

ВАЛЕРИЙ ВЛАДИМИРОВИЧ САПОЖНИКОВ

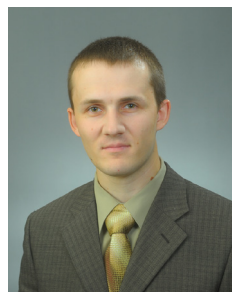
доктор технических наук, профессор,
профессор Петербургского государственного университета
путей сообщения Императора Александра I,
С.-Петербург, Россия;
e-mail: port.at.pgups@gmail.com

**ВЛАДИМИР ВЛАДИМИРОВИЧ САПОЖНИКОВ**

доктор технических наук, профессор,
профессор Петербургского государственного университета
путей сообщения Императора Александра I,
С.-Петербург, Россия;
e-mail: at.pgups@gmail.com

**ДМИТРИЙ ВИКТОРОВИЧ ЕФАНОВ**

доктор технических наук, доцент,
профессор Российского университета транспорта
Москва, Россия;
e-mail: TrES-4b@yandex.ru



«Настраиваемый» индекс оценки деятельности ученого

УДК: 001.92, 001.891

DOI: 10.24411/2079-0910-2020-12011

Предложен новый наукометрический показатель, позволяющий более точно оценивать вклад того или иного ученого в соответствующую предметную область науки, чем широко распространенный индекс Хирша. Новый индекс, так же как и индекс Хирша, представляет собой одно число, просто вычисляется и имеет понятную геометрическую интерпретацию. Приводится алгоритм определения нового наукометрического показателя. Отмечаются его особенности, достоинства и недостатки. Приведены некоторые примеры его подсчета.

Ключевые слова: наукометрические показатели, оценка деятельности ученого, индекс цитирования, индекс Хирша, w-индекс.

Введение

При организации и планировании научно-исследовательской деятельности в последние годы стали широко применяться разнообразные системы показателей, позволяющие дать количественную оценку вкладу ученого в развитие соответствующей предметной области науки [Новиков, 2011]. Такая возможность возникла благодаря колоссальному скачку в области информационных и компьютерных технологий, когда каждая весомая работа попадает в какие-либо наукометрические базы данных, доступные в той или иной форме из любой точки мира, из которой есть вход в интернет.

Существуют разнообразные базы, которые имеют собственные программные средства вычисления различных наукометрических показателей, а также по-разному ценятся в процессе анализа творческой деятельности ученых. Так, например, в Российской Федерации наиболее востребованными считаются база данных российского индекса научного цитирования (РИНЦ), а также две международные наукометрические базы *Web of Science*, или *WoS* (USA), и *Scopus* (EU). При этом вес последних двух, как правило, оценивается выше веса первой, а публикация в издании, индексируемом в *WoS* или *Scopus*, ценится научным сообществом выше публикации в базе данных РИНЦ.

Состав показателей, рассчитываемых всеми базами данных, является примерно одинаковым. Одними из ключевых являются такие показатели, как общее число опубликованных работ (q), общее число цитирований всех работ (N), число цитирований каждой работы (N_j), а также вычисляемые на их основе общие индексы цитирования и важности публикаций. Среди таких индексов наиболее популярным является h -индекс, предложенный Х. Хиршем в 2005 г. и названный впоследствии его именем (индекс Хирша) [Hirsch, 2005]. 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.

Индекс Хирша имеет свои достоинства и свои недостатки, что отмечено не только в последующих публикациях ученых всего мира, но и в оригинальной статье автора данного показателя. Например, h -индекс учитывает число наиболее важных работ ученого, но никак не учитывает число цитирований за пределом величины h : любые цитирования свыше величины h тех публикаций, по которым определен h -индекс, или же любое цитирование $q - h$ публикаций с меньшим, чем h , числом цитирований. Несмотря на это, h -индекс в настоящее время «прижился» и является одним из ключевых вспомогательных показателей в оценке деятельности ученых.

Известно большое число модификаций индекса Хирша и других наукометрических показателей, ориентированных на получение более точной оценки качественных и количественных составляющих работы ученых. Большинство из них в той или иной мере копирует идею Х. Хирша или направлено на уточнение относительных показателей ученого [Попов, 2010; Цыганов, 2013; Штовба, 2013; Полянин, 2014; Михайлов, 2014а, 2014б; Марвин, 2015; Raheel et al., 2018]. В данной статье авторы предлагают читателю ознакомиться с результатами исследований в области оценки научно-исследовательской деятельности ученых и с предложенным новым индексом цитируемости, обозначенным как w -индекс. Новый показатель, так же как и индекс Хирша, достаточно просто вычисляется, представляет собой одно число и имеет простую геометрическую интерпретацию, однако он более точно учитывает

работы ученого, не попадающие в число h^2 цитирований, учитываемых при определении индекса Хирша.

Новый индекс цитируемости

При определении индекса Хирша производится весьма простая операция: работы q_i некоторого ученого ранжируются в порядке невозрастания показателя числа цитируемости каждой из них N_i , а затем находится такое число h , для которого число цитирований каждой из h работ больше, либо равно h , а $h + 1$ работа и остальные работы не учитываются, даже если некоторые из них также имеют по h цитирований. Геометрически эта процедура иллюстрируется рис. 1 [Hirsch, 2005]. Требуется отыскать так называемый квадрат Хирша H , включающий в себя h^2 цитирований первых h работ ученого в ранжированном списке.

Из рис. 1 следует, что h -индекс никак не учитывает $N - h^2$ цитирований всех работ ученого, которые попадают в две крупные области: над квадратом Хирша (область G) и справа от него (область P), что является весомым недостатком.

Интересным для анализа мог бы быть такой показатель, который в какой-то степени учитывает все цитирования ученого. Такой показатель должен правильно отражать значимость работ ученого, иметь понятный геометрический смысл и являться единственным числом, что было бы удобно для анализа. Индекс Хирша получил широкую известность именно по последним двум своим особенностям.

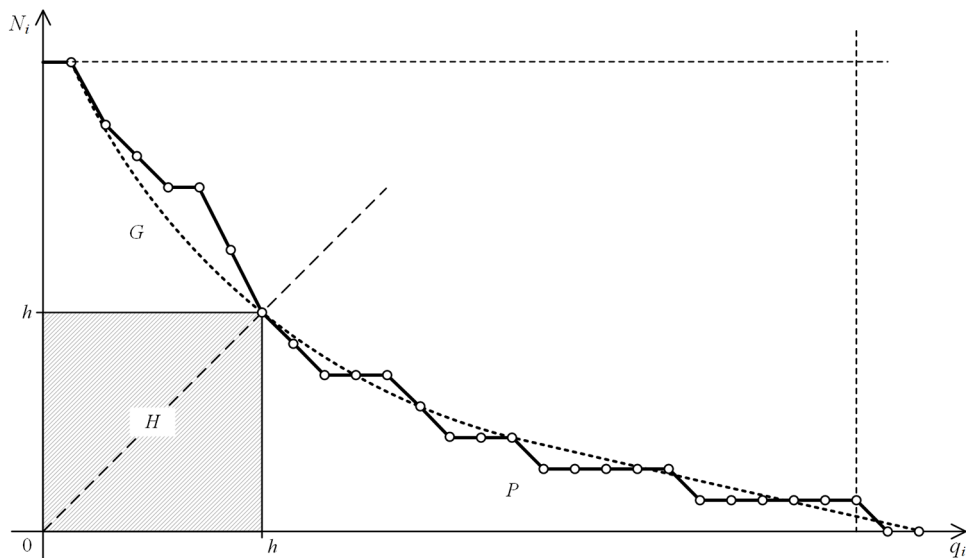


Рис. 1. Геометрическое представление h -индекса

Идея нового индекса заключается в последовательном получении *цепи квадратов Хирша* (далее — просто *цепи Хирша*) и учете среднего числа цитирований тех работ, которые участвуют в его формировании (рис. 2). Цепь Хирша образуется так. Сначала выбираются первые $h_1 = h$ работ, по которым подсчитывается h -индекс, для них определяется среднее число цитирований:

$$(1) \quad w^i = \frac{S_i}{h_i},$$

$$\text{где } S_i = H_i + G_i = \sum_{j=1}^{h_i} N_j,$$

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;
 j — номер работы из рассматриваемых.

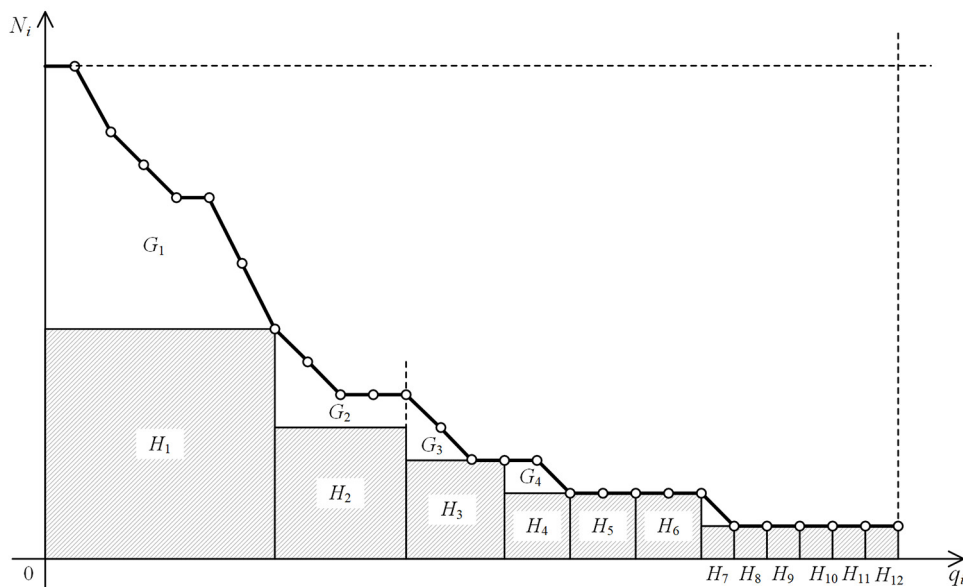


Рис. 2. Геометрическое представление w -индекса

Затем работы, образующие первый квадрат Хирша, исключаются из множества рассматриваемых работ, и процедура повторяется вновь для оставшихся работ, и т. д. до тех пор, пока множество работ автора не станет пустым. Затем определяется сумма значений усредненных чисел цитирования по всем полученным квадратам Хирша:

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

В формуле (2) число k определяет количество итераций подсчета суммы w до тех пор, пока $h_i \neq 0$. Число k также может быть названо *длиной цепи Хирша*, так как определяет, какое число квадратов Хирша должно быть построено при вычислении w -индекса.

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

Примеры подсчета индекса цитируемости

Приведем пример определения w -индекса.

Предположим, что показатели числа работ и их цитируемости некоторого ученого задаются табл. 1.

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	16
Число цитирований каждой работы															
46	38	30	22	10	10	9	9	8	8	8	7	7	6	6	5
Номера работ в ранжированном списке															
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Число цитирований каждой работы															
5	5	4	4	4	3	3	2	2	2	1	1	1	0	0	0

Определим первые $h_1 = h$ работ, образующих первый квадрат Хирша и определяющих, соответственно, индекс Хирша ученого. Из табл. 1 следует, что число $h_1 = 8$. Пользуясь формулой (1), определяем усредненное количество цитирований первых восьми работ ученого:

$$w^1 = \frac{S_1}{h_1} = \frac{46 + 38 + 30 + 22 + 10 + 10 + 9 + 9}{8} = \frac{174}{8} = 21,75.$$

Первые восемь работ исключаются из рассматриваемого множества, и образуется табл. 2.

Таблица 2. Ранжированные работы ученого после удаления $h_1 = 8$ первых работ

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

Процедура вычислений повторяется. Определяем число $h_2 = h$ первых работ из полученного множества работ ученого, по которым подсчитывается новый h -индекс (формируется второй квадрат Хирша). Из данных табл. 2 следует, что это первые шесть работ ученого из усеченного списка. Далее определяем усредненное значение числа цитирований между этими шестью работами:

$$w^2 = \frac{S_2}{h_2} = \frac{8 + 8 + 8 + 7 + 7 + 6}{6} = \frac{44}{6} = 7,333.$$

Первые шесть работ из множества исключаются, и процедура подсчета повторяется. Образуется ряд чисел:

$$w^3 = \frac{S_3}{h_3} = \frac{6+5+5+5}{4} = \frac{21}{4} = 5,25; \quad w^4 = \frac{S_4}{h_4} = \frac{4+4+4}{3} = \frac{12}{3} = 4; \quad w^5 = \frac{S_5}{h_5} = \frac{3+3}{2} = \frac{6}{2} = 3;$$

$$w^6 = \frac{S_6}{h_6} = \frac{2+2}{2} = \frac{4}{2} = 2; \quad w^7 = \frac{S_7}{h_7} = \frac{2}{2} = 2; \quad w^8 = \frac{S_8}{h_8} = \frac{1}{1} = 1; \quad w^9 = \frac{S_9}{h_9} = \frac{1}{1} = 1; \quad w^{10} = \frac{S_{10}}{h_{10}} = \frac{1}{1} = 1.$$

w -индекс ученого определяется как сумма полученных усредненных чисел цитирования по каждому из квадратов Хирша:

$$w = \sum_{i=1}^{10} w^i = 21,75 + 7,333 + 5,25 + 4 + 3 + 2 + 2 + 1 + 1 + 1 = 48,333.$$

Далее приведем три примера подсчета w -индекса, демонстрирующих его особенности.

Рассмотрим случай, когда ученый опубликовал за всю научную деятельность $q = 10$ работ, каждая из которых процитирована по 15 раз (рис. 3). Индекс Хирша этого ученого определяется величиной $h = 10$. Таким образом, h -индекс учитывает $10^2 = 100$ цитирований из 150 цитирований всех работ этого ученого. Использование w -индекса позволяет учесть все цитирования работ:

$$w = \frac{100 + 50}{10} = 15.$$

Следующий пример показывает особенности цитирования $q - h$ работ, не учитываемых при вычислении h -индекса. Пусть ученый имеет 10 научных работ, каждая из которых цитируется по 6 раз (рис. 4). В этом случае h -индекс равен $h = 6$. Он учитывает 36 цитирований из общих 60. Рассчитаем w -индекс:

$$w = \frac{36}{6} + \frac{24}{4} = 6 + 6 = 12.$$

Первое суммируемое число характеризует первые $h = 6$ работ ученого, а второе число — оставшиеся 4 работы.

В обоих рассмотренных случаях, по нашему мнению, оценка деятельности ученого является более полной и точной, так как охватывает область не учтенных при расчете h -индекса работ и их цитирований.

Приведем третий пример, иллюстрирующий особенности учета количества цитирований как из области над квадратом Хирша, так и справа от него. Рассмотрим такой случай. Предположим, имеются два ученых, которые опубликовали 10 и 20 работ соответственно. При этом первый ученый имеет следующую последовательность числа цитирований работ, расположенных в порядке невозрастания числа цитирований: 40, 20, 10, 5, 5, 3, 2, 1, 1, 1 (рис. 5, а). Для второго ученого аналогичная последовательность такова: 10, 9, 6, 4, 3, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 (рис. 5, б). Определим показатели цитируемости этих ученых.

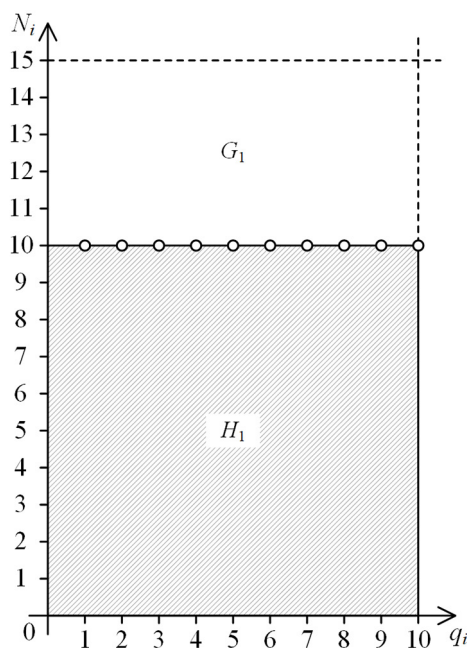


Рис. 3. Графическая иллюстрация к первому примеру

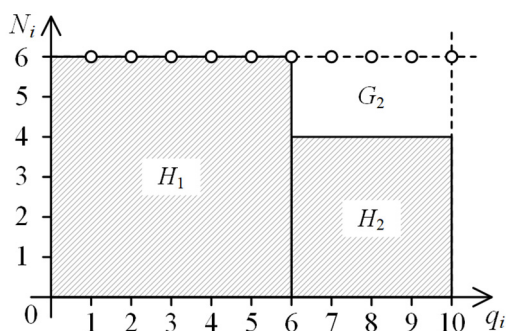


Рис. 4. Графическая иллюстрация ко второму примеру

Индексы Хирша для первого и второго ученых равны соответственно 5 и 4. В данном случае первый ученый по отношению ко второму имеет показатель более высокий. Это согласуется с общими показателями цитируемости, например, по общему числу цитирований всех работ: 88 и 54 цитирования работ первого и второго ученых соответственно. Рассчитаем w -индекс для каждого из них.

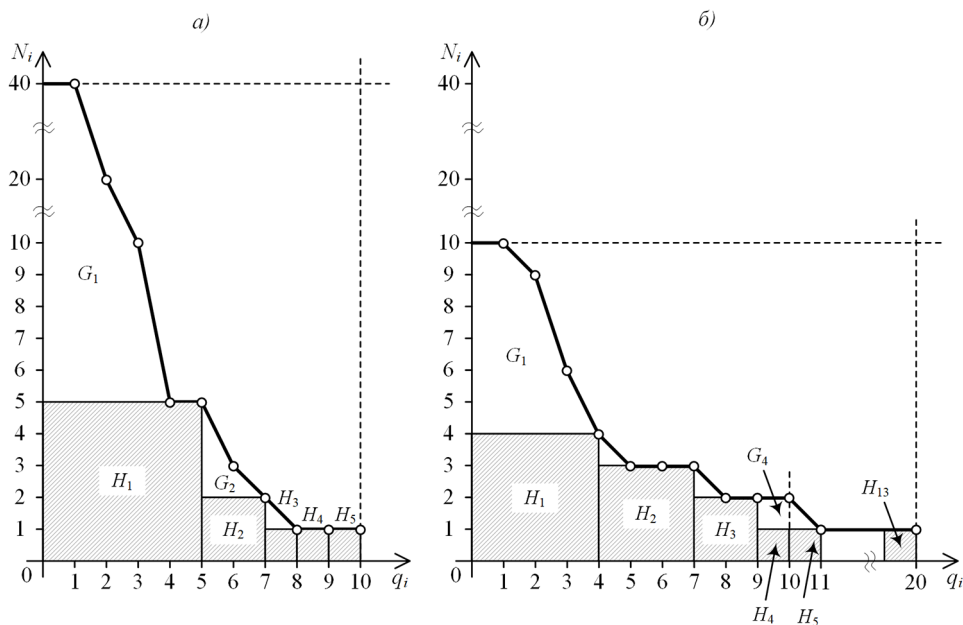


Рис. 5. Графическая иллюстрация к третьему примеру

Для первого ученого имеем:

$$w = \frac{25 + 55}{5} + \frac{4 + 1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} = 16 + 2,5 + 1 + 1 + 1 = 21,5.$$

Для второго ученого:

$$w = \frac{16 + 13}{4} + \frac{9}{3} + \frac{4}{2} + \frac{1 + 1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} = 7,25 + 3 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 22,25.$$

Сравнивая рассчитанные показатели, отмечаем, что второй ученый имеет большее значение w -индекса, чем первый.

Представленный пример показывает основной недостаток w -индекса: при большом числе малоцитируемых работ вклад в значение w -индекса этих работ может оказаться весьма существенным. Поэтому ученый, имеющий гораздо большее значение общего числа цитирований, оказывается по критерию w -индекса в более проигрышной позиции. Учет этого недостатка будет показан далее.

Результаты оценки деятельности ученых с помощью нового индекса

Для примера использования w -индекса для оценки деятельности ученых авторами были произведены расчеты с данными, полученными из базы данных РИНЦ.

Выборка была сделана 25 сентября 2019 г. по 15 ученым, работающим в области автоматизации и вычислительной техники (50.00.00 — Автоматика. Вычислительная техника) и ранжированным в порядке невозрастания h -индекса. Для каждого ученого (далее они обозначены прописными буквами латинского алфавита) были зафиксированы:

- общее число публикаций (q),
- общее число публикаций, которые были процитированы хотя бы один раз (q_c),
- общее число цитирований (N),
- число цитирований по каждой публикации.

Наличие таких данных позволило получить для каждого автора различные статистические показатели (см. табл. 3 и 4). Показатель $q/q_c \times 100\%$ характеризует долю научных работ ученого, которые определяют его рейтинг на основе цитируемости. Только для пяти ученых этот показатель превысил 75% научных работ. А для некоторых ученых он составил менее 50% научных работ. Столбцы H_i , $H_i + G_i$, G_i и P содержат данные о числе цитирований, попадающих в соответствующие области на геометрической интерпретации принципа подсчета w -индекса: H_i характеризует число ссылок для первого квадрата Хирша, $H_i + G_i$ — суммарное число ссылок области H_i и всей области над квадратом Хирша, G_i — число ссылок в области над квадратом Хирша и P — число ссылок на работы, соответствующие правой области за квадратом Хирша. Показатель k (длина цепи Хирша) дает информацию о том, какое количество итераций было проделано по предложенному алгоритму в процессе подсчета w -индекса. Коэффициент N/q характеризует среднее число цитирований в пересчете на одну публикацию, а коэффициент N/q_c — среднее число цитирований работ, которые процитированы хотя бы по одному разу. Столбец w^1 — показатель цитируемости первых h работ автора: $w^1 = \frac{H_i + G_i}{h}$. Показатель w^1 , в отличие от h -индекса, более точно учитывает самые «популярные» работы ученого (a-Index [Цыганов, 2013]). В последнем столбце табл. 4 приведено рассчитанное значение w -индекса.

Сравним между собой показатели некоторых ученых. Например, ученые, занимающие 14-ю и 15-ю позиции в рейтинге по h -индексу, имеют одинаковый индекс Хирша, равный 32. Однако области P у них существенно отличаются (в 16,36 раза), что оказывает влияние на результат подсчета w -индекса: для них эти индексы разнятся в 3,76 раза. Кроме того, на значение w -индекса оказывает влияние длина цепи Хирша и значения H_i для каждого квадрата Хирша. Для рассматриваемых ученых $k = 72$ и $k = 21$. Следует отметить, что доля хотя бы раз процитированных работ у первого ученого выше, чем у второго, что оказывает влияние на длину цепи Хирша и в конечном итоге на результаты подсчета w -индекса.

Значительно в рейтинге поменялись позиции ученых, расположенных в табл. 3 на позициях номер 2 и 8. Несмотря на то, что для первого ученого число $w^1 = 104,647$, а для второго $w^1 = 44,029$, за счет более длинной цепи Хирша (191 против 89) второй ученый имеет большее значение w -индекса. Это как раз иллюстрация рассмотренного выше в примерах третьего случая.

Таблица 3. Статистические данные для расчета w-индекса

Позиция при ранжировании по h-индексу	Ученый	b	b_c	$\%001 \times b/b_c$	N	h -индекс	H	$H + G_1$	G_1	p	k	b/N	b_c/N	w	w-индекс
1	A	377	329	87,268	15246	58	3364	12534	9170	2712	96	40,44	46,34	216,103	509,996
2	B	396	240	60,606	6415	51	2601	5337	2736	1078	89	16,199	26,729	104,647	309,878
3	C	535	301	56,262	4785	45	2025	3776	1751	1009	131	8,944	15,897	83,911	352,828
4	D	491	387	78,819	8093	42	1764	5377	3613	2716	139	16,483	20,912	128,024	470,121
5	E	160	140	87,500	5567	41	1681	4578	2897	989	39	34,794	39,764	111,659	222,192
6	F	295	194	65,763	2846	38	1444	1658	214	1188	68	9,647	14,67	43,632	214,541
7	G	581	364	62,651	4332	37	1369	2011	642	2321	158	7,456	11,901	54,351	400,588
8	H	676	488	72,189	4174	35	1225	1541	316	2633	191	6,175	8,553	44,029	508,633
9	I	246	108	43,902	1817	35	1225	1544	319	273	45	7,386	16,824	44,114	126,148
10	J	336	198	58,929	2320	34	1156	1312	156	1008	74	6,905	11,717	38,588	212,166
11	K	401	252	62,843	6025	33	1089	4848	3759	1177	106	15,025	23,909	146,909	380,547
12	L	386	268	69,430	3020	33	1089	1819	730	1201	113	7,824	11,269	55,121	302,965
13	M	276	244	88,406	5284	32	1024	3671	2647	1613	78	19,145	21,656	114,719	340,489
14	N	366	306	83,607	4170	32	1024	1438	414	2732	72	11,393	13,627	44,938	332,622
15	O	146	70	47,945	1527	32	1024	1360	336	167	21	10,459	21,814	42,500	88,414

Таблица 4. Рейтинг ученых в порядке невозрастания w -индекса

Позиция при ранжировании по w -индексу	Ученый	h -индекс	w -индекс	Позиция при ранжировании по h -индексу
1	<i>A</i>	58	509,996	1
2	<i>H</i>	35	508,633	8
3	<i>D</i>	42	470,121	4
4	<i>G</i>	37	400,588	7
5	<i>K</i>	33	380,547	11
6	<i>C</i>	45	352,828	3
7	<i>M</i>	32	340,489	13
8	<i>N</i>	32	332,622	14
9	<i>B</i>	51	309,878	2
10	<i>L</i>	33	302,965	12
11	<i>E</i>	41	222,192	5
12	<i>F</i>	38	214,541	6
13	<i>J</i>	34	212,166	10
14	<i>I</i>	35	126,148	9
15	<i>O</i>	32	88,414	15

Целесообразно сравнить также ученых, которые имеют весьма близкие значения w -индекса и занимают в «обновленном» рейтинге 1-ю и 2-ю позиции. Число цитирований у первого ученого гораздо больше числа цитирований у второго — в 3,65 раза. Однако за счет более длинной цепи Хирша у второго ученого w -индексы обоих ученых становятся практически равными.

Последние два примера указывают на весомый недостаток w -индекса — при его подсчете на каждом этапе осуществляется различное усреднение числа цитирований по отношению к числу цитируемых работ. С уменьшением значений длин сторон квадратов Хирша это усреднение быстрее приближается к числу цитирований каждой работы. К примеру, для ученого, занимающего первую позицию в обоих рейтингах, на первые $h = 58$ работ приходится 12 534 ссылки, что составляет в среднем на каждую из 58 работ по 216 ссылок, тогда как для квадратов Хирша в конце цепи вычислений на каждую цитируемую единожды работу в суммарное значение w -индекса включается показатель, равный 1 («один к одному»).

Модификации w -индекса

Как отмечено выше, несмотря на положительные качества w -индекса, он имеет недостаток, связанный с возможным большим влиянием на его общее значение малоцитируемых работ автора. Это связано с особенностями подсчета w -индекса и усреднения показателей цитируемости на каждом шаге вычисления. К примеру, можно обратить внимание на область P для ученого, занимающего 8-ю позицию

Таблица 5. Статистические данные для расчета w -индексов со значениями $t = 1, 5, 10$

Позиция при ранжированию по h -индексу	Учебный	N	h -индекс	$h^2/N \times 100\%$	w -индекс	N_1	N_5	N_{10}	$N_1/N \times 100\%$	$N_5/N \times 100\%$	$N_{10}/N \times 100\%$	w^1 -индекс	w^5 -индекс	w^{10} -индекс
1	A	15246	58	22,065	509,996	12534	14435	14796	82,212	94,681	97,048	216,103	307,032	350,130
2	B	6415	51	40,546	309,878	5337	6025	6181	83,196	93,920	96,352	104,647	160,045	188,712
3	C	4785	45	42,32	352,828	3776	4235	4384	78,913	88,506	91,620	83,911	130,161	158,661
4	D	8093	42	21,797	470,121	5377	7026	7445	66,440	86,816	91,993	128,024	212,338	258,921
5	E	5567	41	30,196	222,192	4578	5463	5531	82,235	98,132	99,353	111,659	173,659	192,192
6	F	2846	38	50,738	214,541	1658	2559	2679	58,257	89,916	94,132	43,632	102,525	127,375
7	G	4332	37	31,602	400,588	2011	3478	3848	46,422	80,286	88,827	54,351	134,627	178,755
8	H	4174	35	29,348	508,633	1541	2669	3145	36,919	63,943	75,347	44,029	116,206	166,166
9	I	1817	35	67,419	126,148	1544	1736	1773	84,975	95,542	97,578	44,114	74,314	88,148
10	J	2320	34	49,828	212,166	1312	2052	2157	56,552	88,448	92,974	38,588	96,166	119,666
11	K	6025	33	18,075	380,547	4848	5561	5730	80,465	92,299	95,104	146,909	203,142	232,714
12	L	3020	33	36,060	302,965	1819	2502	2686	60,232	82,848	88,940	55,121	110,048	141,048
13	M	5284	32	19,379	340,489	3671	4720	4971	69,474	89,326	94,076	114,719	182,092	218,389
14	N	4170	32	24,556	332,622	1438	2944	3560	34,484	70,600	85,372	44,938	125,439	181,494
15	O	1527	32	67,060	88,414	1360	1500	1516	89,064	98,232	99,280	42,500	67,914	77,414
Средние значения				36,733	—	—	—	—	67,323	87,566	92,533	—	—	—

в табл. 1: 2 633 цитирования за счет большого числа итераций работы алгоритма при вычислении w -индекса $k = 191$ дало существенное изменение позиции при ранжировании по w -индексу. Учитывая такую особенность нового w -индекса, можно предложить различные его модификации.

Одной из модификаций w -индекса является w^t -индекс, где t — число итераций (циклов) работы алгоритма вычисления w -индекса. Число t выбирается заранее и ограничивает число цитирований всех работ ученого: учитывается только число цитирований первых $\sum_{i=1}^t h_i$ его работ.

Как показывают исследования на реальных статистических данных из РИНЦ, для многих ученых как раз не учитывается весомое количество ссылок первых нескольких «квадратов Хирша» (h_1, h_2, h_3). Например, в табл. 5 приведены результаты расчетов для w^1 -индекса, w^5 -индекса и w^{10} -индекса, полученных, соответственно, при 1, 5 и 10-й итерациях работы предложенного алгоритма. Даже при первой итерации учитывается значительно больше цитирований работ каждого ученого. Например, если в среднем индекс Хирша учитывает только 36,733% цитирований всех работ ученого для рассматриваемой выборки, то w^1 -индекс учитывает в среднем 67,323% цитирований всех работ. Для w^5 -индекса и w^{10} -индекса этот показатель еще выше: получены средние значения в 87,566% и 92,533% всех ссылок на работы ученых, соответственно.

Другим подходом к модификации w -индекса является ограничение длины цепи учитываемых «квадратов Хирша». Например, можно предложить следующую модификацию — w^{-r} -индекс, где число r указывает на число цитирований статей, которые не учитываются при расчетах. Например, w^{-1} -индекс не учитывает все цитирования работ, которые процитированы только по одному разу, w^{-2} -индекс не учитывает все цитирования работ, которые процитированы только один и два раза, и т. д.

В табл. 6 представлены рассчитанные значения w^{-r} -индексов для $r = 1, 2, 3$, из которых следует, что в среднем w^{-1} -индекс учитывает 98,666% цитирований всех работ, w^{-2} -индекс учитывает 97,181% цитирований всех работ и w^{-3} -индекс учитывает 95,512% цитирований всех работ. Напомним, что при расчете именно w -индекса учитываются все 100% цитирований. Таким образом, учет только тех работ, которые процитированы 1, 2 или 3 раза, не столь существенно влияет на общую долю учтенных цитирований, однако весомо изменяет значение самого коэффициента.

Таким образом, модификации w -индекса позволяют фактически осуществлять «настройку» самого коэффициента, исходя из целей рейтинга (например, ранжирования специалистов по значимости работ в пределах организации или, например, в пределах специальности исследований).

Заключение

Предложенный в настоящей работе w -индекс, по мнению авторов, позволяет более точно оценивать вклад ученого в свою предметную область науки, чем широко распространенный индекс Хирша. Новый индекс учитывает не только число цитирований из квадрата Хирша (h^2), но и те $N - h^2$ цитирований всех работ, которые в этот квадрат не попадают. Для различных ученых это может быть весомое число неучтенных цитирований, как первых по рейтингу Хирша h работ, так и остальных $q - h$ работ. Таким образом, w -индекс позволяет учесть и значимость высокоцити-

Таблица 6. Статистические данные для расчета w^r -индексов со значениями $r = 1, 2, 3$

Позиция при ранжированию по h -индексу	Учедный	N	h -индекс	$h^2/N \times 100\%$	w -индекс	N_{-1}	N_{-2}	N_{-3}	$N_{-1}/N \times 100\%$	$N_{-2}/N \times 100\%$	$N_{-3}/N \times 100\%$	w^{-1} -индекс	w^{-2} -индекс	w^{-3} -индекс
1	A	15246	58	22,065	509,996	15194	15129	15066	99,659	99,233	98,819	457,996	425,496	404,496
2	B	6415	51	40,546	309,878	6363	6287	6231	99,189	98,005	97,132	257,878	219,878	201,212
3	C	4785	45	42,320	352,828	4712	4592	4482	98,474	95,967	93,668	279,828	219,828	183,161
4	D	8093	42	21,797	470,121	8023	7930	7848	99,135	97,986	96,973	409,121	362,621	335,287
5	E	5567	41	30,196	222,192	5543	5526	5507	99,569	99,264	98,922	198,192	189,692	183,359
6	F	2846	38	50,738	214,541	2806	2769	2695	98,595	97,294	94,694	174,541	156,041	131,375
7	G	4332	37	31,602	400,588	4223	4134	4061	97,484	95,429	93,744	291,588	247,088	222,755
8	H	4174	35	29,348	508,633	4066	3920	3783	97,413	93,915	90,632	400,633	327,633	281,966
9	I	1817	35	67,419	126,148	1785	1764	1736	98,239	97,083	95,542	94,148	83,648	74,314
10	J	2320	34	49,828	212,166	2274	2229	2157	98,017	96,078	92,974	166,166	143,666	119,666
11	K	6025	33	18,075	380,547	5959	5885	5812	98,905	97,676	96,465	314,547	277,547	253,214
12	L	3020	33	36,060	302,965	2955	2859	2776	97,848	94,669	91,921	237,965	189,965	162,298
13	M	5284	32	19,379	340,489	5240	5203	5122	99,167	98,467	96,934	296,489	277,989	250,989
14	N	4170	32	24,556	332,622	4137	4104	4029	99,209	98,417	96,619	299,622	283,122	258,122
15	O	1527	32	67,060	88,414	1513	1500	1491	99,083	98,232	97,642	74,414	67,914	64,914
Средние значения				36,733	—	—	—	—	98,666	97,181	95,512	—	—	—

руемых публикаций, игнорирование которой зачастую выделяется в качестве недостатка индекса Хирша, и значимость остальных работ автора. Кроме того, преимуществом нового индекса является и сам способ подсчета, связанный с определением «квадратов Хирша» (цепи Хирша) на каждом шаге итерации (от 1 до k).

Важным является и то, что введенный в рассмотрение w -индекс может быть легко модифицирован и «настроен» под конкретную особенность, например, повышение учета цитируемости значимых работ, отбрасывание менее цитируемых работ и т. п.

Использование нового индекса, как и индекса Хирша, к сожалению, не решает проблемы искусственного увеличения показателей («накрутки»), а также не учитывает вопросы самоцитирования, количества соавторов и т. д. Некоторые из особенностей возможно учесть при подсчете комплексных коэффициентов на основе w -индекса, где на каждом шаге итерации, например, учитываются те или иные особенности (например, число соавторов по методике, описанной в [Марвин, 2015]).

В заключение отметим, что новый индекс цитируемости может применяться для оценки деятельности ученых наравне с индексом Хирша, дополнять его и даже в некоторой степени его заменить.

Литература

- Марвин С. В. Альтернативная дробная модификация индекса Хирша, учитывающая количество авторов цитируемых статей // Управление большими системами. 2015. № 56. С. 108–122.
- Михайлов О. В. Новая версия индекса Хирша — j -индекс // Вестник Российской академии наук. 2014а. Т. 84. № 6. С. 532–535. DOI: 10.7868/S0869587314060085.
- Михайлов О. В. О возможной модификации индексов Хирша и Эгга с учетом соавторства // Социология науки и технологий. 2014б. Т. 5. № 3. С. 48–56.
- Новиков Д. А. Методология управления. М.: Либроком, 2011. 128 с.
- Полянин А. Д. Недостатки индексов цитируемости и Хирша и использование других наукометрических показателей // Математическое моделирование и численные методы. 2014. № 1. С. 131–144.
- Попов С. Как улучшить индекс Хирша? // Троицкий вариант. 2010. 22 июня. № 56. С. 3.
- Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами. 2013. № 44. С. 248–261.
- Штовба С. Д., Штовба Е. В. Обзор наукометрических показателей для оценки публикационной деятельности ученого // Управление большими системами. 2013. № 44. С. 262–278.
- 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. Iss. 46. P. 16569–16572. URL: <https://doi.org/10.1073/pnas.0507655102>.
- Raheel M., Ayaz S., Afzal M. T. Evaluation of h -index, Its Variants and Extensions Based on Publication Age & Citation Intensity in Civil Engineering // Scientometrics. 2018. Vol. 114. Iss. 3. P. 1107–1127. URL: <https://doi.org/10.1007/s11192-017-2633-2>.

The “Adjusted” Index for the Scientist Performance Assessment

VALERIY V. SAPOZHNIKOV

Emperor Alexander I St Petersburg state transport university,
St Petersburg, Russia;
e-mail: port.at.pgups@gmail.com

VLADIMIR V. SAPOZHNIKOV

Emperor Alexander I St Petersburg state transport university,
St Petersburg, Russia;
e-mail: at.pgups@gmail.com

DMITRIY V. EFANOV

Russian university of transport,
Moscow, Russia;
e-mail: TrES-4b@yandex.ru

The article proposes a new scientometric indicator, which makes it possible to get a more accurate assessment of the contribution of a particular scientist to the relevant subject area of science than using the H-index. The new index, as well as the H-index, is a single number, it is simply calculated and has a clear geometrical interpretation. The article provides an algorithm for determining a new scientometric indicator, notes its features, advantages and disadvantages. The article gives some examples of its calculation.

Keywords: scientometric indicators, scientist performance assessment, citation index, H-index, w-index.

References

- Hirsch, J. E. (2005). An Index to Quantify an Individual's Scientific Research Output, *Proceedings of National Academy of Sciences of the USA*, vol. 102, iss. 46, pp. 16569–16572. URL: [https:// doi.org/10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102).
- Marvin, S. V. (2015). Al'ternativnaya drobnaya modifikatsiya indeksa Hirsha, uchityvayushchaya kolichestvo avtorov tsitiruyemykh statey [Alternative fractional modification of hirsch index to care for authors' count in article cited], *Upravleniye bol'shimi sistemami*, no. 56, pp. 108–122 (in Russian).
- Mikhaylov, O. V. (2014a). Novaya versiya indeksa Hirsha — *j*-indeks [New Hirsch index version — *j*-index], *Vestnik Rossiyskoy akademii nauk*, vol. 84, no. 6, pp. 532–535. URL: [https:// doi:10.7868/S0869587314060085](https://doi.org/10.7868/S0869587314060085) (in Russian).
- Mikhaylov, O. V. (2014b). O vozmozhnoy modifikatsii indeksov Hirsha i Egga s uchetom soavtorstva [Possible modification of the Hirsch and Egg indices, taking into account co-authorship], *Sotsiologiya nauki i tekhnologii*, vol. 5, no. 3, pp. 48–56 (in Russian).
- Novikov, D. A. (2011). *Metodologiya upravleniya* [Management methodology], Moskva, Librokom (in Russian).
- Polyanin, A. D. (2014). Nedostatki indeksov tsitiruемости i Hirsha i ispol'zovaniye drugikh naukometricheskikh pokazateley [Disadvantages of citation and Hirsch indices and the use of other

scientometric indicators], *Matematicheskoye modelirovaniye i chislennyye metody*, no. 1, pp. 131–144 (in Russian).

Popov, S. (2010). Kak uluchshit' indeks Hirsha? [How to improve the Hirsch index?], *Troitskiy variant*, no. 56, June 22, p. 3 (in Russian).

Shtovba, S. D., Shtovba, E. V. (2013). Obzor naukometricheskikh pokazateley dlya otsenki publikatsionnoy deyatel'nosti uchenogo [A survey on scientometric indicators for assessment of researcher's publication activity], *Upravleniye bol'shimi sistemami*, no. 44, pp. 262–278 (in Russian).

Tsyganov, A. V. (2013). Kratkoye opisaniye naukometricheskikh pokazateley, osnovannykh na tsitiruyemosti [Brief review of main scientometric indices based on citations], *Upravleniye bol'shimi sistemami*, no. 44, pp. 248–261 (in Russian).

Raheel, M., Ayaz, S., Afzal, M. T. (2018). Evaluation of *h*-index, its Variants and Extensions Based on Publication Age & Citation Intensity in Civil Engineering, *Scientometrics*, vol. 114, iss. 3, pp. 1107–1127. URL: <https://doi.org/10.1007/s11192-017-2633-2>.