

H-indeks kao novi scientometrijski indikator

H-index as a new scientometric indicator

Maja Jokić

Nacionalna i sveučilišna knjižnica, Zagreb

National and University Library, Zagreb, Croatia

Sažetak

U tekstu se raspravlja o indikatorima vrednovanja znanstvenog rada znanstvenika i časopisa. Naglasak je na značenju i važnosti novog scientometrijskog indikatora *h*-indeksa, koji je uveo fizičar Hirsch 2005. godine i koji su prihvatile vodeće citatne baze kao i šira znanstvena zajednica.

Ključne riječi: bibliometrijski/scientometrijski indikatori; *h*-indeks; čimbenik odjeka (IF)

Vrednovanje znanstvene djelatnosti najčešće se mjeri znanstvenom produktivnošću i njenim odjekom mjerenim citatnim analizama. Citatne analize uključuju mjerenja broja citata, vrste citata, samocitata, npr. autora, koautora, institucije, zemlje, časopisa, ili neovisnih citata. S obzirom na vrednovanje statusa znanstvenika, institucije ili zemlje, nije svejedno u kojim časopisima se objavljuju rezultati istraživanja, u kojoj mjeri su zapaženi, tko ih je zapazio i to registrirao citiranošću. Stoga se često kao indikator u vrednovanju znanstvenog rada pojedinog znanstvenika ili institucije koristi i status časopisa, s obzirom na čimbenik odjeka* (engl. *impact factor*, IF) u kojem je rad objavljen, kao i status časopisa koji određeni rad citiraju. Međutim, upotreba IF, a naročito tzv. standardnog ili Garfieldovog IF, kao jednog od osnovnih indikatora u vrednovanju nečijeg rada, upućuje na zaključak da se radi o nerazumijevanju njegovog pravog značenja. IF časopisa je mjera za frekvenciju kojom je „prosječni članak” u časopisu citiran u određenom razdoblju. Čimbenik odjeka je pomagalo u određivanju kvalitete časopisa, a ne u određivanju kvalitete pojedinog članka, odnosno procje-

Abstract

In this text we discuss about indicators for evaluation of scientific work of both, the individual author, and journal as a whole. Emphasis is put on the value and significance of a new scientometric indicator, *h*-index, which was introduced by physicist Hirsch in 2005, and which was adopted both, by leading citation database providers and by the wider scientific community.

Keywords: bibliometric/scientometric indicators; *h*-index; impact factor

Evaluation of scientific work is most often done by measuring scientific productivity and its impact, through citation analysis. Citation analysis includes measuring the number of citations, types of citations, self-citations (among authors, coauthors, institutions, countries or journals) and “independent” citations. In the evaluation of the status of author, institution or country, it is significant in which journals were the research results published, to what degree were they noticed and who noticed them and formalised that by citing. The status of the journal in which the research is published, as well as the status of the citing journal are some of frequently used indicators for evaluating individual scientists and institutions and are gained through the use of impact factor (IF). However, to use IF in general, and so called standard or Garfield IF in particular, as a basic indicator for the evaluation of an institution or individuals work is to misunderstand its real meaning. Journal IF is a measure of a frequency with which an “average” article was cited during a certain period of time. IF helps in evaluation of a journal’s quality. It is not to be used for evaluation of a single article, or of a

* Stav autora je da je pojam *čimbenik odjeka* (engl. *Impact factor*, IF) primjereniji u odnosu na često korišten doslovni prijevod *čimbenik utjecaja*. Naime, konceptualno IF-om se ne može mjeriti utjecaj jer kroz citate na razini broja, ne možemo tvrditi da je određeni rad imao utjecaj. No, zasigurno možemo tvrditi da su citirani radovi iz nekog časopisa kroz citate imali odjek i da je to utjecalo na vrijednost IF-a.

ne kvalitete pojedinog znanstvenika. IF časopisa može biti samo potencijalni pokazatelj vrijednosti članka i to zato što se pretpostavlja da je prošao strogi recenzijski postupak, a prava vrijednost tog članka dobiva se *a posteriori* tj. brojem citata i potencijalnim utjecajem toga rada na vrijednost IF časopisa.

Fizičar J.E. Hirsch (1), svjestan nedostataka dosadašnjih indikatora vrednovanja znanstvene produktivnosti, broja objavljenih radova i odjeka mjenog kroz ukupan broj citata, prosječan broj citata po radu, broja radova s iznad prosječnim brojem citata, potencijalne vrijednosti radova objavljenih u časopisima s određenim IF, uveo je indikator kojim se može mjeriti širi odjek i prepoznatljiviji utjecaj rada pojedinog znanstvenika, odnosno časopisa. Predložio je samo jedan broj "h-indeks," kao jednostavan i koristan način da se okarakterizira znanstvena aktivnost istraživača. Znanstvenik ima određeni h-indeks ako je svaki od njegovih N_p radova dobio najmanje h citata, dok ostalih ($N_p - h$) radova ima $\leq h$ citata. To u praksi znači da ako autor ima h-indeks 10, tada je objavio 10 ili više radova, pri čemu je njegovih 10 radova dobilo najmanje po 10 citata, dok su njegovi ostali radovi citirani manje od deset puta. Ukupan broj citata u ovom slučaju može biti najmanje 100.

H-indeks, kao scientometrijski indikator, u osnovi služi za usporedbu znanstvenika samo iz istog područja i približno istog radnog iskustva, a ista tvrdnja se odnosi i na časopise. Naime, dva znanstvenika sa sličnim h-indeksima usporedivi su u smislu njihove ukupne znanstvene produktivnosti i odjeka, čak i ako je njihov ukupan broj radova i citata jako različit. Odnosno, usporedbom dva znanstvenika (približno istog radnog iskustva) sa sličnim brojem objavljenih radova i/ili sličnim ukupnim brojem citata, ali različitim h-indeksima, govori u prilog veće „prepoznatljivosti“ znanstvenika s visokim h-indeksom (1). Prema mišljenju Brauna i suradnika (2) h-indeks kombinira na specifičan i izbalansiran način efekte „količine“ (broj objavljenih radova) i „kvalitete“ (broj citata).

Batista i suradnici (3) smatraju da h-indeks ima nekoliko prednosti. Kombinira produktivnost s odjekom, nije osjetljiv na ekstremne vrijednosti u smislu radova bez citata odnosno na radove s natprosječnim brojem citata i izravno omogućava identifikaciju najrelevantnijih radova s obzirom na broj dobivenih citata. Nisu rijetke situacije da znanstvenik objavi nekoliko značajnih radova i da ti radovi dobiju ekstremno puno citata, međutim njegov h-indeks ne bude naročito visok. Čest je slučaj da znanstvenici s visokim h-indeksom rade timski i objavljuju radove s velikim brojem autora (većim od 50) te da se međusobno citiraju, kao što je slučaj npr. u području visoko-energijske fizike. Batista i suradnici (3), kao i van Raan (4), upozoravaju da je kod h-indeksa važno istražiti utjecaj broja autora na ukupan broj citata. Ovi autori su dokazali da što je veći broj autora, to je veći broj samocitata, čime se iz-

single scientist. A journal IF can potentially be an indirect measure for the value of an article, because we can suppose it has passed a strict review procedure, but the real value is gained *a posteriori*, through citation counts and its influence on the IF.

Some standard scientific productivity indicators include the number of publications, impact measured through the total number of citations received, the average number of citations per paper, the number of papers with above average citation count and the potential values of articles gained through IF of journals that have published them. J. E. Hirsch (1), a physicist, was well aware of the shortcomings of these indicators and proposed a new one which would show the recognizable impact of a single scientist, although it can be used for journals as well. He has proposed a single number, the "h index," as a particularly simple and useful way to characterize the scientific output of a researcher. A scientist has index h if h of his N_p papers have at least h citations each and the other ($N_p - h$) papers have $\leq h$ citations each. In practice, this means, that if an author has an h -index of 10, then he has 10 papers published that have a minimum of 10 citations each. The minimal possible total citation count in this case is 100.

As a scientometric indicator, h -index mainly serves for comparing scientists from the same disciplines and with similar work experience. The same can be said when using it for journals. Namely, two individuals with similar h -indices are comparable in terms of their overall scientific impact, even if their total number of papers or citations is very different. Conversely, comparing two individuals (preferably of similar length of work experience in science) with a similar number of total papers, or of total citation count, and very different h -values, the one with the higher h is likely to be the more accomplished scientist (1). According to Braun *et al.* (2) h -index combines the effect of "quantity" (number of publications) and "quality" (citation rate) in a rather specific, balanced way.

Batista *et al.* (3) consider h -index to have several advantages. It combines impact and productivity, it is not sensitive to extreme values in the sense of articles with no citations or hyper-cited articles, and it allows the direct identification of the most relevant works with regard to citations. This indicator does not show highly cited or hyper-cited articles, as well as the articles with citation counts which are below the index. One can frequently see situations where a scientist publishes some significant papers, which have an extremely high citation count, but his or hers h -index is not especially high. A frequent case with scientists who have a high h -index is that they work as a team and quote each other. This is the case, for example, in the field of high energy physics, where the number of authors is often higher than 50. Batista *et al.* (3) and van Raan (4) warn that when using h -index, it is

ravno može povećavati h -indeks, ukoliko se samocitati ne isključuju. S druge strane, važno je imati na umu, da je za neka uža znanstvena područja, npr. koja se tek razvijaju, samocitiranost logična i očekivana pojava.

Kad se sve navedeno uzme u obzir, h -indeks u osnovi definira prepoznatljivost, odnosno konzistentnost pojedinog znanstvenika, odnosno časopisa, u određenom području. U ovom slučaju prepoznatljivost znači da znanstvenik ima veći broj radova, pri čemu su svi dobili relativno veliki broj citata i to tzv. neovisnih citata. Neovisni citati su citati koje autor dobiva od njemu nepoznatih kolega izvan svoje institucije, a u slučaju malih zemlja, izvan svoje zemlje.

Kao i za ostale indikatore vrednovanja znanstvenog rada, tako je i za h -indeks pri interpretaciji vrijednosti važno voditi računa, ne samo o disciplini ili području (5), nego i granama, kao i o aktualnosti problematike kojom se znanstvenik bavi. Hirsch (1) na temelju svojih izračuna, predlaže kao orijentir za vrednovanje fizičara vodećih svjetskih istraživačkih sveučilišta za napredovanje u izvanrednog profesora $h \sim 12$, za redovitog profesora $h \sim 18$, a članstvo u Nacionalnoj akademiji za znanost Sjedinjenih Američkih Država (engl. *National Academy of Science of the United States of America*), prosječan $h \sim 45$, osim nekih iznimaka. Predlaže da kao mjerilo za uspješnog znanstvenika fizičara, s 20 godina istraživačkog rada, h -indeks bude 20, dok h -indeks od 40 indicira „izvanrednog znanstvenika u izuzetno uspješnom laboratoriju“. Navodi i primjere fizičara nobelovaca, čije se vrijednosti h -indeksa kreću od 70 to 90. Prosječan h -indeks fizičara kandidata za Nobelovu nagradu u dvadesetogodišnjem razdoblju od 1985.-2005. bio je 35.

Prema Hirschu, najcitiranijih 10 znanstvenika iz područja bioznanosti, u razdoblju od 1983.-2002., imalo je medijan h -indeksa 57, što je znatno više nego za fizičare. No, bioznanosti su preširoko područje da bi se olako uspoređivao h -indeks molekularnog biologa i biologa koji se bavi ekologijom, ili biorazličitošću, odnosno floristikom ili zoologijom.

Cronin i Meho (6) proveli su istraživanje usporedbe h -indeksa i ukupnog broja citata za područje informacijskih znanosti. Analizirali su 31-og znanstvenika s najviše citata s Fakulteta informacijskih znanosti u SAD, u razdoblju od 1999.-2005. godine, prema *SSCI most cited IS scholars*. Raspon vrijednosti njihovih h -indeksa bio je od 5 do 20, uz činjenicu da su isključili samocitate. Dokazali su da postoji pozitivna korelacija između h -indeksa i broja citata, čime se sugerira da je ukupan broj citata zaista pouzdan indikator odjeka i utjecaja radova pojedinih znanstvenika. Srednja vrijednost h -indeksa za informacijske znanosti iznosila je 11. Oppenheim (7) je analizirao britanske znanstvenike iz područja knjižničnih i informacijskih znanosti i dobio srednju vrijednost h -indeksa 7.

Jokić i Šuljok (8) objavile su rezultate istraživanja h -indeksa doktora znanosti iz prirodnih i društvenih znanosti u

important to investigate the impact of number of authors on total citation count. These authors have proven that the higher the number of authors, the higher the number of self-citations which, unless they are excluded from the final count, may inflate the h -index. On the other hand, it is important to notice, that for some scientific fields, and especially those in development, self citation is a logical and expected phenomenon.

When we take previously said into consideration, h -index mainly defines the recognisability and consistency of an individual scientist or journal in a specific discipline. In this case, "recognisability" means that a recognisable scientist has a relatively higher number of papers of which every one has a significant independent citation count. Independent citations are citations author receives from unknown colleagues from other institutions or, in case of small countries, from other countries.

Considering h -index, similarly to other indicators for evaluating scientific work, it is important to be aware of the scientific discipline as well, as its branches and the topicality of the work (5). For the promotion of physicists on leading research universities, Hirsch (1), as guidelines based on his calculations, suggests $h \sim 12$ for an associate professor, $h \sim 18$ for a full professor, and an average $h \sim 45$ for a membership in the National Academy of Sciences of the United States of America, although he allows for exceptions. He suggests, that h -index for a successful physicist with 20 years of scientific activity, should be 20, while h -index of 40 indicates an "outstanding scientist in a major top research laboratory". He is giving an example of physicists, Nobel Prize winners, whose h -indices range from 70 to 90. An average h -index for a physicist, a candidate for Nobel Prize in a 20 years period from 1985-2005 was 35.

According to Hirsch, the most quoted top 10 scientists in the life sciences in the period 1983-2002, had a median h -index of 57, what is significantly higher than for physicists. However, the life sciences are too large area where we cannot lightly compare h -index of molecular biologists and environmental or biodiversity biologists, or biologists in the fields of floristics or zoology.

Cronin & Meho (6) compared h -indices and the total citation counts for the information sciences. They analysed 31 most cited authors, according to *SSCI most cited IS scholars*, from faculties of information sciences in the US in the period 1999-2005. After excluding the self citations, the h -index values where ranged from 5-20. They have proven that there is a positive correlation between the citation count and h -index, which suggest that total citation count is a reliable indicator of impact and influence of individual scientists' works. The average h -index for information sciences was 11. Oppenheim (7) has analysed British scientists in library and information science discipline and got an average h -index of 7.

Hrvatskoj, u razdoblju od 1996.-2005. godine. Za društvene znanosti kao cjelinu, dobivene vrijednosti *h*-indeksa kretale su se u rasponu od 1 do 6, s tim da je 57,9% citiranih autora imalo *h*-indeks 1, dok je *h*-indeks 4 ili veći od 4 imalo je svega 9% autora. Vrijednosti *h*-indeksa za znanstvenike iz prirodnih znanosti kao cjeline, bile su u rasponu od 1 do 20 i mogle su se pripisati 94,6% autora. Fizičari su imali *h*-indeks veći od 4 u 61%, kemičari u 56,3%, biolozi u 41,3% te matematičari u 19,8% slučajeva. Važno je napomenuti da samocitirati u ovom istraživanju nisu isključivani iz tehničkih razloga.

Za potrebe ovog prikaza, istražene su vrijednosti *h*-indeksa radova objavljenih u citatnoj bazi WoS (engl. *Web of Science*) za područje kliničke kemije (*Clinical chemistry* i *Medicinal chemistry*) u razdoblju od 1995. do 2008. godine. Ukupno je bilo objavljeno 8.675 radova. Prvih 10 najproduktivnijih autora, čiji se broj radova kretao od 21 do 33, imalo je raspon *h*-indeksa od 8 do 19. Za usporedbu, da bi ilustrirali koliko je važno razdoblje koje se istražuje, kao i aktualnost znanstvene problematike, analizirali smo radove iz iste citatne baze u razdoblju od 1985.-2005. godine. U tom razdoblju, 10 najproduktivnijih autora iz istog područja imalo je raspon radova od 20-27, a *h*-indekse od 6-17, bez provjere i isključivanja samocitata. Zanimale su nas i vrijednosti *h*-indeksa hrvatskih znanstvenika i stručnjaka iz ovog područja. Najproduktivniji autor, uključujući i 2008. godinu, imao je 21 rad u WoS-u, a *h*-indeks 3. Važno je spomenuti da je gotovo 50% radova objavljeno u 2007. i 2008. godini, koji nisu niti mogli dobiti veći broj citata, čime je trenutna vrijednost *h*-indeksa relativno niska, u usporedbi s navedenim najproduktivnijim znanstvenicima i stručnjacima iz navedenih područja.

Znanstvena zajednica je pokazala veliki interes za *h*-indeks kao scientometrijski indikator, tako da citatane baze Scopus, a ubrzo i WoS, uz broj radova, broj citata, prosječan broj citata nude i automatski izračun *h*-indeks uključujući sve vrste citata.

Osim za autore, *h*-indeks se počeo sve više koristiti i kao indikator za vrednovanje časopisa. Braun i suradnici (2) su napravili usporedbu određenih časopisa prema IF i *h*-indeksu. Rezultati su pokazali da su časopisi *Physical Review Letters*, *Astrophysical Journal* i *Journal of the American Chemical Society* među prvih 20 po iznosu *h*-indeksa, dok prema IF, ta tri časopisa nisu bila među prvih 100 časopisa, što pokazuje nepotpunu korelaciju ova dva indikatora.

Za potrebe ovog teksta istražila sam najprestižnije časopise, tj. časopise s najvišim IF, koji se bave problematikom kliničke kemije, indeksirane u WoS-u u razdoblju 1995. do 2007. godine. Pokazalo se da je časopis *Clinical Chemistry* s IF 4.803 za 2007. godinu imao *h*-indeks 80; *Current Medicinal Chemistry* s IF 4.944 imao je *h*-indeks 33; *Journal of Medicinal Chemistry* s IF 4.895 imao je *h*-indeks 30; *Clinical Chemistry and Laboratory Medicine* s IF 2.618 imao je *h*-indeks 16; *Clinica Chimica Acta* s IF 2.601 imao je *h*-in-

Jokić and Šuljok (8) analysed *h*-indices of PhDs in natural and social sciences in Croatia for the period 1996-2005. For social sciences as a whole, *h*-indices were in range 1-6, where the 57.9% of authors had an *h*-index of 1, while only 9% of the authors had an *h*-index of 4 or greater. 94.6% *h*-index values for scientists in natural sciences were in range 1-20. The physicists had *h*-index $4 \geq$ in 61% of the cases, chemists in 56.3%, and biologists in 41.3% and mathematicians in 19.8% of the cases. Out of technical reasons, this study did not differentiate independent citations from self citations.

For this short overview we have analysed the *h*-index values of papers covered by WoS (*Web of Science*) for *Clinical chemistry* and *Medicinal chemistry* from 1995 to 2008. The total number of papers was 8,675. The top ten most productive authors had the number of papers in range from 21 to 33 and an *h*-index in range from 8 to 19. For comparison and illustration of importance of the time span being analysed, we also analysed the papers covered by WoS from 1985 to 2005. In this period the top ten authors had the number of papers in range from 20 to 27, and *h*-indices in range from 6 to 17. The study did not exclude self-citations. We were also interested in *h*-indices of Croatian scientists and experts in this field. The most productive author in the period 1995-2008 had 21 papers covered by WoS and his *h*-index was 3. We should mention that almost 50% of papers were published in 2007 and 2008 and were not in the position to get a significant number of citations. This makes the current *h*-index value relatively low in comparison to the most productive scientists and experts in the researched fields in this period.

The scientific community has shown much interest for *h*-index as a scientometric indicator. A Scopus citation index was the first to offer automatic *h*-index calculation in addition to already present paper count, citation count and average citations. WoS soon followed.

Except for authors, *h*-index is used more and more as an indicator for the evaluation of journals. Braun *et al.* (2) compared certain journals by both *h*-index and IF. The results demonstrated that the journals *Physical Review Letters*, *Astrophysical Journal* and *Journal of the American Chemical Society* were in the top 20 by *h*-index. These three journals ranked outside the top 100 by impact factor, which demonstrates incomplete correlation of these two indicators.

For this text, we have researched the most prestigious journals (with the highest IF) in the *clinical chemistry* and *medicinal chemistry* subject fields, indexed by WoS for the time span 1995.-2007. The results were these: *Clinical Chemistry* with IF 4.803 for 2007 had an *h*-index 80, *Current Medicinal Chemistry* had IF 4.944 and *h*-index 33, *Journal of Medicinal Chemistry* had IF 4.895 and *h*-index 30, *Clinical Chemistry and Laboratory Medicine* had IF

deks 15. Samo na osnovi navedenih podataka može se zaključiti da su podatci o *h*-indeksu časopisa dopuna IF i doprinose cjelovitijoj slici prepoznatljivosti časopisa za određeno područje.

Na temelju svih iznesenih činjenica, *h*-indeks je zasigurno jedan od indikatora koji doprinose ukupnoj procjeni znanstvenog rada pojedinog znanstvenika, institucije, područja, časopisa itd. Ne bi bilo dobro promatrati ga zasebno, odnosno neovisno o predmetnom području, dužini radnog vijeka znanstvenika, znanstvenoj produktivnosti, koautorstvu, ukupnom broju citata i vrsti citata i ostalim relevantnim parametrima.

2.618 and *h*-index 16, *Clinica Chimica Acta* had IF 2.601 and *h*-index 15. Just by analysing this data we can see that *h*-index value of a journal, when used alongside IF, can help in forming a clearer picture of a journal's status within a specific discipline.

Taking into account already mentioned facts; *h*-index is one of the indicators which can help in evaluating scientific work of an individual scientist, institution, discipline, and journal. However, it should not be considered independently of the scientific discipline, the length of a scientist's work experience, scientific productivity, co-authorships, the citation count, the types of citations and other relevant parameters.

Adresa za dopisivanje:

Maja Jokić
Nacionalna i sveučilišna knjižnica u Zagrebu
Ulica Hrvatske bratske zajednice b.b.
10 000 Zagreb
e-pošta: mjokic@nsk.hr

Corresponding author:

Maja Jokić
National and University Library in Zagreb
Ulica Hrvatske bratske zajednice b.b.
10000 Zagreb
Croatia
e-mail: mjokic@nsk.hr

Literatura/References

1. Hirsch JE. An indeks to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005;102:16569-72.
2. Braun T, Glanzel W, Schubert A. A Hirsch-type indeks for journals. *Scientist* 2005;19:8.
3. Batista PD, Campiteli MG, Kinouchi O. Is it possible to compare researchers with different scientific interests? *Scientometric*. 2006;68:179-89.
4. van Raan AFJ. Comparison of the hirsch-indeks with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*. 2006;67:491-502.
5. Egghe L. Dynamic indeks: The hirsch indeks in function of time. *JASIST*. 2007;58:452-4.
6. Cronin B, Meho, L. Using the *h*-indeks to rank influential information scientists. *JASIS*. 2006;57:1275-8.
7. Oppenheim C. Using the *h*-indeks to rank influential british researchers in information science and librarianship. *JASIST*. 2007;58:197-301.
8. Jokić M, Suljok A. [Produktivnost i njezin odjek prema citatnim bazama ISI-ja i Scopusa za razdoblje 1996-20005]. In: [Onkraj mitova o prirodnim i društvenim znanostima: sociološki pogled]. Ed. Prpic K, Zagreb: Institut za društvena istraživanja u Zagrebu; 2008, pp. 133-58. (in Croatian)