

УДК 001.92

В. В. Сапожников, д-р техн. наук
Вл. В. Сапожников, д-р техн. наук

Кафедра «Автоматика и телемеханика на железных дорогах»
Петербургский государственный университет путей сообщения
Императора Александра I, Санкт-Петербург

Д. В. Ефанов, д-р техн. наук

ООО «ЛокоТех-Сигнал»
Кафедра «Автоматика, телемеханика и связь на железнодорожном транспорте»
Российский университет транспорта, Москва

НОВЫЙ ПОДХОД К РАСЧЕТУ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ УЧЕНЫХ

Предложен новый индекс оценки работы ученого, максимально полно учитывающий количество цитирований всех его работ. Новый показатель деятельности ученых, так же как индекс Хирша, выражается в виде единственного числа, имеет простой алгоритм подсчета и понятную геометрическую интерпретацию. Приведены примеры подсчета нового показателя, обозначены некоторые его особенности.

оценка деятельности ученого, наукометрический показатель, индекс цитирования, индекс Хирша.

DOI: 10.20295/2412-9186-2019-5-4-505-514

Введение

При управлении научно-исследовательской деятельностью во всем мире принято использовать различные системы показателей, ориентированные так или иначе на количественную оценку вклада того или иного ученого в определенную область науки. Такой подход стал возможен благодаря развитию вычислительной техники и компьютерных технологий, совершенствованию интернета и созданию электронных баз данных, учитывающих результаты научной деятельности. Показатели оценки стараются использовать на всех этапах процесса управления научно-исследовательской деятельностью – как на этапах организации и планирования, так и на этапах стимулирования ученых и контроля их работы [1].

В Российской Федерации при оценке деятельности ученых широко применяют данные из таких наукометрических баз, как Web of Science (USA), Scopus (EU) и РИНЦ (РФ). При этом в качестве наиболее важных показателей в каждой из приведенных выше баз данных выбираются одни и те же, а именно: число опубликованных работ (q), общее число цитирований всех работ (N), индексы цитирования и важности публикаций. К по-

следним относятся различные вычисляемые показатели, наиболее известным из которых является индекс Хирша (h -индекс), введенный в рассмотрение в 2005 году [2]. Ученый имеет h -индекс, равный h , в том случае, если h его работ процитированы не менее чем h раз, а остальные $q-h$ работ цитируются не более чем h раз. Такой показатель имеет свои достоинства и свои недостатки, что отмечено в оригинальной статье Х. Хирша. Например, h -индекс учитывает число наиболее важных работ ученого, но никак не учитывает число цитирований за пределом величины h : если работа из числа публикаций, по которым определен h -индекс, цитируется большее количество раз или любое цитирование $q-h$ публикаций с меньшим, чем h , числом цитирований. Несмотря на это h -индекс в настоящее время стал привычным и наиболее часто используется при оценке результатов научной деятельности.

Известно большое число модификаций индекса Хирша и других наукометрических показателей, ориентированных на получение более точной оценки качественных и количественных составляющих работы ученых, уточняющих или модернизирующих ранее использовавшиеся показатели. Большинство из них в той или иной мере копируют идею Х. Хирша или направлены на уточнение относительных показателей деятельности ученого [3, 4].

Данная статья посвящена описанию нового наукометрического показателя, который позволяет комплексно учитывать все цитирования всех публикаций ученого из имеющихся в базе данных. При этом новый показатель, обозначенный как w -индекс, так же как и h -индекс, является одним числом, однако более полно характеризующим вклад автора в развитие науки в соответствующей предметной области.

1. Новый наукометрический показатель

Индекс Хирша получил широкое распространение, по всей видимости, благодаря простоте вычисления и доступности его интерпретации. На рис. 1 изображено графическое представление h -индекса: на оси абсцисс расположены публикации, ранжированные по числу цитирований в невозрастающем порядке (q_i), а на оси ординат – число цитирований каждой публикации (N_i) [2]. Из геометрической интерпретации следует, что из N цитирований всех p работ некоторого автора индекс Хирша учитывает только h^2 цитирований первых h работ. На рис. 1 эти работы и их цитирования образуют квадрат Хирша с длиной стороны, равной h .

Графическая форма h -индекса наглядно показывает его недостатки – не учитывается $N-h^2$ цитирований всех работ, цитирования которых попадают в две области: G (над квадратом Хирша) и P (справа от квадрата Хирша).

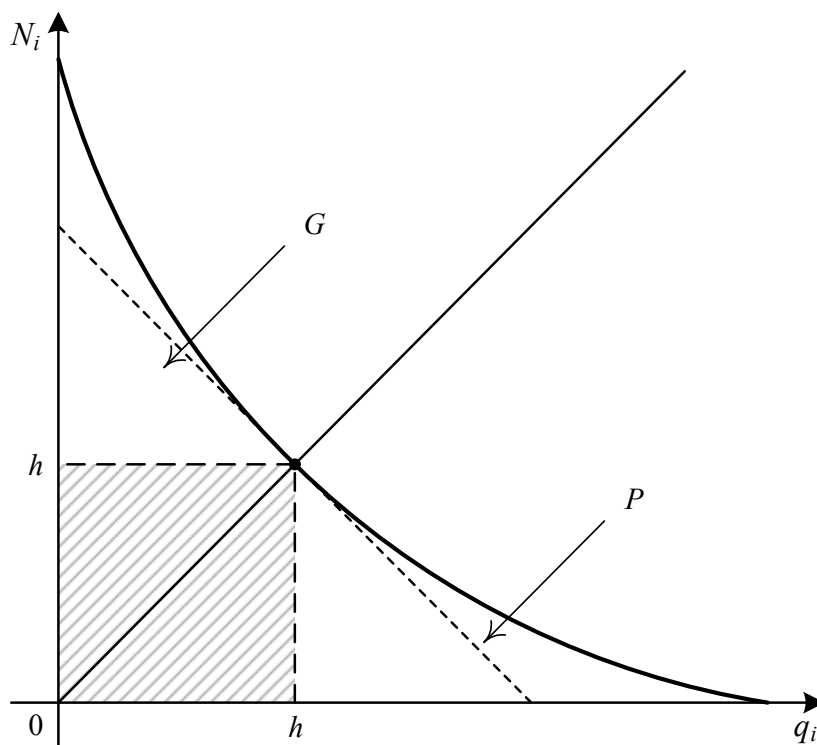


Рис. 1. Геометрическое представление h -индекса

Какой-либо иной индекс, отражающий деятельность некоторого автора, по нашему мнению, должен отвечать следующим критериям:

- 1) правильно отражать значимость всех работ;
- 2) учитывать всё множество цитируемых работ;
- 3) иметь понятный физический смысл;
- 4) быть простым, т. е. представлять собой некоторое число, характеризующее работу автора.

Индекс w , отвечающий представленным критериям, может быть получен по следующему алгоритму.

1. Производится ранжирование работ автора в порядке невозрастания числа их цитирований.

2. Полагается $i = 1$, $w = 0$.

3. Формируется множество работ ученого.

4. По публикациям ученого из имеющегося множества определяется h -индекс и выделяется $h_i = h$ первых в списке работ с числом цитирований, равным h .

5. Подсчитывается общее число цитирований h_i первых работ $S_i = H_i + G_i$ (здесь H_i и G_i – число цитирований h_i работ, попадающих в i -й квадрат Хирша, и число цитирований работ, попадающих в область над i -м квадратом Хирша):

$$S_i = \sum_{t=1}^{h_i} N_t, \quad (1)$$

где t – каждая конкретная работа из рассматриваемых.

6. Полученное число делится на количество цитируемых работ:

$$w_i = \frac{S_i}{h_i}. \quad (2)$$

7. Определяется число:

$$w = w + w_i. \quad (3)$$

8. Первые h_i работ автора исключаются из множества его работ, и формируется множество оставшихся работ.

9. Полагается $i = i + 1$.

10. Проверяется условие $h_i = 0$? Если да, то подсчет числа w окончен, если нет, то осуществляется переход к выполнению п. 4.

11. Конец алгоритма.

Процедура вычисления w -индекса может быть выражена формулой:

$$w = \sum_{i=1}^k \frac{h_i^2 + G_i}{h_i} = \sum_{i=1}^k \frac{H_i + G_i}{\sqrt{H_i}} = \sum_{i=1}^k w_i. \quad (4)$$

В формуле (4) число k определяет количество итераций подсчета суммы w до тех пор, пока $h_i \neq 0$.

Алгоритм подсчета w -индекса может быть проиллюстрирован геометрической интерпретацией (рис. 2). По оси абсцисс отложены ранжированные по числу цитирований (N_i) все публикации автора (q_i). На каждом этапе итерации при работе алгоритма подсчета w -индекса осуществляется усреднение числа цитирований первых h_i публикаций.

Число H_1 представляет собой h_1 -индекс автора, полученный при рассмотрении первой итерации алгоритма. Число H_2 представляет собой h_2 -индекс автора, полученный во второй итерации алгоритма, и т. д.

2. Вычисление нового наукометрического показателя и его особенности

Приведем пример подсчета w -индекса. Предположим, что статистические показатели некоторого ученого задаются таблицей 1.

На первом этапе определяется $h_1 = h$ первых работ из полного множества работ ученого, по которым подсчитывается h -индекс. Из данных таблицы 1 это первые восемь его работ. Определим число цитирований h_1 работ:

$$S_1 = 28 + 25 + 23 + 17 + 16 + 16 + 10 + 8 = 143.$$

По формуле (2) определяем число w_1 :

$$w_1 = \frac{S_1}{h_1} = \frac{143}{8} = 17,875.$$

Удаляем из множества всех работ $h_1 = 8$ первых по цитированию и образуем таблицу 2.

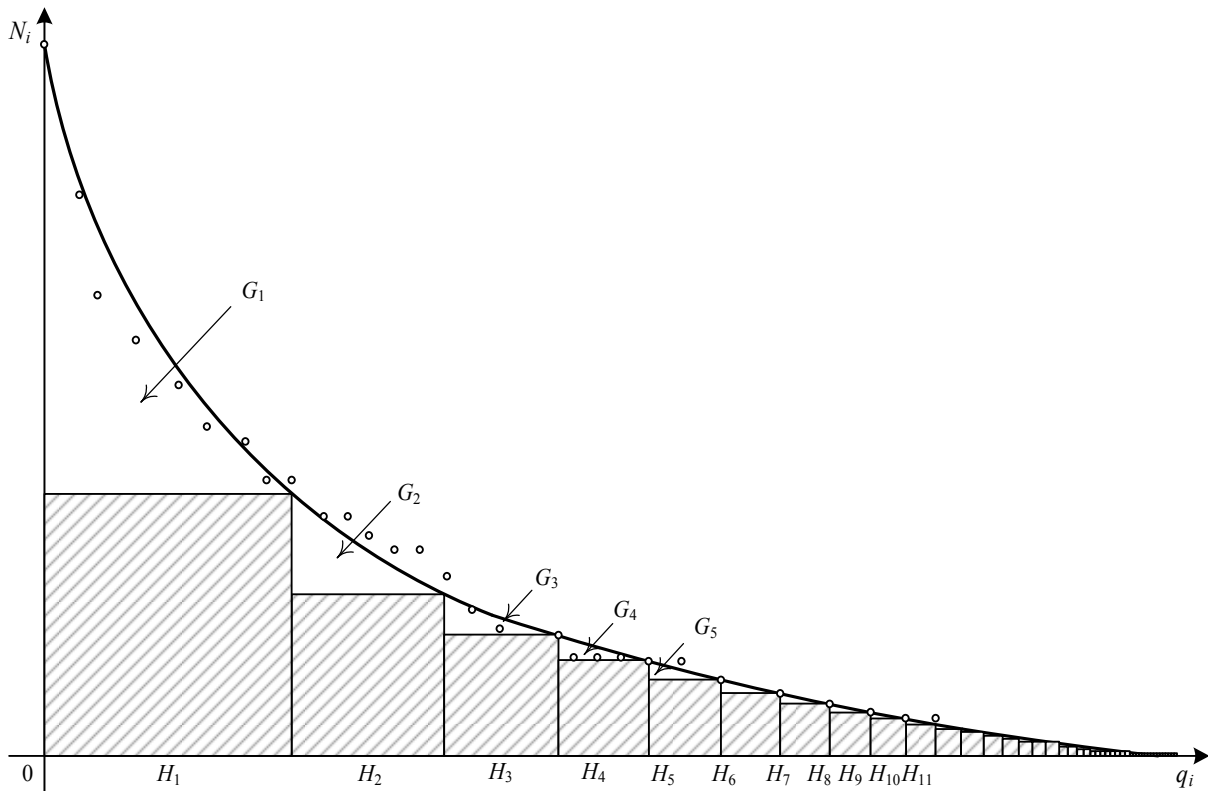


Рис. 2. Геометрическое представление w -индекса

Таблица 1. Исходные данные к расчету

Номера работ в ранжированном списке														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Число цитирований каждой работы														
28	25	23	17	16	16	10	8	8	8	7	7	7	7	6
Номера работ в ранжированном списке														
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Число цитирований каждой работы														
5	5	4	3	3	3	2	2	2	2	1	1	1	0	0

Определяем $h_2 = h$ первых работ из полученного подмножества работ ученого, по которым подсчитывается h -индекс. Из данных таблицы 2 это первые шесть его работ. Определим число цитирований h_2 работ:

$$S_2 = 8 + 8 + 7 + 7 + 7 + 7 = 44.$$

Таблица 2. Ранжированные работы ученого после первой итерации работы алгоритма

Номера работ в ранжированном списке										
1	2	3	4	5	6	7	8	9	10	11
Число цитирований каждой работы										
8	8	7	7	7	7	6	5	5	4	3
Номера работ в ранжированном списке										
12	13	14	15	16	17	18	19	20	21	22
Число цитирований каждой работы										
3	3	2	2	2	2	1	1	1	0	0

По формуле (2) определяем число w_2 :

$$w_2 = \frac{S_2}{h_2} = \frac{44}{6} = 7,333.$$

Удаляем из рассматриваемого подмножества работ $h_2 = 6$ первых по цитированию и образуем таблицу 3.

Таблица 3. Ранжированные работы ученого после второй итерации работы алгоритма

Номера работ в ранжированном списке															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Число цитирований каждой работы															
6	5	5	4	3	3	3	2	2	2	2	1	1	1	0	0

Вновь определяем те работы, которые образуют квадрат Хирша: $h_3 = h$ первых работ из полученного подмножества работ ученого. Из таблицы 3 это первые четыре его работы. Число цитирований h_3 работ

$$S_3 = 6 + 5 + 5 + 4 = 20,$$

значит

$$w_3 = \frac{S_3}{h_3} = \frac{20}{4} = 5.$$

Удаляем из рассматриваемого подмножества работ $h_3 = 4$ первых по цитированию и образуем очередную таблицу (табл. 4).

Из табл. 4 следует, что $h_4 = 3$, $S_4 = 9$ и $w_4 = 3$.

Дальнейшие шаги работы алгоритма приводят к тому, что таблица становится еще более простой. Далее их приводить не будем, а выпишем ряд получаемых величин:

$$h_5 = 2, S_5 = 4 \text{ и } w_5 = 2;$$

$$h_6 = 2, S_6 = 4 \text{ и } w_6 = 2;$$

$$h_7 = 1, S_7 = 1 \text{ и } w_7 = 1;$$

$$h_8 = 1, S_8 = 1 \text{ и } w_8 = 1;$$

$$h_9 = 1, S_9 = 1 \text{ и } w_9 = 1.$$

Таблица 4. Ранжированные работы ученого после третьей итерации работы алгоритма

Номера работ в ранжированном списке											
1	2	3	4	5	6	7	8	9	10	11	12
Число цитирований каждой работы											
3	3	3	2	2	2	2	1	1	1	0	0

Конечное значение w -индекса определяется как сумма промежуточных (см. формулу (4)):

$$w = \sum_{i=1}^k w_i = 17,875 + 7,333 + 5 + 3 + 2 + 2 + 1 + 1 + 1 = 40,208.$$

Таким образом, w -индекс учел все 227 цитирований ученого, тогда как h -индекс учел только 64 ссылки на его работы.

Из рис. 2 наглядно следует, что h -индекс никак не учитывает область цитирований над квадратом Хирша (область G_1), а также все цитирования $q-h_1$ работ автора.

Приведем здесь два примера, иллюстрирующих возможности учета при вычислении w -индекса числа цитирований из областей G и P .

Первый пример показывает, как w -индекс учитывает число цитирований из области G . Предположим, что некоторый ученый опубликовал за всю научную деятельность 5 работ, каждая из которых процитирована по 100 раз. В этом случае h -индекс этого ученого будет равен 5. При этом учтено будет 25 цитирований из 500. Рассчитаем w -индекс:

$$w = \frac{25 + 475}{5} = 100.$$

Таким образом, w -индекс показывает среднее число цитирований по отношению к числу публикаций. В данном случае этот показатель равен отношению числа цитирований (N) к числу опубликованных работ (q). Од-

нако для авторов с бóльшим числом публикаций, не цитируемых как минимум h раз, этот коэффициент дает более точную оценку.

Приведем второй пример, иллюстрирующий влияние числа цитирований из области P на показатель деятельности ученого. Предположим, что автор опубликовал 20 научных работ, каждая из которых процитирована по 10 раз. В этом случае h -индекс ученого будет равен 10. Будут учтены 100 ссылок на его работы из 200. Рассчитаем w -индекс:

$$w = w_1 + w_2 = \frac{100}{10} + \frac{100}{10} = 20.$$

В данном случае оценка, по нашему мнению, является более точной, так как охватывает область неучтенных при расчете h -индекса работ и их цитирований.

Заключение

Предложенный в настоящей работе w -индекс является простым и понятным и более точно в сравнении с h -индексом отражает «картину» деятельности ученых.

Так же как индекс Хирша, новый индекс сохраняет недостаток, связанный с необходимостью дополнительного учета самоцитируемых работ и не решает проблемы искусственного увеличения показателей («накрутки»). Тем не менее процедура «накрутки» гораздо меньше влияет на значение w -индекса, чем на значение h -индекса. Следует отметить еще один недостаток w -индекса, заключающийся в том, что при его подсчете на каждом этапе итерации осуществляется различное усреднение показателей цитируемости работ. Число цитирований работ из правой области цитируемости (см. рис. 1) может дать существенный вклад в значение w -индекса (являться более весомым, чем число цитирований первых h работ, не учтенных при подсчете индекса Хирша). По-видимому, возможны модификации w -индекса, учитывающие эту особенность. Например, может быть предложен w' -индекс, рассчитываемый не для всех цитируемых работ, а только для тех работ, которые процитированы не менее двух раз (или для работ, процитированных не менее t раз), или же w'' -индекс, рассчитываемый путем ограничения числа шагов итерации по предложенному алгоритму.

Несмотря на недостатки, w -индекс, в отличие от индекса Хирша, позволяет учесть проблему высокой цитируемости авторов с малым количеством работ, а также авторов с большим количеством малоцитируемых работ и, по нашему мнению, более точно отражает вклад ученого в развитие его предметной области наук.

Библиографический список

1. Новиков Д. А. Методология управления. – М. : Либроком, 2011. – 128 с.
2. Hirsch J. E. An Index to Quantify an Individual's Scientific Research Output // Proceedings of National Academy of Sciences of the USA. – 2005. – Vol. 102. – Issue 46. – Pp. 569–572. – doi.org/10.1073/pnas.0507655102.
3. Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами. – 2013. – № 44. – С. 248–261.
4. Штовба С. Д. Обзор наукометрических показателей для оценки публикационной деятельности ученого / С. Д. Штовба, Е. В. Штовба // Управление большими системами. – 2013. – № 44. – С. 262–278.

Valery V. Sapozhnikov
Vladimir V. Sapozhnikov

«Automation and Remote Control on Railways» department
Emperor Alexander I St. Petersburg State Transport University, St. Petersburg

Dmitry V. Efanov
“LocoTech-Signal” LLC,
«Automation, Remote Control and Communication on Railway Transport»
Russian University of Transport, Moscow

The new approach for the scientists research index

The article proposes a new index for evaluating the scientist work, which fully takes into account the number of citations of all his works. Like the h-index a new index expressed as a single number. The index has a simple calculation algorithm and a clear geometric interpretation. Examples of calculating a new indicator are given and some features are indicated.

scientist performance assessment; author-level metrics; citation index; H-index.

References

1. Novikov D. A. (2011) Management methodology [Metodologiya upravleniya], Moscow, Librokom, 2011. – 128 p.
2. Hirsch J. E. (2005) An Index to Quantify an Individual's Scientific Research Output // Proceedings of National Academy of Sciences of the USA. – 2005. – Vol. 102. – Issue 46. – Pp. 569–572. – doi.org/10.1073/pnas.0507655102.
3. Tsyganov A. V. (2013) Brief review of main scientometric indices based on citations [Kratkoe opisanie naukometriceskih pokazatelej, osnovannyh na citiruemosti], Large-Scale Systems Control [Upravlenie bol'shimi sistemami], 2013, Issue 44. – Pp. 248–261.
4. Shtovba S. D., Shtovba E. V. (2013) A survey on scientometric indicators for assessment of researcher's publication activity [Obzor naukometriceskih pokazatelej dlya ocenki publi-kacionnoj deyatel'nosti uchenogo], Large-Scale Systems Control [Upravlenie bol'shimi sistemami], 2013, Issue 44. – Pp. 262–278.

*Статья представлена к публикации членом редколлегии Х. Христовым.
Поступила в редакцию 02.09.2019, принята к публикации 17.09.2019.*

САПОЖНИКОВ Валерий Владимирович – д-р техн. наук, профессор кафедры «Автоматика и телемеханика на железных дорогах» Петербургского государственного университета путей сообщения Императора Александра I.
e-mail: port.at.pgups1@gmail.com

САПОЖНИКОВ Владимир Владимирович – д-р техн. наук, профессор кафедры «Автоматика и телемеханика на железных дорогах» Петербургского государственного университета путей сообщения Императора Александра I.
e-mail: sapozhnikov-at@yandex.ru

ЕФАНОВ Дмитрий Викторович – д-р техн. наук, доцент, руководитель направления систем мониторинга и диагностики ООО «ЛокоТех-Сигнал», профессор кафедры «Автоматика, телемеханика и связь на железнодорожном транспорте» Российского университета транспорта.
e-mail: TrES-4b@yandex.ru

© Сапожников Вал. В., Сапожников В. В., Ефанов Д. В., 2019