, 2011)
- d.** Jednom kada je proces ekstrakcije završen i kada je stvorena mreža odnosa između podataka potrebno je normalizirati podatke. Za taj proces koriste se različite mjere kao što su Jaccardov Indeks, Equivalence Indeks, Saltonov kosinus i dr.
- e.** Nakon procesa ekstrakcije podataka mogu se koristiti različite tehnike izgradnje znanstvenih mapa, poput multimedijskog skaliranja, algoritama klastera i dr.
- f.** Jednom kada je normalizacija podataka završena možemo započeti s analizom podataka koja nam omogućuje otkrivanje korisnog znanja iz sirovih podataka, mreža i mapa podataka. Postoje, naravno, različite metode analiza kao što su mrežna analiza, vremenska ili longitudinalna analiza, analiza učinka itd. Svaka od tih vrsta analiza nam omogućava otkrivanje drugačijih pogleda i znanja.
- g.** Na kraju analize znanstvenog mapiranja proces je vizualizacije podataka. Vizualizacija se koristi da bi se dobiveni rezultati analize mogli prezentirati kroz znanstvenu mapu. Vizualizacija podataka je vrlo važan proces jer nam omogućuje dobro razumijevanje izlaznih podataka analize, a ona se može postići korištenjem raznih vizualizacijskih alata kao što su heliocentrične mape, geocentrične mape, geometrijski modeli, tematske vizualizacije, vizualizacije gdje blizina stavki predstavlja njihovu sličnost itd. Naposljetku, kada je analiza znanstvenog mapiranja završena, analitičari trebaju interpretirati rezultate analize koristeći svoje znanje i iskustvo, a na taj način

mogu otkriti novo korisno znanje koje kasnije može biti iskorišteno za donošenje novih odluka.

Sada kada smo ukratko objasnili tijek jedne znanstvene analize korištenjem bibliometrijskih alata za znanstveno mapiranje pokušat ćemo ukratko opisati i usporediti dva poznata bibliometrijska alata, a to su BibExcel i SciMAT.

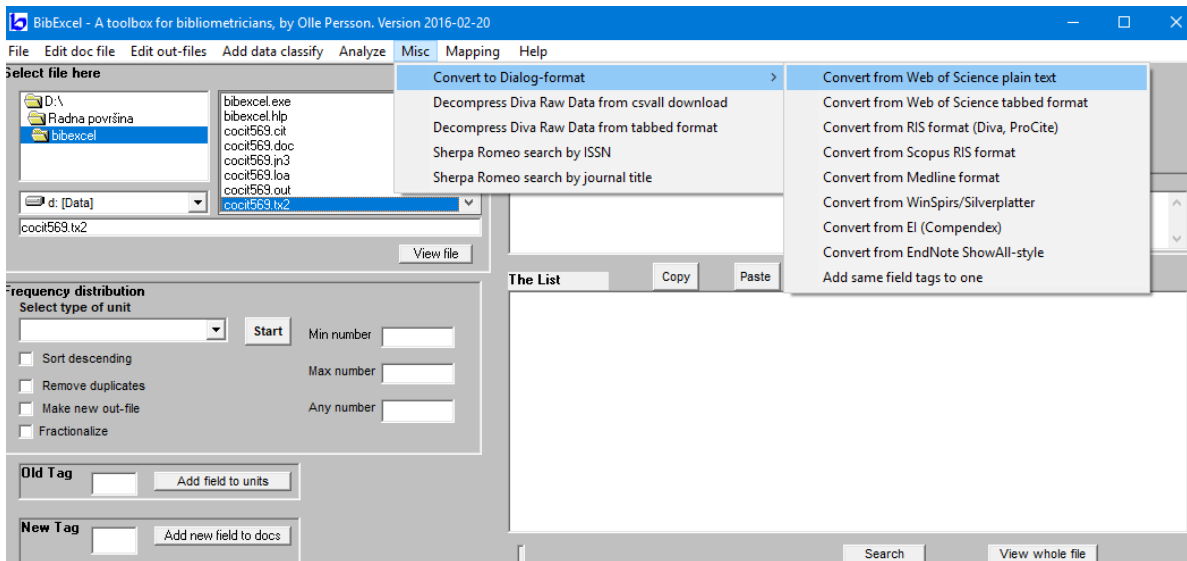
#### **4.1. BIBEXCEL**

BibExcel je svestrani bibliometrijski alat razvijen od strane Olle Perssona na Sveučilištu Umeå u Švedskoj u kojem je moguće napraviti većinu vrsta bibliometrijskih analiza. Bibexcel omogućuje jednostavnu interakciju s drugim programima kao što je program za vizualizaciju Pajek, nadalje Excel, IBM-ov program za napredne statističke analize SPSS itd. BibExcel nudi korisniku visok stupanj fleksibilnosti što je jedna od prednosti ovoga alata. BibExcel može, na primjer, koristiti podatke iz drugih izvora, a ne samo sa Web of Science pa tako može koristiti i podatke koji nisu bibliografski zapisi. No, ta navedena fleksibilnost ima i svoju cijenu jer može uzrokovati da korisnici dožive BibExcel kao alat koji se teško koristi. U daljnjem tijeku ovoga rada opisati ćemo osnovne funkcije BibExcel-a kroz analizu članaka preuzetih s Web of Science, a koji citiraju znanstvene publikacije Ollea Perssona. Opis navedenog alata pokušati ćemo strukturirati u četiri cjeline. U prvoj cjelini opisati ćemo kako se podatci za analizu preuzimaju sa Web of Science i na koji način moraju biti strukturirani. U drugoj cjelini ćemo pobliže objasniti OUT datoteku jer svaki postupak u Bibexcel-u kreira novu vrstu datoteke, a navedena se OUT datoteka uvijek kreira prva. U trećoj cjelini ćemo ukratko opisati osnovne analitičke funkcije navedenog alata, a u zadnjoj, četvrtoj cjelini ćemo objasniti na koji se način podatci mogu izvesti u program za vizualizaciju Pajek. Naravno, BibExcel sadrži mnogo više funkcija, no za potrebe ovoga rada fokusirati ćemo se na ove četiri navedene cjeline.

#### 4.1.1. PRIPREMA I UVOZ PODATAKA

Pilkington (2018) u svom tekstu navodi, a tome smo se i sami uvjerali, da je instalacija BibExcel-a prilično jednostavna. Potrebno je samo sve pomoćne datoteke, uz izvršnu datoteku, smjestiti u istu mapu na tvrdom disku računala.

Podatci koji su preuzeti sa Web of Science potrebno je spremirati u obliku \*.txt datoteke. Sljedeći korak je pretvoriti \*.txt datoteku u DIALOG format datoteku, a to ćemo učiniti opcijom „Misc → Convert to dialog format → Convert from Web of Science“. Na taj način stvoriti ćemo datoteku sa ekstenzijom \*.doc. U \*.doc datoteci bibliografski zapis započinje i završava sa dvije okomite crte „||“, a svaki je bibliografski zapis sastavljen od nekoliko bibliografskih polja kao na primjer polje autora „AU“, polje citata koje je obilježeno ili sa „CR“ ili sa „CD“, polje naslova „TI“, polje godine izdavanja „PY“ itd. Unutar bibliografskog zapisa bibliografska su polja odijeljena jednom okomitom crtom „|“, a ukoliko u jednom bibliografskom polju ima nekoliko jedinica, tada su one odijeljene znakom točka zarez „.“.



Slika 9, Pretvaranje \*.txt datoteke u datoteku DIALOG formata  
Preuzeto sa: BibExcel – A toolbox for bibliometricians, by Olle Persson. Version  
2016-02-20

```

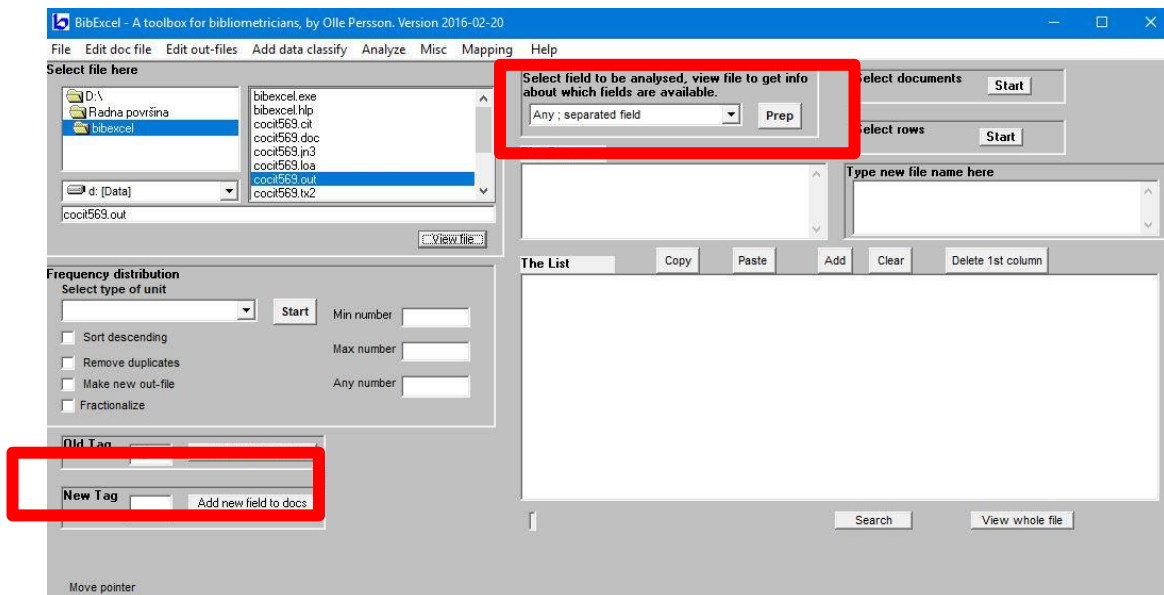
PT- 3]
AU- Su YM; Hsu PY; Pai NY|
AF- Su Yu-Min; Hsu Ping-Yu; Pai, Ning-Yao|
TI- An approach to discover and recommend cross-domain bridge-keywords in document banks|
SO- ELECTRONIC LIBRARY|
LA- English|
DT- Article|
DE- Databases; Data handling; Information retrieval; Cluster analysis|
ID- AUTHOR COCITATION ANALYSIS; INTELLECTUAL STRUCTURE; CLASSIFICATION; NETWORKS; SCIENCE; SEARCH; SPACE|
AB- Purpose - The co-word analysis method is commonly used to cluster-related keywords into the same keyword domain. In other words, traditional co-word analysis cannot cluster the same |
In more than one complete keyword set, the same keywords may end up appearing in different clusters Findings - Results of this study show that the CKP method can discover bridge-keywords |
CI- [Hsu, Ping-Yu] Natl Cent Univ, Dept Business Adm, Tao Yuan, Taiwan; [Su, Yu-Min] Natl Chengchi Univ, Dept Comp Sci, Taipei 11623, Taiwan; [Pai, Ning-Yao] Natl Chiao Tung Univ, Inst I |
RP- Hsu, PY, Natl Cent Univ, Dept Business Adm, Tao Yuan, Taiwan|
EM- pyhsu@ngt.ncu.edu.tw|
CD- *ASS COMP MACH, 2009, ACM COMP CLASS SYST; *ASS COMP MACH, 2009, ACM PORT; AHLGREN P, 2003, V54, P550, J AM SOC INF SCI TEC; AVANCINI H, 2004, P67, P IADIS C APPL COMP; CHANG CC, 2008 |
P 1 ACM INT C DIG LI; SHERE AA, 2002, V29, P1, KNOWL ORGAN; TANAKA H, 2000, V11, P73, GENOME INFORMATICS; VILLARROEL M, 2002, V20, P386, ELECTRON LIBR; WHITE HD, 1981, V32, P163, J AM SO |
NR- 32|
TC- 0|
PU- EMERALD GROUP PUBLISHING LIMITED|
PI- BINGLEY|
PA- HOWARD HOUSE, WAGON LAWE, BINGLEY BD16 1WA, W YORKSHIRE, ENGLAND|
SN- 0264-0473|
J9- ELECTRON LIBR|
J1- Electron. Libr|
PV- 2010|
VL- 28|
IS- 5|
BP- 669|
EP- 687|
DI- 10.1108/02640471011081951|
PG- 19|
SC- Information Science & Library Science|
GA- 6979C|
JN- ELECTRONIC LIBRARY, 2010, V28, N5, P669-687|
UT- ISI:000285526700004 ER ||

```

Slika 10, Prikaz bibliografskog zapisa u \*.doc datoteci u Bibexcel-u  
Preuzeto sa: cocit596.doc

#### 4.1.2. KREIRANJE OUT DATOTEKE

Prethodno opisanim postupkom stvorili smo datoteku koja sadrži bibliografske zapise čitljive u Bibexcel-u i sada smo spremni stvoriti i OUT datoteku. Stvaranje OUT datoteke uvijek je prvi korak kod bibliografske analize pomoću alata BibExcel. Persson, Danell i Schneider (2009) u svom članku napominju da je ovo trenutak kada moramo odlučiti od kojih će bibliografskih polja OUT datoteka biti sačinjena. U našem slučaju to će biti datoteka temeljena na polju autora. To ćemo učiniti tako da ćemo u kućicu „Old tag“ upisati „AU“, a u kućicu „Select field to be analyzed“ izabrati ćemo „Any; separated field“.



Slika 11, OUT file baziran na polju autora  
Preuzeto sa: BibExcel – A toolbox for bibliometricians, by Olle Persson. Version 2016-02-20

Nakon što smo kliknuli gumb „*Prep*“ BibExcel će stvoriti OUT datoteku temeljenu na polju autora, a svaki redak u OUT datoteci sadržavati će broj dokumenata u kojima je neki autor citirao Ollea Perssona. Iz slike 12 vidljivo je da je u analizi obuhvaćeno sveukupno 390 dokumenata u kojima je citiran Olle Persson, a da su prvi dokument u OUT datoteci napisali Levitt, JM i Thelwall, M.

Document identification number	Authors
1	Levitt JM
1	Thelwall M
2	Hsu PY
2	Shiau WL
2	Su YM
2	Yang SC
3	Anegon FD
3	Guerrero-Bote VP
3	Olmeda-Gomez C
3	Ovalle-Perandones MA
3	Perianes-Rodriguez A
....	....
....	....
388	Stefaniak B
389	Vlachy J
390	Ellis D

Slika 12, Struktura OUT datoteke temeljene na polju autora (prema Persson, 2009)

Ukoliko u OUT datoteci želimo napraviti izračun frekvencijske distribucije i to je u BibExcel-u moguće. Na primjer, ukoliko se naša OUT datoteka sastoji od citiranih dokumenata, u BibExcel-u možemo izabrati između metode brojanja i to „cjelobrojnog brojanja“ i „frakcijskog brojanja“. Metodu frakcijskog brojanja prilično je lako razumjeti. Naime, ukoliko nam se OUT datoteka sastoji od članaka koji su napisani u koautorstvu dvaju autora, a citirali su Ollea Perssona, tada će svaki od njih dobiti po pola boda, a ukoliko je članak napisan u koautorstvu tri autora, tada će svaki od autora članka koji je citirao Perssona dobiti po trećinu boda. Persson, Danell i Schneider (2009) u svom članku to zorno prikazuju u tablici koju ćemo priložiti kao sliku 13 u ovome radu.

Author	Whole counts	Fractional counts
Persson O	15	7,999
Leydesdorff L	14	9
Glanzel W	13	7,416
Meyer M	13	9,833
Melin G	12	7,916
Zitt M	8	3,332
White HD	7	5,5
Rousseau R	6	2,366
Gomez I	6	2,199
Zuccala A	6	4,5
Cronin B	6	4,333
Moya-Anegon F	5	1,999
Morris SA	5	2,083
Bassecoulard E	5	2,166
Herrero-Solana V	5	1,749

Slika 13, Cjelobrojno i frakcijsko brojanje u frekvencijskoj distribuciji BibExcel-a (prema Persson, 2009)

#### 4.1.3. OSNOVNE ANALITIČKE FUNKCIJE BIBEXCEL-A

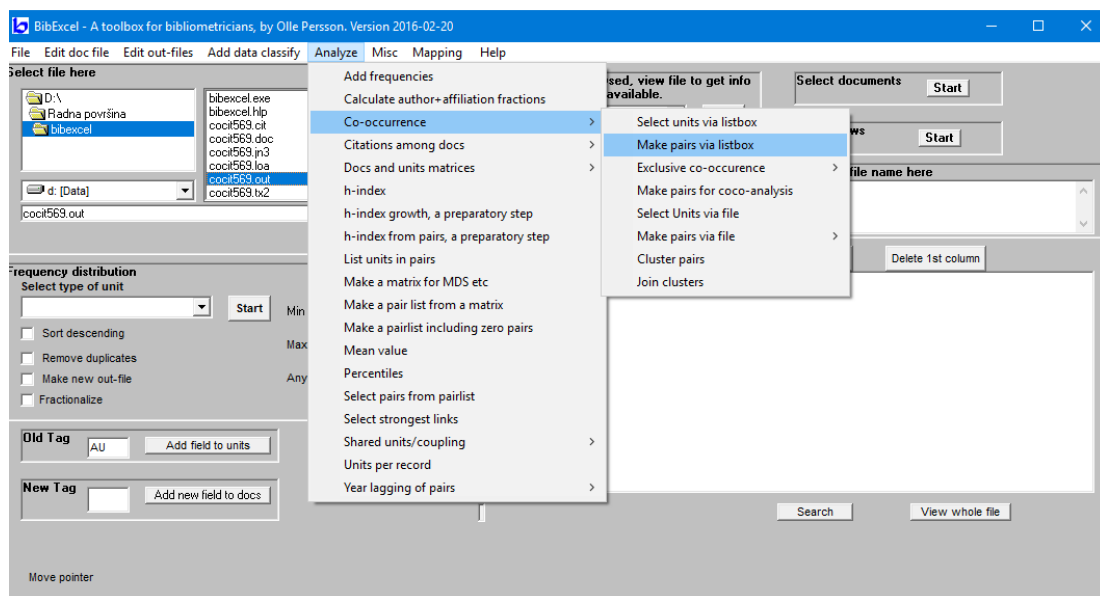
Do sada smo se fokusirali na pripremu podataka za analizu u BibExcel-u, a sada ćemo pobliže pojasniti nekoliko specijaliziranih funkcija koje nam omogućavaju analize citatnih mreža, a možda od toga i najvažnije, nekoliko analiza supojavnosti.

Pojam supojavnosti u bibliografskom zapisu obično označava zajedničku pojavu dviju jedinica u istom polju metapodataka, dakle kada se riječi X i Y pojavljuju zajedno u polju naslova ili kada se autori Z i W pojavljuju zajedno u polju autora. Očito je da nam supojavnost dvaju zapisa nije zanimljiva, ali kada se ta pojava javlja na više zapisa govorimo o analizi supojavnosti. Supojavnost bismo tako mogli definirati kao zajedničko pojavljivanje parova jedinica u više bibliografskih zapisa (Persson, Danell i Schneider, 2009). Datoteka na kojoj ćemo isprobati funkciju supojavnosti jest naša OUT datoteka. Trebalo bi zapamtiti, u svom članku napominju Persson, Danell i Schneider (2009), da OUT datoteka sadrži određene ekstrahirane jedinice iz metapodatkovnih polja DOC datoteke. Stoga te ekstrahirane jedinice i definiraju tip analize supojavnosti koju je moguće načiniti. Na primjer, ukoliko nam se OUT datoteka sastoji od autora ekstrahiranih iz DOC datoteke, tada ćemo moći napraviti analizu supojavnosti autora. Važno je zapamtiti, napominje u svom radu Pilkington (2018), da neke ekstrahirane jedinice imaju jako malu frekvenciju pojavnosti i one će biti nebitne za analizu supojavnosti. U svrhu otkrivanja i brisanja takvih jedinica sa niskom frekvencijom biti će nam potrebna CIT datoteka. Ovdje također Persson, Danell i Schneider (2009) ističu da je važno zapamtiti kako je CIT datoteka ona koja



sadrži frekvencijsku analizu jedinica sadržanih u OUT datoteci. U CIT datoteci ćemo, slično kao i u Excelu, odabrati neprekinuti niz jedinica tako da odaberemo prvu jedinicu u nizu i zatim pritisnemo tipku *SHIFT* te odaberemo zadnju jedinicu u nizu. Ukoliko pak želimo pridodati neke od jedinica iz CIT datoteke koje nisu u nizu sa već označenima, tada je potrebno pritisnuti tipku *CTRL* i označiti bilo koju od jedinica u CIT datoteci.

Sljedeći korak nam je identificirati relacije supojavnosti među označenim jedinicama. Prvo trebamo odrediti na kojoj će se datoteci analiza napraviti, a to je uvijek OUT datoteka. Sljedeći korak je sama analiza supojavnosti u BibExcel-u, a ona se pokreće preko izbornika „*Analyze* → *Co-occurrence* → *Make pairs via listbox*“ kao što je vidljivo i iz priložene slike 14.



Slika 14, Analiza supojavnosti u Bibexcel-u  
Preuzeto sa: BibExcel – A toolbox for bibliometricians, by Olle Persson. Version 2016-02-20

Ishod analize supojavnosti u BibExcel-u je COC datoteka što je skraćenica od eng. co-occurrence. COC datoteka sadrži frekvencije zajedničkog pojavljivanja jedinica odabranih za analizu temeljenih na svim mogućim uparivanjima jedinica iz OUT datoteke da bi se utvrdilo koliko se puta jedinice pojavljuju zajedno kroz zapise što možemo vidjeti iz slike 15 temeljene na članku Persson, Danell i Schneider (2009).

COC freq.	Cited reference #1	Cited reference #2
42	KATZ JS 1997	MELIN G 1996
42	BEAVER DD 1978	BEAVER DD 1979
34	WHITE HD 1981	WHITE HD 1998
33	PERSSON O 1994	WHITE HD 1981
33	BEAVER DD 1978	LUUKKONEN T 1992
31	BEAVER DD 1979	LUUKKONEN T 1992
30	KATZ JS 1997	LUUKKONEN T 1992
28	MELIN G 1996	NEWMAN MEJ 2001

Slika 15, Izlazni podatci analize supojavnosti autora temeljene na COC datoteci u BibExcel-u (prema Persson, 2009)

Uz analizu supojavnosti u BibExcel-u je još moguće napraviti analizu kocitiranosti, analizu bibliografske uparenosti gdje se gleda ima li dva ili više dokumenta koja u svojim bibliografijama spominju neki treći dokument, analizu koautorstva te analizu zajedničke pojave riječi u naslovu i/ili sažetku članaka, no za potrebe ovoga rada nećemo detaljnije ići u opise navedenih analiza.

#### 4.1.4. KREIRANJE PAJEK DATOTEKE U BIBEXCEL-U

Pajek je besplatan program namijenjen nekomercijalnoj uporabi za operativni sustav Windows, a rabi se za analizu i vizualizaciju velikih mreža.

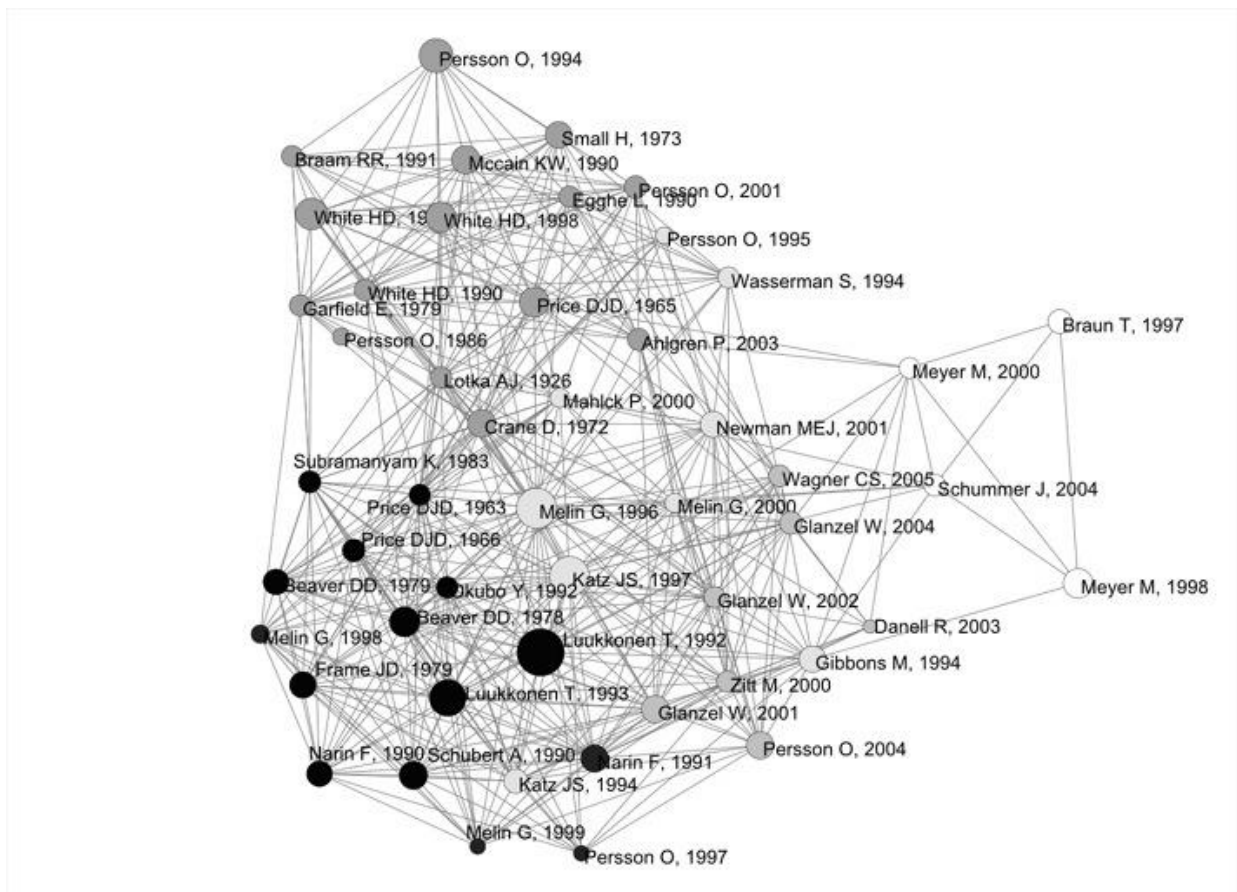
U BibExcel-u možemo stvoriti tri vrste Pajek datoteka, a to su:

- a) Mrežne datoteke (NET datoteka)
- b) Vektorske datoteke (VEC datoteka)
- c) Datoteke koje sadrže informacije o članstvu u klasteru (CLU datoteka)

U nastavku ćemo opisati kako stvoriti te datoteke u BibExcel-u kako bismo ih mogli koristiti za vizualizaciju položaja Ollea Perssona u prostoru dokumenata koje analiziramo. Kao završnu vizualizaciju koristiti ćemo sliku 15 iz članka Persson, Danell i Schneider (2009).

- a) Ukoliko želimo stvoriti navedene datoteke za Pajek u BibExcel-u potrebno je slijediti određene korake. Uvijek je prvo potrebno stvoriti NET datoteku koja se stvara iz nedavno spomenute COC datoteke i to iz menija: „*To Pajek → Create net-file from coc-file*“. Kao posljedica ove naredbe nastaje NET datoteka, a nakon što smo nju stvorili moguće je nastaviti stvarati VEC datoteku ili CLU datoteku i to bilo kojim redoslijedom.
- b) VEC datoteku stvaramo iz CIT datoteke naredbom „*To Pajek → Create vec-file*“. U vizualizaciji intelektualnog konteksta Ollea Perssona; Persson, Danell i Schneider (2009), u svom članku žele prikazati koji su dokumenti najviše citirani tako da su kreirali vektorsku datoteku koja sadrži informacije o citatnoj frekvenciji kocitiranih radova.
- c) Treća vrsta datoteke koja se u BibExcel-u može stvoriti za Pajek je CLU datoteka. Ona sadrži informacije o klasterima, a može se kreirati pomoću algoritma kojeg je razvio Olle Persson (Persson party cluster) (Persson, Danell i Schneider, 2009). Da bismo napravili particije na matrici kocitata odabiremo COC datoteku i iz menija odabiremo „*Analyze → Co-occurrence → Cluster pairs*“. BibExcel će tada stvoriti tri datoteke koje sadrže informacije o klasterima stvorenim iz COC datoteke. Te datoteke imaju ekstenzije \*.pe2, \*.p3 i \*per, a datoteka koja nam je potrebna za stvaranje CLU datoteka je \*pe2 datoteka koju označujemo, te iz menija odabiremo naredbu „*To Pajek → Create clu-file*“.

Uvoz NET, VEC i CLU datoteka u Pajek isto je tako jednostavan. Nakon što otvorimo sve tri datoteke u Pajek-u kroz opciju menija „*Draw → Draw-Partition-Vector*“ stvoriti će se vizualizacija naše analize u BibExcel-u kao što je vidljivo na slici 16 (Persson, Danell i Schneider, 2009).



Slika 16, Vizualizacija intelektualnog sadržaja Ollea Perssona u Pajek-u (prema Persson, 2009)

Sada kada je analiza znanstvenog mapiranja pomoću Pajek-a gotova možemo konstatirati kako je iz generirane mape vidljivo da se u gornjem lijevom kutu nalaze publikacije primarno korištene u kontekstu informacijskih znanosti, a čvorovi u tom klasteru obojani su svjetlo sivom bojom. Nadalje, u srednjem i donjem kutu mape nalaze se publikacije koje se bave istraživačkom suradnjom i internacionalizacijom znanosti čiji su čvorovi obojani crnom bojom, a naposljetku, na desnoj strani mape nalaze se članci koji se bave evaluacijom istraživanja i transferom tehnologije čiji su čvorovi obojani bijelom bojom. (Persson, Danell i Schneider, 2009).

#### 4.2. SciMAT – SCIENCE MAPPING ANALYSIS SOFTWARE TOOL

SciMAT je noviji softverski alat otvorenog koda koji izvodi analizu znanstvenog mapiranja unutar longitudinalnog (uzdužnog) okvira. SciMAT je osmišljen u istraživačkoj grupi Sci2s na Sveučilištu u Granadi u Španjolskoj od strane M.J. Cobo, A.G. López-Herrere, Herrere-Viedmana i F. Herrere, a podržava ga Španjolsko Ministarstvo prosvjete i znanosti. SciMAT sadrži različite module koji

pomažu analitičarima da provedu sve korake znanstvenog mapiranja. SciMAT sadrži tri ključna obilježja u odnosu na druge alate za znanstveno mapiranje znanosti pa tako i BibExcel, a to su: a) snažan procesni modul za čišćenje sirovih bibliometrijskih podataka, b) koristi bibliometrijske mjere za proučavanje odjeka svakog proučavanog elementa i c) sadrži snažan čarobnjak za konfiguraciju analize.

#### **4.2.1. GLAVNE KARAKTERISTIKE SCIMAT-A**

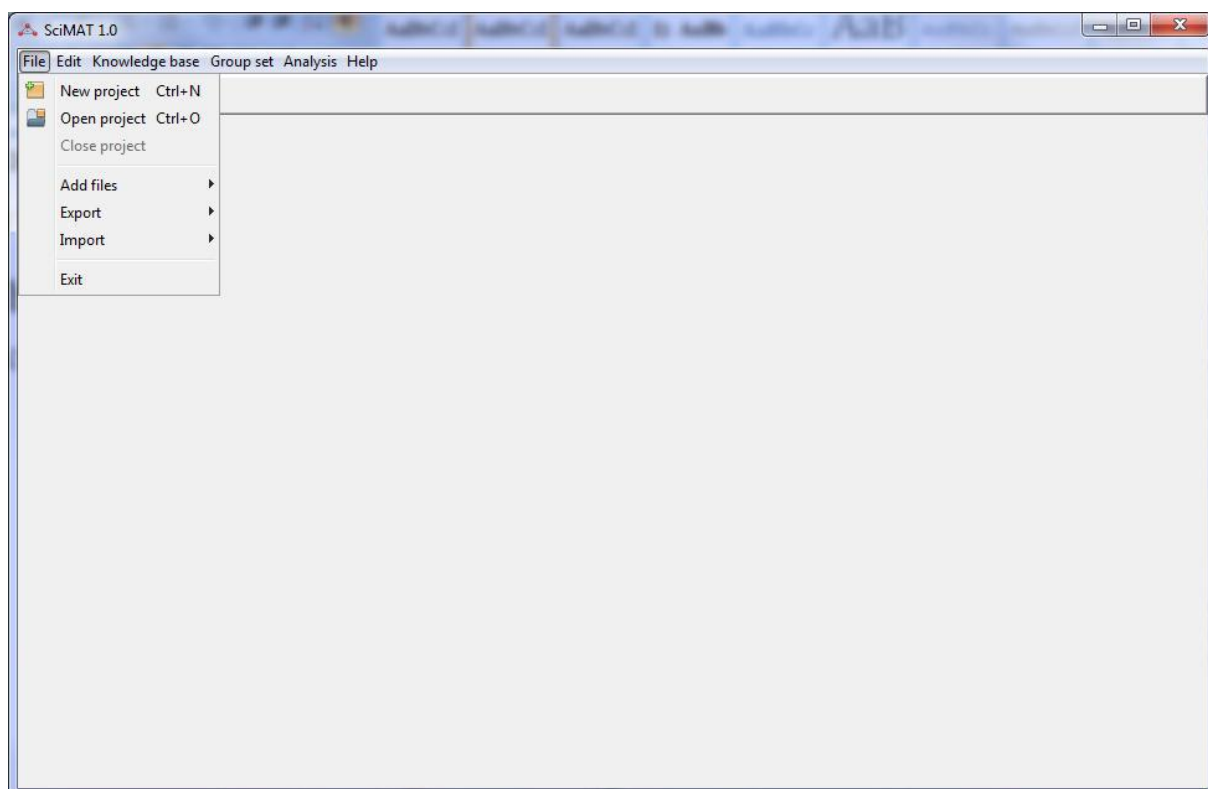
- Uključuje sve potrebne module za izvršavanje svih potrebnih koraka tijekom rada u znanstvenom mapiranju (za razliku od BibExcel-a).
- U njega su implementirane metode kojima je moguće izgraditi većinu bibliometrijskih mreža, njihovo normaliziranje za izgradnju znanstvenih mapa, odnosno karata, koristeći algoritme za klasteriranje te različite tehnike za vizualizaciju istih korisne za interpretaciju izlaznih podataka.
- Sadrži širok raspon alata za predobradu podataka kao što su alati za otkrivanje dvostrukih i pogrešno napisanih podataka, alati za redukciju viška podataka itd. (Cobo, López-Herrera, Herrera-Viedma i Herrera, 2012).
- U svom članku (Cobo i dr., 2012) također navode da SciMAT može generirati znanstvene mape obogaćene bibliografskim mjerama na temelju citata kao što su: h-indeks, g-indeks, hg-indeks ili g2-indeks.
- SciMAT je podijeljen u tri različita modula:
  1. Modul posvećen upravljanju baza podataka i njihovim entitetima
  2. Modul odgovoran za provođenje analiza znanstvenog mapiranja i
  3. Modul za vizualizaciju generiranih rezultata i mapa

#### **4.2.2. BAZA ZNANJA**

SciMAT generira bazu znanja iz skupa znanstvenih dokumenata u kojima su pohranjeni odnosi različitih entiteta koji se odnose na svaki dokument, a to su autori, ključne riječi, časopisi, reference i dr. Ovakva struktura pomaže analitičarima u uređivanju i procesuiranju baze znanja kako bi mogli poboljšati kvalitetu podataka i kako bi na kraju mogli dobiti bolje rezultate analize znanstvenog mapiranja. Baza znanja se sastoji od 16 entiteta. Glavni entitet je sam dokument koji sadrži informacije kao što su naslov, sažetak, citate itd. Dokument sadrži i niz drugih informacija kao što su autor, pripadnost, godina izdavanja i svaka se od tih informacija smatra entitetom u bazi znanja.

#### 4.2.2.1. UPRAVITELJ BAZE ZNANJA

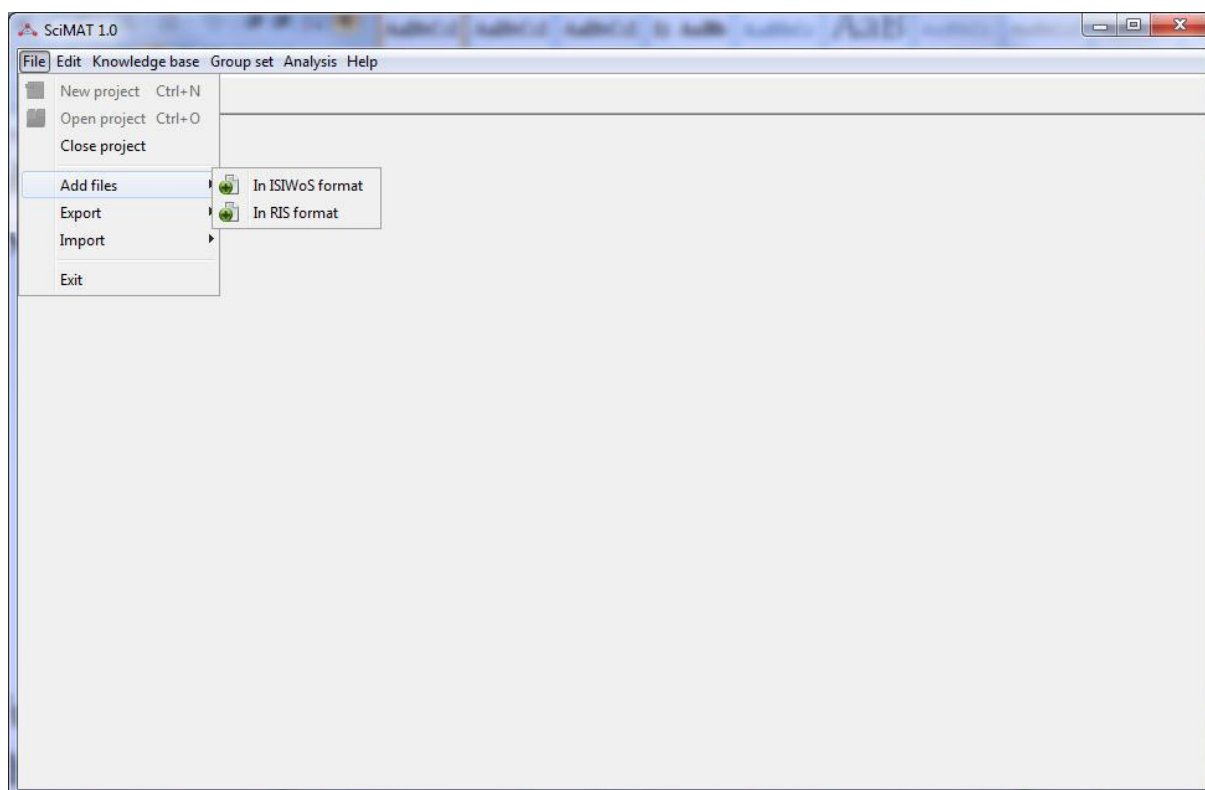
Upravitelj baze znanja odgovoran je za njenu izgradnju kako za uvoz podataka iz različitih bibliografskih izvora, tako i za čišćenje podataka i popravljavanje mogućih pogrešaka u uvezenim entitetima. Sam se upravitelj može smatrati prvom fazom u predprocesuiranju podataka. Ukoliko je odabran novi projekt tada će se pojaviti novi prozor koji će zahtijevati upisivanje puta prema kojem će se baza znanja spremiti kao što je vidljivo na slici 16 i naravno ime same baze znanja. Uz to, moguće je navesti i ekstenziju navedene datoteke.



Slika 17, Pokretanje nove baze znanja u SciMAT-u

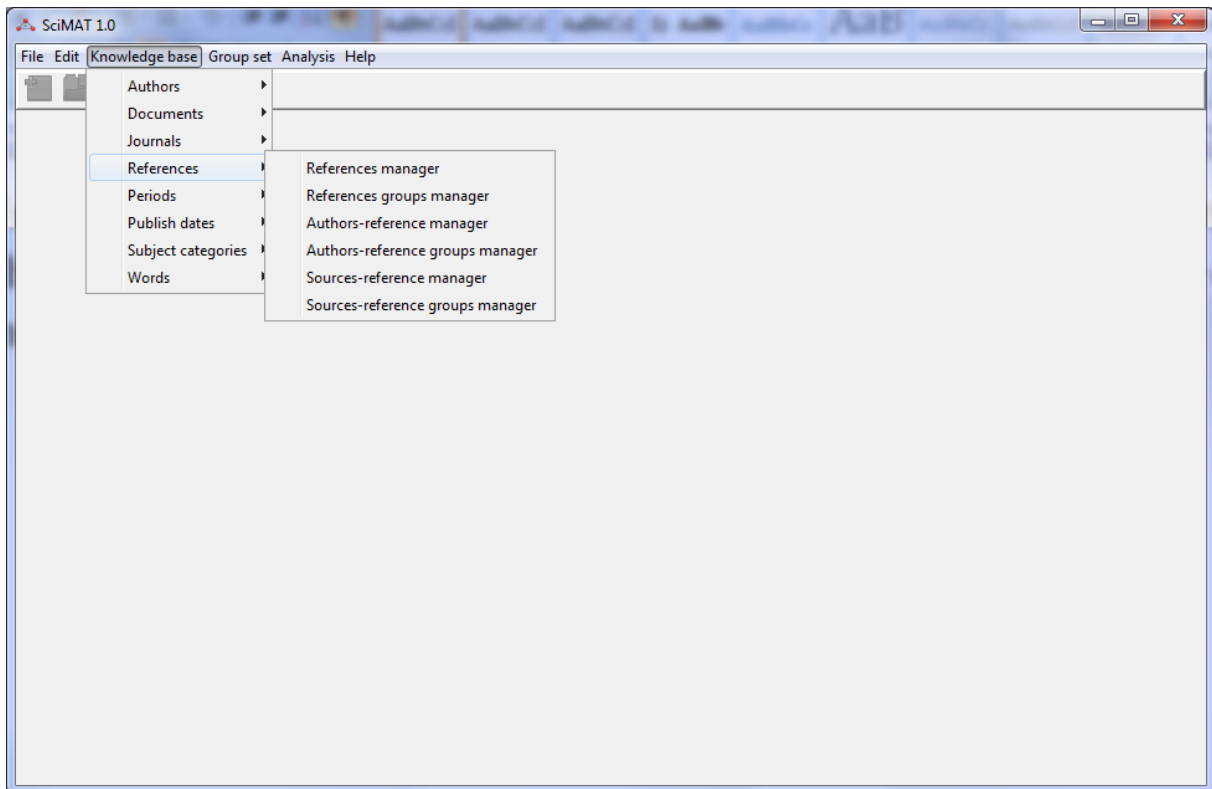
Preuzeto sa: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>

SciMAT koristi SQLite bazu podataka za pohranu baze znanja izgrađenu u prethodnom koraku. Zahvaljujući tim mogućnostima baza se znanja može otvoriti s preglednikom baze podataka koji čita SQLite datoteke. Opcija dodavanja datoteka omogućuje korisniku dodavanje bibliografskih podataka, izvezenih iz bibliografskih baza podataka u bazu znanja. SciMAT može čitati bibliografske podatke izvezene u ISI Web of Knowledge formatu (ISI-CE) ili pak RIS (Scopus) formatu.

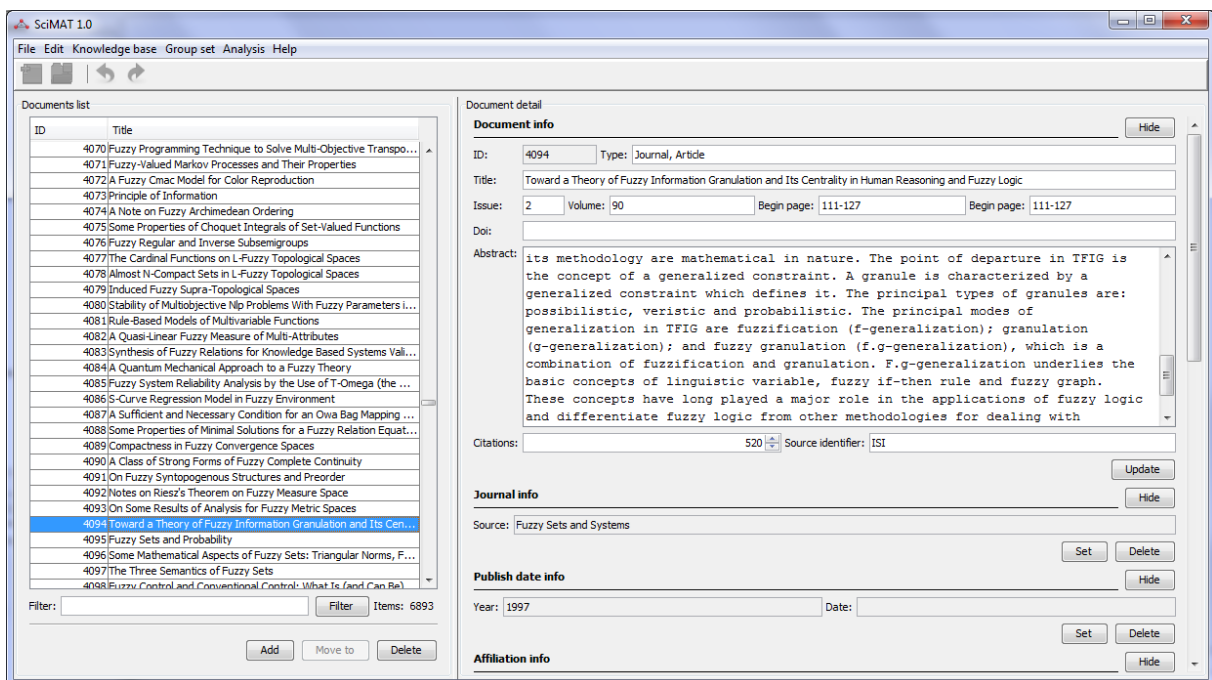


Slika 18, Uvoz baze znanja u ISI ili RIS formatu u SciMAT-u  
Preuzeto sa: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>

Pod izbornikom Baza znanja moguće je pronaći upravitelj za 16 entiteta, a uz to postoji i upravitelj za svaki od tih entiteta. Zahvaljujući tim upraviteljima svaki je entitet moguće uređivati, a uz to je moguće i mijenjati njihove attribute. Primjer Upravitelja baze znanja i Upravitelja dokumenta prikazan je na slikama 19 i 20. Kao što je vidljivo na slici 20, gdje je prikazan Upravitelj dokumenata, na popisu dokumenata na lijevoj strani odabran je jedan od najcitiranijih članaka, a na desnoj se strani nalaze povezane informacije kao što su naslov, sažetak, publikacijski podatci, citati itd. Upravitelj dokumenata nam dopušta i dodavanje novog dokumenta, kao što je vidljivo na slici 21, a to se može učiniti ručno, ispunjavajući svaki entitet, brisanje skupa dokumenata ili pridruživanje više dokumenata u jedan skup. (Cobo, López-Herrera, Herrera-Viedma i Herrera, 2012). Trebalo bi naglasiti da svi upravitelji imaju istu strukturu. Na lijevoj se strani prikazuje popis entiteta, dok se na desnoj strani prikazuju polja odabranog entiteta i njegovi odnosi s ostalim entitetima.

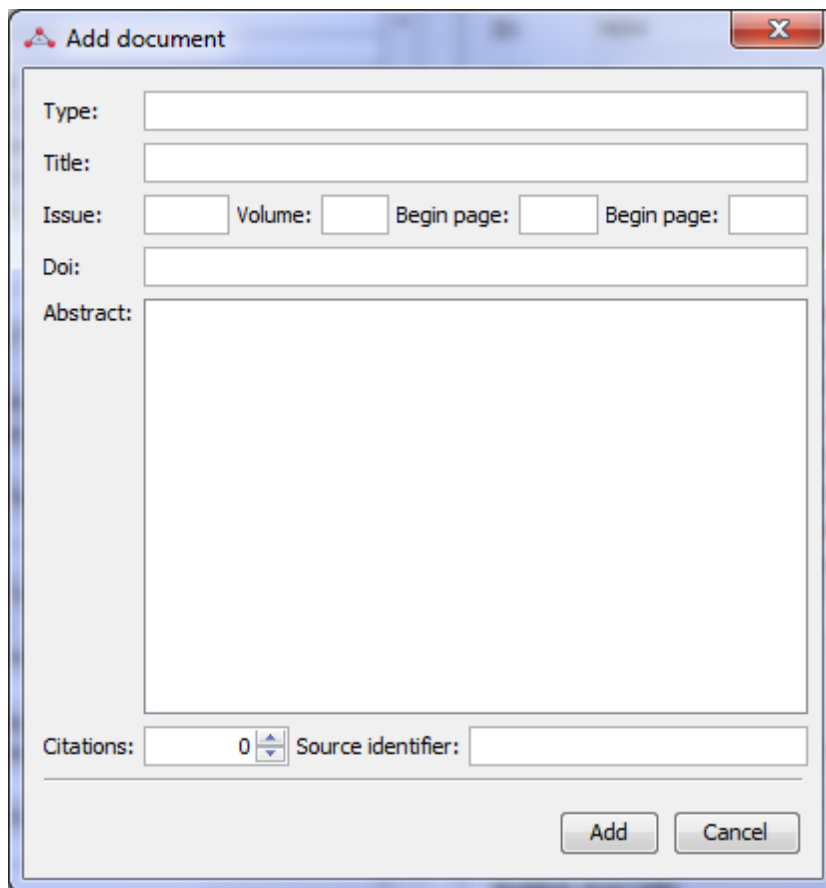


Slika 19, Upravitelj baze znanja sa prikazanim entitetima u SciMAT-u  
 Preuzeto sa: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>



Slika 20, Upravitelj dokumenta sa informacijama o dokumentu na desnoj strani u SciMAT-u  
 Preuzeto sa: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>





Slika 21, Dodavanje novog dokumenta preko Upravitelja dokumenata u SciMAT-u  
Preuzeto sa: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>

#### **4.2.3. ČAROBNJAK ZA KONFIGURACIJU ANALIZE ZNANSTVENOG MAPIRANJA**

Ovaj čarobnjak jedan je od najvažnijih modula grafičko korisničkog sučelja SciMAT-a. Odgovoran je za stvaranje određene konfiguracije za provođenje znanstvenog mapiranja znanosti na način da korisnik može odabrati metode i algoritme koji će se koristiti da bi se obavio svaki pojedini korak. Nakon što korisnik odredi željenu konfiguraciju Cobo i dr. (2012) u svom članku navode da se navedena konfiguracija šalje modulu koji je odgovoran za analizu koji će pomoću SciMAT API-a obaviti analizu znanstvenog mapiranja. Naposljetku, rezultati se pohranjuju u datoteku i zatim šalju modulu za vizualizaciju.

Čarobnjak za konfiguraciju analize sastoji se od jedanaest uzastopnih koraka:

- i. U prvom koraku korisnik bi trebao odabrati razdoblje koje želi analizirati. Svako pojedino razdoblje će proizvesti zasebnu kartu/mapu. Ta će razdoblja biti korištena ili u longitudinalnoj (uzdužnoj) ili vremenskoj analizi kako bi se mogla proučiti strukturalna evolucija polja.
- ii. U drugom koraku potrebno je odabrati jedinice analize. Kao jedinice analize korisnik može odabrati bilo koju od 5 grupa koje postoje u bazi znanja: autorska grupa, referentna grupa autora, grupa izvora, referentna grupa i grupa riječi.
- iii. U trećem koraku odrađuje se redukcija podataka, a tu SciMAT dopušta filtriranje podataka koristeći minimalni frekvencijski prag.
- iv. Četvrti korak omogućuje odabir načina na koji će se mreža izgraditi. To može biti supojavnost ili uparenost.
- v. Peti korak označava smanjenje mreže gdje se za svako razdoblje u analizi određuje vrijednost početka i kraja razdoblja.
- vi. U šestom koraku se obavlja normalizacija mreže.
- vii. Sedmi korak označava odabir algoritma za klasteriranje koji se koristi za dobivanje mape i njezinih pripadajućih klastera i podmreža.
- viii. Osmi korak odnosi se na odabir između 5 mogućih alata za mapiranje dokumenata u analizi izvedbe.
- ix. Deveti korak jest odabir performansi i kvalitativnih bibliometrijskih mjera kao što su minimum, maksimum, suma, prosječna citiranost ili kompleksnijih mjera kao što su h-indeks, g-indeks, hg-indeks.
- x. Deseti korak označava odabir mjera sličnosti korištenih za izgradnju znanstvene mape i mapa koje se s njom preklapaju.
- xi. Najzad, jedanaesti korak je odgovoran za analizu znanstvenog mapiranja i taj se proces u bilo kojem trenutku može prekinuti. Na samom kraju analiza se treba spremati kroz novi prozor koji će se pojaviti, a krajnji rezultat analize se tada vizualizira kroz modul za vizualizaciju.

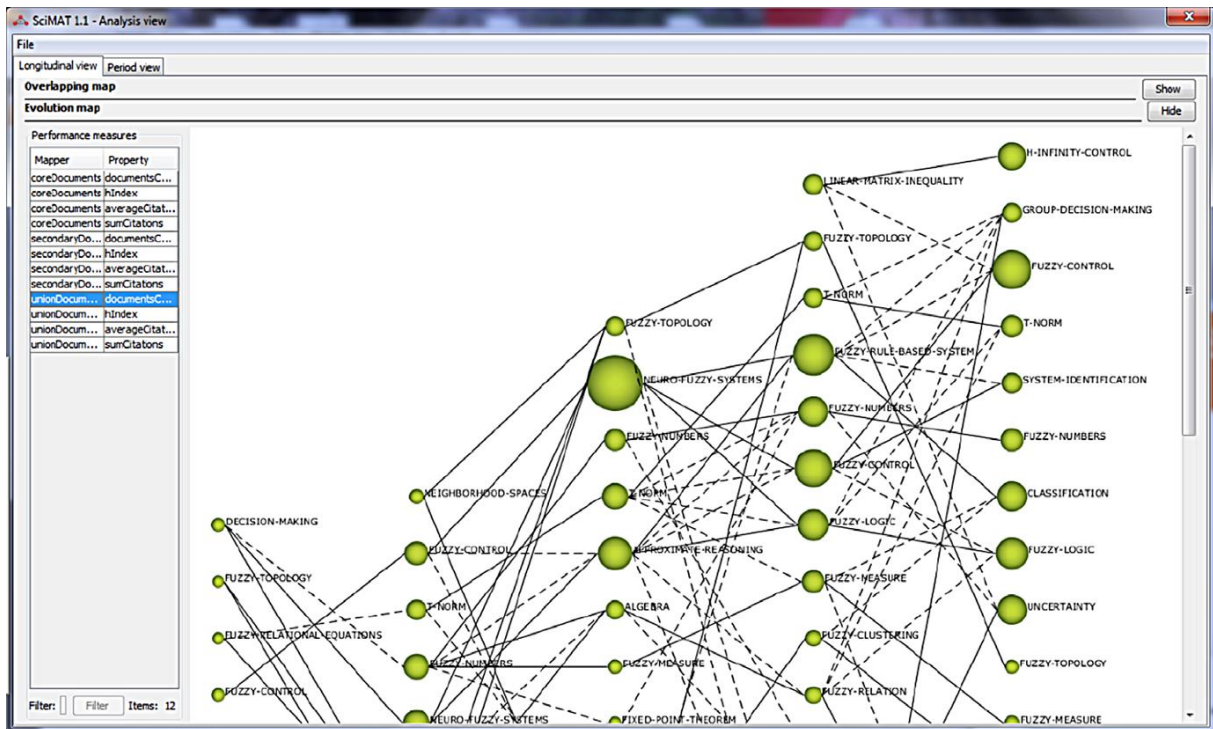
#### **4.2.4. MODUL ZA VIZUALIZACIJU**

Nakon što je odrađena analiza znanstvenog mapiranja rezultati se vizualiziraju kroz modul za vizualizaciju. Modul za vizualizaciju ima dva prikaza:

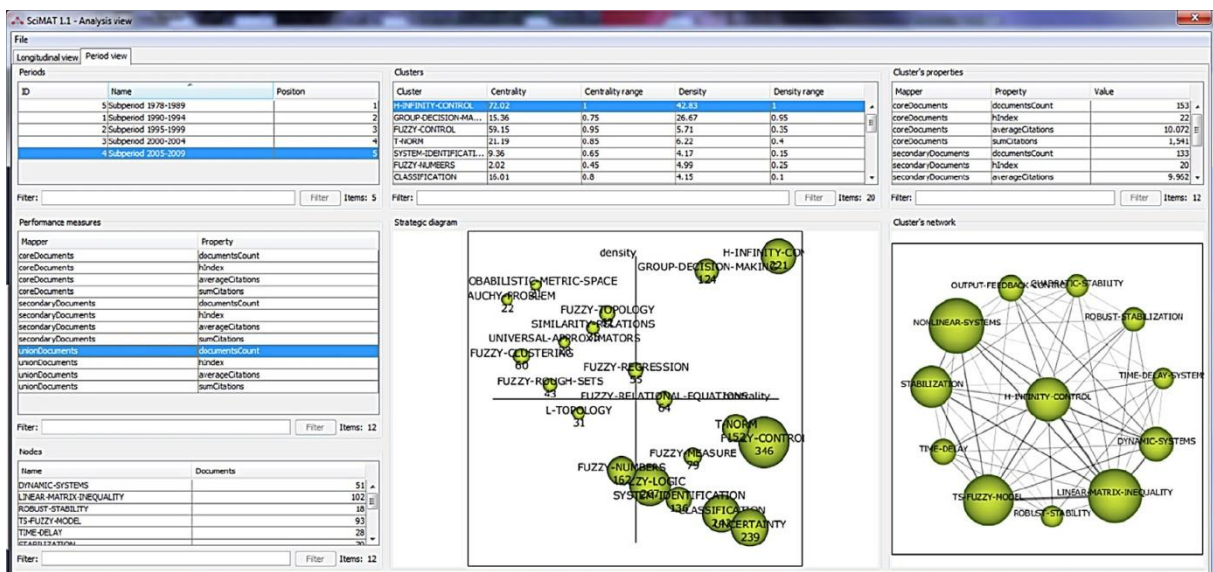
- 1.) Longitudinalni (uzdužni) prikaz u kojem se prikazuju mapa preklapanja u gornjem dijelu i evolucijska mapa u donjem dijelu prikaza. Ovaj način prikaza

omogućuje nam otkriti evoluciju klastera tijekom različitih razdoblja, napominju Cobo i dr. u svom radu (2012).

- 2.) Vremenski prikaz prikazuje detaljne informacije za svako razdoblje, njihov strateški dijagram te bibliografske mjere, mrežu i pripadajuće čvorove za svaki klaster, što je jasnije vidljivo na slikama 22 i 23.



Slika 22, Longitudinalni (uzdužni) prikaz vizualizacije znanstvene mape u SciMAT-u mapiranja (prema Cobo, M.J. i dr.,2012)



Slika 23, Vremenski prikaz vizualizacije znanstvene mape u SciMAT-u mapiranja (prema Cobo, M.J. i dr.,2012)

Modul za vizualizaciju svoja izvješća može izraditi u HTML-u i LaTeX formatu, a izvesti ih može u PNG i SVG formatu tako da ih korisnik kasnije može uređivati. Uz to, same mreže klastera i mape mogu se izvesti i u Pajek formatu.

### **4.3. USPOREDNA ANALIZA BIBLIOMETRIJSKIH ALATA BIBEXCEL I SciMAT**

Nakon što smo prošli kroz osnovne karakteristike dvaju bibliometrijskih alata, BibExcel-a i nešto novijeg SciMAT-a, vrijeme je da damo zaključne opaske o navedenim bibliometrijskim alatima. Oba bibliometrijska alata pokazuju različite karakteristike, na primjer SciMAT posjeduje poseban i to vrlo moćan modul za vizualizaciju analiza znanstvenog mapiranja, dok BibExcel uopće nema mogućnost vizualizacije nego se izlazni rezultati iz BibExcel-a moraju u svrhu vizualizacije izvesti u neki drugi program kao na primjer Pajek. Ono što bismo mogli konstatirati, a da je podjednake kvalitete u oba analizirana alata, jest vrlo jednostavno preuzimanje sa mrežnih stranica. Tako se BibExcel jednostavno preuzima sa mrežne stranice <http://homepage.univie.ac.at/juan.gorraiz/bibexcel/> dok se SciMAT isto tako vrlo jednostavno preuzima sa mrežne stranice <http://sci2s.ugr.es/scimat/>. Uz to, bez većih komplikacija oba se alata mogu lagano spremirati na tvrdi disk računala od kuda se mogu i pokrenuti. Samo grafičko korisničko sučelje oba alata je donekle zastarjelo i ne baš intuitivno te nas pomalo podsjeća na grafičko korisničko sučelje dobro nam poznatog Informacijskog sustava visokih učilišta (ISVU). Također, možemo naglasiti da oba alata imaju donekle detaljna korisnička uputstva, no moramo priznati da nisu niti približno detaljna koliko bi trebala biti kako bi korisnici iz navedenih alata mogli iskoristiti njihov maksimum. Ono što bi mogli konstatirati jest da BibExcel ima neznatno bolja uputstva od SciMAT-a, ali niti ona nisu dovoljno detaljna za potpuno korištenje navedenog alata. Moglo bi se reći da su uputstva za oba alata napisana za korisnike koji već imaju priličnog iskustva u bibliografskim analizama, dok bi korisnici početnici mogli imati problema čitajući samo uputstva bez da ulože priličnog truda u samostalno proučavanje navedenih alata. Jedna od očitih prednosti SciMAT-a nad BibExcel-om jest ta što je SciMAT programiran u Java-i, a samim time može ga se pokrenuti preko operativnog sustava Windows, macOS ili Linux, dok se BibExcel može pokrenuti samo preko operativnog sustava Windows. Ono što SciMAT čini privlačnijim jest i to što kroz modul vizualizacije omogućava i longitudinalni (uzdužni) prikaz, što BibExcel ne podržava. Sama se analiza u oba alata donekle razlikuje, no

to prepuštamo analitičarima da odluče koji od navedenih alata bi im mogao biti privlačniji. Za kraj bismo samo naglasili da ne postoji niti jedan bibliografski softverski alat koji bismo smatrali najboljim. Prema tom zaključku, a s time se slaže i Cobo i dr. (2011), smatramo da bi se kvalitetna analiza znanstvenog mapiranja trebala napraviti korištenjem različitih bibliografskih alata (ne samo BibExcel-om ili SciMAT-om) kako bi se na što bolji način uspjela prikupiti nova znanja iz različitih perspektiva. Na posljetku, svaki od alata ima mogućnost različite vizualizacije dobivenih rezultata, a na analitičaru je da odluči koji od oblika mu je najpogodniji i najprimjereniji u danom trenutku. Takva suradnja između različitih bibliografskih softverskih alata može nam dati jednu pozitivnu sinergiju koja nam omogućava izvući skriveno znanje iz gomile nepovezanih podataka.

## **ZAKLJUČAK**

Neprijeporno je da o ovoj temi postoji neizmjereno mnogo izvora informacija i da bi se ovakav rad mogao napisati na nekoliko načina, a da se ne izgubi smisao same teme i rada. Nakon podužeg razmišljanja o tome kako koncipirati strukturu ovoga rada odluka je pala da se pokušaju dotaknuti glavni pojmovi bibliometrije kao primarne teme rada, povijesni razvoj ostalih informetrijskih grana i naravno glavne zakone bibliometrije s obzirom da bi ovaj rad trebao predstavljati uvod u ovu tematiku za osobe koje se s ovom temom nisu još susrele. Da bi ovaj rad bio što kompaktnije zaokružen u cjelinu, na kraju prvoga teorijskog dijela rada smo dodatno pojasnili pojam faktor odjeka, bez kojega se ovaj rad ne bi mogao u potpunosti shvatiti. Želeći sistematizirati rad u što boljem i jasnijem obliku, krenulo se od povijesnog razvoja bibliometrije pa sve do sadašnjosti kada se spominje krajnji doseg današnje bibliometrije, a to su bibliografske baze podataka kao sekundarni izvori informacija koje se temelje na citatnim analizama članaka iz znanstvenih časopisa. Doduše, pišući ovaj rad uvidjeli smo u nekoliko navrata da, koliko god proširivali temu, nikada nećemo moći obuhvatiti, po nama, sve važne informacije za ovakav rad, osim ako se ne odlučimo, kao i neki autori citirani u ovome radu, o ovoj temi napisati knjigu. S druge strane, ovaj rad bi trebao prikazivati osnovne značajke, zakonitosti, zanimljivosti i probleme bibliometrije kao discipline informetrije, stoga smatramo da smo uspjeli u donekle konciznom obliku zainteresiranim osobama približiti temu na zanimljiv način. Nakon što smo detaljnije pojasnili sve teorijske pojmove bibliometrije

kao znanstvene discipline unutar informetrije bilo je vrijeme da detaljnije opišemo dva bibliometrijska softverska alata, a odluka je pala na BibExcel i SciMAT. U tom drugom, praktičnom dijelu rada htjeli smo pobliže zainteresiranoj publici približiti mogućnosti koje ovi alati nude te pri samom kraju drugog, praktičnog dijela rada, načiniti usporednu analizu istih kako bi zainteresirane osobe na jednom mjestu imale informacije na temelju kojih će se u budućnosti, možda, odlučiti i koristiti neki od dva opisana alata za neku njihovu buduću bibliometrijsku analizu. Bilo kako bilo, po mom mišljenju smatram da smo uspjeli ovako opsežnu temu, kako teorijski tako i praktično, sažeti u oblik koji će biti prihvatljiv široj zainteresiranoj publici.

## LITERATURA

1. Åström, F. i dr. Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday // Special volume of the e-zine of the ISSI, 05-S, 2009.  
URL: <http://portal.research.lu.se/portal/files/5902071/1458992.pdf>  
(15.07.2018)
2. Bawden, D. and Robinson L. Introduction to Information Science // Facet Publishing, London: 2012. Str. 165 – 187.
3. BibExcel. URL: <http://homepage.univie.ac.at/juan.gorraiz/bibexcel/>  
(02.06.2018)
4. Cobo, M.J. i dr. Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools. // Journal of the American Society of Information Science and Technology, 63(8), 2011.  
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21525> (15.07.2018.)
5. Cobo, M.J. i dr. SciMAT: A New Science Mapping Analysis Software Tool. // Journal of the American Society of Information Science and Technology, 62(7), 2012.  
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22688> (16.07.2018.)
6. Cobo, M.J. i dr. SciMAT Version 1.0 User guide.  
URL: <http://sci2s.ugr.es/scimat/software/v1.01/SciMAT-v1.0-userGuide.pdf>  
(16.07.2018.)
7. Hood, W. W. and Wilson, S. C. The literature of bibliometrics, scientometrics, and informetrics // Scientometrics, 52 (1), Budimpešta: 2001. Str. 291 – 314.
8. Jokić, M. Bibliometrijski aspekti vrednovanja znanstvenog rada // Sveučilišna knjižara, Zagreb: uredio Damir Mikuličić, 2005.
9. Olujčić-Vuković, V. Kvantitativna istraživanja procesa stvaranja, prijenosa i korištenja informacija – nužnost jedinstvenog pristupa // Vjesnik bibliotekara Hrvatske 50 (1/2), Zagreb: 2007. Str 27 – 42.
10. Pečarić, Đ. Razvoj informacijskih znanosti u Hrvatskoj. Bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978. – 2007. Dr. sc. Sveučilište u Zagrebu; 2010.
11. Pilkington, A. BibExcel: Quick Start Guide to Bibliometrics and Citation Analysis. Innovation Discovery: Network Analysis Of Research And Invention

Activity For Technology Management. // World Scientific Publishing, 2018. Str. 585 – 600.

12. Pehar, F. Od statističke bibliografije do bibliometrije. Povijest razvoja kvantitativnog pristupa istraživanju pisane riječi // Libellarium: časopis za istraživanje pisane riječi, knjige i baštinskih ustanova, 3 (1), Zadar:, 2011. Str 1 – 28.
13. SciMAT – Science Mapping Analysis Tool. URL: <http://sci2s.ugr.es/scimat/> (02.06.2018.)



# Bibliometrijska analiza pomoću alata BibExcel i SciMAT

## SAŽETAK

Ovaj završni rad pisan je u svrhu upoznavanja zainteresirane publike s pojmom i osnovnim teorijskim značajkama bibliometrije te osnovama korištenja i usporednim opisom dva bibliometrijska alata BibExcel-a i SciMAT-a. Cilj rada biti će sažeti široku teorijsku tematiku u jednu konciznu i kompaktnu cjelinu koja će uz usporedni opis dvaju navedenih alata dati svim zainteresiranima mogućnost upoznavanja sa disciplinom bibliometrije te mogućnost odabira korištenja jednog od dvaju navedenih bibliometrijskih alata temeljem usporednog opisa istih. Prvi, teorijski dio, objasniti će povijesni razvoj bibliometrije te će dati uvid u tri najvažnija bibliometrijska zakona. Završni dio teorijskog dijela objasniti će pojmove faktora odjeka i indeksa brzine citiranja. Drugi, praktični dio završnog rada, odnositi će se na usporedni opis dvaju bibliometrijskih alata, BibExcel-a i SciMAT-a, koji će se sastojati od prikaza načina instalacije navedenih alata na samo računalo, pojašnjenja koliko ih je lako ili teško koristiti, na kojim se tehnologijama koriste, pojašnjenja imaju li detaljne i koliko kvalitetne upute za korištenje, vizualne usporedbe navedenih alata te općenitih prednosti i nedostataka jednog i drugog alata. Na ovaj način, uz teorijski i praktični dio završnog rada, omogućiti ćemo zainteresiranoj publici upoznavanje s teorijom discipline bibliometrije te, kroz usporedbu dvaju navedenih bibliometrijskih alata, mogućnost odabira jednog od njih, ukoliko će im u budućnosti zatrebati u svrhu bibliometrijske analize nekoga od traženih pojmova.

**Ključne riječi:** Bibliometrija, Lotkin zakon, Bradfordov zakon, Zipfov zakon, citatna analiza, faktor odjeka, SciMAT, BibExcel, kocitiranost, supojavnost, frekvencijska distribucija.

# **Bibliometric analysis using BibExcel and SciMAT tools**

## **SUMMARY**

This paper was written for the purpose of introducing the public with the concept and the basic theoretical features of bibliometry and the basics of use and comparative description of two bibliometric tools, BibExcel and SciMAT. The aim of this paper will be to summarize the broad theoretical knowledge into a concise and compact entirety that will provide interested students with a comparative description of those two tools mentioned above and the opportunity to familiarize themselves with the discipline of bibliometry giving them the possibility of choosing one of two bibliometric tools based on their comparative description. The first, theoretical part of the paper will explain the historical development of bibliometry and provide insight into the three most important bibliometric laws. Final stage of the theoretical part of this paper will explain the terms of impact factor and citation rate index. The second, practical part of this paper will refer to a comparative description of two bibliometric tools, BibExcel and SciMAT, which will consist of describing how to install those tools on a computer and explaining of how easy or difficult it is to use them and which technologies are they based on and clarifying whether they have or not detailed user manuals. Also, this part of the paper will offer visual comparisons of those tools and the general advantages and disadvantages of one and the other. In this way, with the theoretical and the practical part of this paper we will enable the interested audience to get acquainted with the theory of bibliometrics and by comparing the two bibliometric tools listed above give the possibility of selecting one of them if in the future bibliometric analysis of some required terms will be needed.

**Key words:** Bibliometrics, Lotka's law, Bradford's law, Zipf's law, citation analysis, impact factor, SciMAT, BibExcel, co-citation, co-appearance, frequency distribution.