УДК 004.09+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 НТЦ «Комплексные системы мониторинга», Санкт-Петербург
Высшая школа транспорта
Институт машиностроения, материалов и транспорта
Санкт-Петербургского политехнического университета
Петра Великого, Санкт-Петербург
Кафедра «Автоматика, телемеханика и связь
на железнодорожном транспорте»
Российский университет транспорта, Москва*

ИТЕРАЦИОННЫЕ МЕТОДЫ ФОРМИРОВАНИЯ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ С УЧЕТОМ ПОЛНОГО МНОЖЕСТВА ЦИТИРУЕМЫХ РАБОТ

Проанализированы широко известные наукометрические показатели, основанные на учете числа цитирований работ ученого. Предложены некоторые способы формирования нового наукометрического показателя, учитывающие полное множество цитирований работ ученого. Эти способы имеют простую и понятную геометрическую интерпретацию и дают значения показателей в виде единственного числа, что удобно с позиции сравнения показателей различных ученых и формирования рейтингов. Продемонстрирована эффективность использования итерационного принципа формирования индекса цитируемости, при котором выделяется некоторое количество первых работ в ранжированном в порядке невозрастания числа цитирований перечне работ ученого, усреднение числа цитирований, исключение из рассмотрения этих работ и повторение всех процедур с заданным способом выделения первых работ на каждом шаге итерации. Способ выделения первых работ из списка позволяет выделить основание цепи промежуточных показателей.

В работе авторов предлагается в качестве основания брать индекс Хирша на каждом шаге итерации (выделять основание Хирша) и строить цепь убывающих индексов с усреднением числа цитирований на каждой итерации. Такой подход позволяет более полно учесть вклад ученого в развитие соответствующей предметной области. В качестве основания может выбираться и другое число, например, осуществляться определение индекса Эгге или Михайлова. Основание может быть и переменным – в статье рассмотрены такие случаи. Расчеты, приведенные в ней, дополняются экспериментальными результатами, полученными при использовании статистической информации из базы данных российского индекса научного цитирования. Предложенный

итерационный способ формирования индекса цитирования позволяет осуществлять «настройку» индекса и, при необходимости, не учитывать малоцитируемые работы или учитывать только определенный процент цитирований от общего их количества. Это может быть актуальным для конкретных предметных областей науки и научных или образовательных организаций.

Оценка деятельности ученых, наукометрические показатели, учет числа цитирований, индекс Хирша, индекс Эгге, индекс Михайлова

DOI: 10.20295/2412-9186-2021-7-1-137-166

Введение

С развитием компьютерной техники и информационных технологий появилась возможность не только доступа к большому объему научных материалов (препринтов, статей, монографий, результатов исследований и пр.), но и различного рода оценки деятельности ученых. В этой связи особое внимание уделяется индивидуальным показателям деятельности ученых, позволяющим оценить их вклад в конкретную предметную область [1–5]. Существуют десятки наукометрических показателей, учитывающих ту или иную «деталь» в деятельности ученых: число опубликованных работ, количество цитирований каждой из них, вклад соавторов, уровень издания, конкретную тематику научных работ и т. д. [6–10]. Учеными по всему миру вводятся наукометрические показатели и модифицируются уже известные. В отдельных организациях создаются специализированные издания, освещающие основные достижения в области наукометрии и управления научной деятельностью [11–20].

Подавляющее большинство, если не все наукометрические показатели не позволяют оценить «на равных» вклад ученых в различных областях науки. К примеру, абсолютно бесполезно сравнивать наукометрические показатели из области биологии и из области автоматизации: в первой работает куда большее количество ученых и публикуется гораздо большее число научных работ, чем во второй. Однако термин «оценка научной деятельности» достаточно часто упоминается в научных организациях и учебных заведениях, а наукометрические показатели становятся чуть ли не основным мерилем работы сотрудников. Более того, вопросы наукометрии в управлении научной деятельностью регламентируются на государственном уровне. Так, в Российской Федерации (как и в подавляющем большинстве стран) приоритет в оценке результативности и эффективности деятельности ученых отдается числу публикаций в журналах, индексируемых в международных наукометрических базах Web of Science (США) и Scopus (ЕС). Это зафиксировано в постановлении Правительства РФ от 8 апреля 2009 года № 312 «Об оценке и о мониторинге результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения». Есть и более свежий документ — «Методика расчета качественного показателя государственного задания “Комплексный балл публикационной результативности”» (утвержден Минобрнауки РФ 30 декабря 2019 года).

Среди разнообразия вспомогательных показателей оценки деятельности ученых наиболее популярным является индекс Хирша (h -индекс), предложенный профессором физики Калифорнийского университета в Сан-Диего Хорхе Хиршем 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. Индекс Хирша имеет и достоинства и недостатки, отмеченные как самим первооткрывателем в первой работе, так и в публикациях других ученых, развивавших и развивающих идею Хирша [7, 9, 13, 14]. Так, индекс Хирша учитывает число наиболее цитируемых публикаций ученого, но никак не учитывает те работы, которые оказываются за пределами первых h работ.

Данная статья освещает результаты исследований авторов в области построения нового показателя оценки деятельности ученых на основе учета всех цитирований всех его работ. За пределами исследований остались возможные вариации с учетом количества соавторов, уровня издания, числа самоцитирований и пр. Эти вопросы актуальны и поднимаются в большом числе предыдущих публикаций и с легкостью могут быть адаптированы к использованию с предлагаемыми показателями. Следует также отметить, что все рассуждения далее дополняются результатами экспериментов с выборками из базы данных российского индекса научного цитирования (РИНЦ). Но предлагаемые показатели могут рассчитываться и для других баз данных, вплоть до использования только внутри отдельных организаций с их репозиториями.

1. Известные наукометрические показатели

Во введении отмечается, что наиболее известным наукометрическим показателем является индекс Хирша [2]. Он имеет довольно простую геометрическую интерпретацию. Если ранжировать все работы автора, имеющиеся в определенной базе данных, по невозрастанию числа их цитирований, и нанести работы в виде точек на координатной плоскости, где оси абсцисс будут соответствовать научным работам (q_i), а оси ординат — число их цитирований (число n_i), то можно выделить квадрат с длиной стороны h . Этот квадрат (h -квадрат Хирша) будет учитывать h^2 цитирований первых h научных работ ученого. Однако все остальные работы и их цитирования окажутся за пределом данного квадрата.

Зададимся произвольным статистическим рядом, охватывающим результаты деятельности некоторого абстрактного ученого, и ранжируем их в порядке невозрастания числа цитирований научных работ (табл. 1). Всего ученым опубликованы 32 научные работы, процитированные 209 раз.

Располагая представленные в таблице 1 данные на координатной плоскости в виде точек, получаем геометрическую интерпретацию деятельности ученого. На данной плоскости нетрудно выделить квадрат Хирша. Длина его стороны для рассматриваемого случая равна $h=8$. Таким образом, индекс Хирша учитывает в рассматриваемом примере $h^2=64$ цитируемости ученого. Это составляет чуть

Таблица 1. Исходные данные к примерам

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															
16	15	15	14	13	12	10	9	8	8	8	7	7	6	6	6
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	5	5	4	4	3	3	3	3	3	2	2	1	1	0

более 30 % от общего числа цитирований работ ученого. При этом неучтенными остались 40 цитирований первых восьми работ (как раз образующих квадрат Хирша), находящихся над квадратом Хирша, что составляет около 20 % цитирований всех работ ученого, а также 105 цитирований работ из области, расположенной справа от квадрата Хирша, что составляет около 50 % цитирований всех работ ученого. Индекс Хирша учел только малую часть деятельности ученого, выделив лишь наиболее популярные работы автора. Остальные работы оказались фактически исключены (как будто автор их никогда не публиковал).

Индекс Хирша как раз и задумывался автором для выделения наиболее популярных работ, однако в дальнейшем так прочно вошел в обиход, что сегодня оказывается важнейшим показателем, на который ориентируются во всем мире. Для выделения среди научных работ автора самых популярных публикаций был разработан аналог h -индекса, вычисляемый следующим образом [21]. Ученый имеет g -индекс, равный g , если g его работ процитированы в совокупности не менее чем g^2 раз, а $g+1$ работа цитируется уже менее $(g+1)^2$ раз. Сегодня этот индекс часто называют по имени ученого, введшего его в рассмотрение, — индекс Эгге (g -индекс). Индекс Эгге нивелирует важный минус индекса Хирша, связанный с недостаточно полным учетом наиболее популярных работ ученого. Как было показано в примере выше, индекс Хирша не учел большое число цитирований работ в области над h -квадратом.

Модифицируем индекс Эгге так, как это предложено в работе [22]: будем его определять аналогично индексу Хирша и выделять g работ с числом цитирований каждой не менее величины g^2 . Далее будем использовать только модифицированный индекс Эгге и обозначать его также через g -индекс. На рисунке 1 для тех же данных, что и в предыдущем примере, получена геометрическая интерпретация g -индекса. На плоскости сформирован прямоугольник размером $g \times g^2$, в который попадает 27 цитирований первых трех работ ученого. Назовем его g -прямоугольником Эгге. Оказались неучтенными около 9 % цитирований трех наиболее популярных работ и практически 78 % цитирований 29-ти менее популярных работ.

Рис. 1. Геометрическая интерпретация известных показателей цитируемости

В большом числе научных работ на тему использования h - и g -индексов, например в [22], обсуждаются вопросы целесообразности использования индекса Эгге в ряде предметных областей (даже в общенациональном масштабе), т. к. его значение даже для ведущих ученых оказывается крайне низким. Известен [22] способ модификации h - и g -индексов, позволяющий получить новый показатель цитируемости — j -индекс (индекс Михайлова). Ученый имеет j -индекс, равный j , в том случае, если j работ ученого процитированы не менее чем $j^{3/2}$ раз, а остальные $Q-j$ работ цитируются не более чем $j^{3/2}$ раз. Как показывает практика, в рассматриваемом случае гораздо труднее определить верхнюю границу числа цитирований: получаемые значения зачастую оказываются дробными. Автор данного показателя предлагает определять целое снизу от получаемой величины. Несмотря на это, поскольку индекс определяет всего лишь верхнюю границу учитываемых цитирований, можно условиться определять и целое сверху от получаемого значения $j^{3/2}$ (именно это и сделано авторами данной работы в дальнейшем изложении результатов).

Геометрически j -индекс изображается на плоскости цитирований в виде прямоугольника размером $j \times j^{\frac{3}{2}}$ — j -прямоугольника Михайлова. Получаемый прямоугольник имеет значение длины малой стороны, лежащее в промежутке $h > j > g$. Число цитирований, учитываемых j -индексом, больше, чем число учитываемых g -индексом. Для приведенного ранее примера j -индекс учитывает 60 цитирований работ (около 29 % от общего числа цитирований, что сравнимо с аналогичным показателем h -индекса). Не учтенными оказываются 13 цитирований над j -прямоугольником (примерно 6 % от общего числа цитирований) и 136 цитирований работ, расположенных в правой области от прямоугольника (примерно 65 % от общего числа цитирований).

Описанные выше показатели направлены на уточнение оценки цитируемости наиболее популярных работ ученого. Все данные показатели имеют один общий недостаток — не учитывают все цитирования наиболее популярных работ и во все обходят стороной все цитирования наименее популярных работ ученого. Отсюда следуют возможности модификации представленных индексов. Часть таких модификаций известна. Например, известен показатель, усредняющий общее число цитирований первых h работ ученого — a -Index [12]. Значение данного индекса для рассматриваемого примера равно 13.

Известны и другие разнообразные модификации индекса Хирша и различные производные показатели, основанные на цитировании работ ученого [9, 12, 13]. Все они, однако, не дают целостной картины в оценке деятельности ученого и дополняют друг друга.

Исследования показали, что эффективным является использование итерационного принципа формирования индекса цитируемости, при котором выделяется некоторое количество первых работ в ранжированном в порядке невозрастания числа цитирований перечне работ ученого, усреднение числа

цитирований, исключение из рассмотрения этих работ и повторение всех процедур с заданным способом выделения первых работ на каждом шаге итерации. Способ выделения первых работ из списка позволяет выделить основание цепи промежуточных показателей.

Далее обсуждаются результаты использования итерационного принципа формирования наукометрического показателя.

2. Итерационные методы формирования наукометрических показателей

2.1. Распределение числа цитирований между группами работ

Проанализируем возможности формирования такого наукометрического показателя, который бы учитывал все цитирования всех работ ученого. Наиболее простым представляется следующий подход.

Алгоритм 1. Правила подсчета r^δ -индекса:

1. Работы ученого ранжируются в порядке невозрастания числа их цитирований.

2. Выбирается целое число δ ($Q \geq \delta \geq 2$).

3. Работы автора в сформированном списке делятся на δ равных или примерно равных групп.

4. Для каждой группы подсчитывается среднее число цитирований работ:

$$w^i = \frac{n_\delta}{q_\delta} = \frac{\sum_{j=1}^{j_{\max}} n_j}{q_\delta}, \quad (1)$$

где q_δ — число научных работ в группе δ ; $j_{\max}$ — максимальный индекс работы в рассматриваемой группе.

5. Подсчитывается сумма полученных показателей для каждой группы:

$$w = \sum_{i=1}^{i=\delta} w^i. \quad (2)$$

Определим r^δ -индекс для рассмотренного выше примера при $\delta=4$. Каждая из групп будет иметь по восемь научных работ. В таблице 2 приводится расчет. Для рассматриваемого примера r^δ -индекс определяется величиной 26,125, что более полно оценивает вклад ученого в предметную область.

Тем не менее, подобный показатель, хоть и более полно учитывающий число цитирований работ, не может использоваться для исчерпывающей оценки деятельности ученого. Возникает вопрос выбора числа δ , который никак не формализован. Кроме того, как видно даже из простого примера, для первой группы работ происходит существенно большее усреднение числа цитирований, чем для последующих групп.

Таблица 2. Подсчет r^4 -индекса

δ	n_δ	q_δ	w^i
1	104	8	13
2	56	8	7
3	34	8	4,25
4	15	8	1,875
w	209	32	26,125

С целью оценки особенностей подсчета r^4 -индекса авторы провели эксперименты с данными из базы РИНЦ. Выборка сделана 25 сентября 2019 года по 15 ученым, работающим в области автоматики и вычислительной техники (50.00.00 – «Автоматика. Вычислительная техника»). Для каждого ученого зафиксированы показатели цитируемости по каждой из представленных работ. Более свежая выборка не производилась, поскольку все расчеты выполнялись для различных индексов и важно было сохранить неизменными исходные данные к расчетам. В опубликованной ранее работе авторов [23] присутствуют экспериментальные результаты по данной выборке – можно провести корректное сравнение результатов.

В таблице 3 приведены полученные результаты при вычислении r^4 -индексов для выбранных 15 ученых.

Анализируя данные таблицы 3, отмечаем, что для различных ученых распределение числа цитирований между работами оказывается различным. К примеру, в первую четверть работ у ученых D , E и K попадает всего две работы, а у ученого M – три, и между ними распределена четверть всех цитирований, тогда как для других ученых это число больше. Соответственно, усреднение между малым количеством работ дает больший вклад, чем для других ученых, и существенно влияет на рейтинг. Во вторую и третью четверти попадает куда большее количество работ, что приводит к еще большему усреднению показателей цитируемости. В четвертой четверти оказывается подавляющее количество работ, однако вклад их в общее значение показателя оказывается несущественным и на рейтинг кардинально не влияющим.

2.2. Итерационный принцип подсчета наукометрического показателя

Рассмотрим другой подход к формированию показателя оценки деятельности ученого на основе цитируемости. Он подразумевает использование в качестве «основания» индекса Хирша, но базируется на итерационном принципе формирования [23, 24].

Алгоритм 2. Правила подсчета w -индекса:

1. Работы ученого ранжируются в порядке невозрастания числа их цитирований.

Таблица 3. r^A -индексы для рассмотренных ученых

Позиция при ранжировании по h -индексу	Ученый	Q	N	h	δ				n_δ				w^i				w
					1	2	3	4	1	2	3	4	1	2	3	4	
1	A	329	15246	58	5	13	25	286	3557	4078	3837	3774	711,4	313,692	153,48	13,196	1191,768
2	B	240	6415	51	9	14	19	198	1561	1626	1614	1614	173,444	116,143	84,947	8,152	382,686
3	C	301	4791	45	10	14	18	259	1120	1265	1230	1176	112	90,357	68,333	4,541	275,231
4	D	387	8105	42	2	17	44	324	1976	2060	2060	2009	988	121,176	46,818	6,201	1162,195
5	E	140	5579	41	2	10	21	107	1330	1456	1396	1397	665	145,6	66,476	13,056	890,132
6	F	194	2846	41	14	18	21	141	687	737	723	699	49,071	40,944	34,429	4,957	129,401
7	G	364	4332	37	16	25	47	276	1055	1103	1105	1069	65,938	44,12	23,511	3,873	137,442
8	H	488	4228	35	22	36	88	342	1075	1056	1054	1043	48,864	29,333	11,977	3,05	93,224
9	I	108	1817	35	8	11	11	78	419	494	442	462	52,375	44,909	40,182	5,923	143,389
10	J	198	2358	34	14	16	23	145	591	582	603	582	42,214	36,375	26,217	4,014	108,82
11	K	252	6024	33	2	5	17	228	1215	1671	1629	1509	607,5	334,2	95,824	6,618	1044,142
12	L	268	3020	33	8	17	30	213	760	765	744	751	95	45	24,8	3,526	168,326
13	M	244	5284	32	3	10	30	201	1185	1418	1369	1312	395	141,8	45,633	6,527	588,96
14	N	306	4170	32	21	36	62	187	1061	1030	1054	1025	50,524	28,611	17	5,481	101,616
15	O	70	1527	32	7	8	10	45	384	358	379	406	54,857	44,75	37,9	9,022	146,529

2. Полагается $i=1$.
3. Определяются индекс Хирша ученого и число $h_i=h$.
4. Подсчитывается среднее число цитирований первых $h_i=h_1$ работ:

$$w^i = \frac{s_i}{h_i}, \quad (3)$$

где s_i — общее число цитирований h_i работ.

5. Запоминается число w^i , а первые h_i работ исключаются из рассмотрения и удаляются из множества работ и цитирований.

6. Рассматривается следующая итерация при $i:=i+1$.

7. Проверяется наличие работ в списке с цитированием больше 0. Если они есть, то повторяются пункты 3–6 алгоритма. Если нет, можно переходить к пункту 8.

8. Подсчитывается сумма полученных показателей w^i :

$$w = \sum_{i=1}^{i=k} w^i, \quad (4)$$

где k — количество итераций подсчета суммы до того, как «очередной» квадрат Хирша не будет иметь нулевую сторону ($h_i \neq 0$).

Параметр k является крайне важным. Назовем его *длиной цепи*, т. к. он определяет, какое число h -квадратов должно быть построено при вычислении w -индекса.

Приведем пример подсчета w -индекса на исходных данных, представленных в таблице 1.

Первые $h_1=h$ работ дают следующий результат:

$$w^1 = \frac{s_1}{h_1} = \frac{16+15+15+14+13+12+10+9}{8} = \frac{104}{8} = 13.$$

После удаления первых $h_1=8$ научных работ получаем таблицу 4.

Таблица 4. Ранжированные работы после удаления первых $h_1=8$ научных 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	6	5	5	5	5
Номера работ в ранжированном списке											
13	14	15	16	17	18	19	20	21	22	23	24
Число цитирований каждой работы											
4	4	3	3	3	3	3	2	2	1	1	0

Индекс Хирша по данным таблицы 4 равен $h_2=6$. Таким образом, имеем:

$$w^2 = \frac{s_2}{h_2} = \frac{8+8+8+7+7+6}{6} = \frac{44}{6} = 7,333.$$

Удаляем первые $h_2=6$ работ ученого, что дает новый список, для которого повторяются указанные действия, и видим такой ряд промежуточных значений:

$$w^3 = \frac{s_3}{h_3} = \frac{27}{5} = 5,4, \quad w^4 = \frac{s_4}{h_4} = \frac{13}{3} = 4,333, \quad w^5 = \frac{s_5}{h_5} = \frac{9}{3} = 3,$$

$$w^6 = \frac{s_6}{h_6} = \frac{6}{2} = 3, \quad w^7 = \frac{s_7}{h_7} = \frac{4}{2} = 2, \quad w^8 = \frac{s_8}{h_8} = \frac{1}{1} = 1, \quad w^9 = \frac{s_9}{h_9} = \frac{1}{1} = 1.$$

Длина цепи (в данном случае — Хирша) равна $k=9$.

Суммируя значения, получаем итоговое значение w -индекса:

$$w = w^1 + w^2 + \dots + w^8 + w^9 = 40,066.$$

На рисунке 2 представлена геометрическая интерпретация w -индекса на данных описанного примера. Серым фоном выделены квадраты Хирша, получаемые на каждом этапе итерации.

Введенный w -индекс более полно учитывает все множество работ ученого. Примеры, демонстрирующие его преимущества перед известными показателями, даны в [23]. Основной недостаток w -индекса в том, что при большом числе малоцитируемых работ их вклад в значение w -индекса может оказаться весьма существенным. Поэтому один ученый, имеющий большее значение общего числа цитирований, чем другой, оказывается по критерию w -индекса в проигрышной позиции.

В таблице 5 приводятся результаты экспериментов с той же выборкой, что и ранее для h -индекса. Более полный учет деятельности ученых позволил изменить их положение в рейтинге и указать более обоснованные позиции. К примеру, ученый H имеет 676 публикаций (Q), среди которых процитировано хотя бы раз 488 работ (Q^{1+}), его индекс Хирша равен $h=35$. Аналогичный индекс Хирша имеет ученый I с числом публикаций 246, среди которых процитировано хотя бы единожды 108 работ.

В обновленном рейтинге значения w -индекса для ученых существенно разнятся: 508,633 и 126,148, что более справедливо. Индекс Хирша для большого числа ученых совпадает, тогда как w -индекс может совпасть только в очень редких случаях и всегда позволяет ранжировать ученых.

На рисунке 3 показаны два рейтинга ученых по h -индексу и предлагаемому w -индексу. Позиции ученых изменились, однако при этом учтены все цитирования всех работ, а не только h^2 цитирований первых работ.

Рис. 2. Геометрическая интерпретация w-индекса

Таблица 5. w-индексы для рассмотренных ученых

Позиция при ранжи- ровании по h -индексу	Ученый	Q	Q^{1+}	$Q^{1+}/Q \times 100\%$	N	h -индекс	w-индекс
1	A	377	329	87,268	15246	58	509,996
2	B	396	240	60,606	6415	51	309,878
3	C	535	301	56,262	4785	45	352,828
4	D	491	387	78,819	8093	42	470,121
5	E	160	140	87,5	5567	41	222,192
6	F	295	194	65,763	2846	38	214,541
7	G	581	364	62,651	4332	37	400,588
8	H	676	488	72,189	4174	35	508,633
9	I	246	108	43,902	1817	35	126,148
10	J	336	198	58,929	2320	34	212,166
11	K	401	252	62,843	6025	33	380,547
12	L	386	268	69,43	3020	33	302,965
13	M	276	244	88,406	5284	32	340,489
14	N	366	306	83,607	4170	32	332,622
15	O	146	70	47,945	1527	32	88,414

Рис. 3. Позиции ученых в рейтингах по h - и w -индексам

2.3. Использование различных оснований при формировании индекса цитирования

Исследования показали, что в качестве основания при формировании итерационного показателя может быть использован любой другой показатель — аналог h -индекса [25]. Обозначим основание w -индекса через β и внесем его в обозначение индекса — $w(\beta)$ -индекс.

В качестве основания берется любое число β , для которого может быть сформирован квадрат $\beta \times \beta$ или прямоугольник $\beta \times \alpha\beta$, где $\alpha\beta$ — число цитирований β -ой работы в списке работ, ранжированном в невозрастающем порядке по числу цитирований (α — любое число, $\alpha \geq 1$). Другими словами, величина β определяет некоторый базовый показатель, который используется при формировании последовательности чисел $w^1, w^2, \dots, w^k$. Сравним $w(\beta)$ -индексы с различными основаниями на примере выбора в качестве основания традиционных h -, g - и j -индексов ($w(h)$ -индекс рассмотрен в предыдущем параграфе).

Рассмотрим тот же пример, что и ранее.

На рисунках 4 и 5 показаны цепи для индексов с различными основаниями.

Длина цепи при подсчете $w(g)$ -индекса оказывается равной $k=19$. При этом прямоугольники располагаются достаточно часто, вследствие чего среднее

Рис. 4. Геометрическая интерпретация $w(g)$ -индекса Рис. 5. Геометрическая интерпретация $w(j)$ -индекса

число цитирований становится более близким к общему числу цитирований работ в каждой итерации. Сравните это с особенностью формирования $w(h)$ -индекса (рис. 2). Указанное обстоятельство приводит к существенному увеличению итогового значения показателя. Для рассматриваемого примера он равен $w(g)=100,833$, что в 2,517 раза больше значения показателя $w(h)=40,066$.

Выбор в качестве основания числа j из j -индекса приводит к другому результату. Длина цепи уменьшается до величины $k=14$. Соответственно, малые длины j -прямоугольников оказываются большими, чем малые длины g -прямоугольников. Усреднение в каждой итерации оказывается более далеким, чем общее число цитирований работ в итерации. Значение $w(j)=68,85$. Это в 1,718 раза больше, чем значение $w(h)$ -индекса, и в 1,465 раза меньше, чем значение $w(g)$ -индекса для рассматриваемого примера. Общие закономерности, однако, сохраняются вне зависимости от того, какое основание установлено при формировании показателя. В таблице 6 приведены результаты подсчетов для экспериментальных данных из базы РИНЦ. Отметим тот факт, что при использовании различных оснований в процессе подсчета $w(\beta)$ -индекса характер изменения позиций в рейтинге сохраняется (рис. 6). По величине для одного ученого $w(h)$ -индекс будет иметь всегда меньшее значение, чем $w(j)$ -индекс и $w(g)$ -индекс; последний имеет самую большую величину среди трех показателей.

Таблица 6. $w(\beta)$ -индексы для рассмотренных ученых

Позиция при ранжировании по h -индексу	Ученый	Q	N	h	j	g	$w(h)$	$w(j)$	$w(g)$
1	A	329	15246	58	28	16	509,996	1117,957	1841,744
2	B	240	6415	51	21	11	309,88	642,52	1074,976
3	C	301	4791	45	19	10	352,828	682,668	1030,401
4	D	387	8105	42	18	10	491,12	1027,014	1580,698
5	E	140	5579	41	17	10	222,192	524,398	859,386
6	F	194	2846	41	12	6	212,866	459,776	721,017
7	G	364	4332	37	14	7	400,589	790,199	1184,658
8	H	488	4228	35	12	7	508,634	999,461	1279,456
9	I	108	1817	35	12	7	126,147	270,167	412,124
10	J	198	2358	34	11	6	212,166	437,275	683,203
11	K	252	6024	33	17	11	379,921	718,701	1066,92
12	L	268	3020	33	13	7	302,965	591,535	873,281
13	M	244	5284	32	17	10	340,489	680,175	1068,975
14	N	306	4170	32	12	7	332,622	762,167	1190,751
15	O	70	1527	32	12	7	88,414	198,919	320,29

Рис. 6. Позиции ученых в рейтингах по $w(\beta)$ -индексам

2.4. Настраиваемые показатели учета цитируемости

Вернемся к основному недостатку $w(\beta)$ -индексов — большему влиянию малоцитируемых работ на значение итоговой суммы. Точнее этот фактор можно отметить как более близкое значение усреднения к числу цитирований работ в итерации с малым значением основания. С целью уменьшения этого влияния могут быть использованы различные «настройки» $w(\beta)$ -индексов, связанные с ограничением числа итераций при подсчете суммарного значения w или же с выбором в качестве оснований чисел $\alpha = a\beta$, где a — произвольное целое число. Ограничение числа итераций дает возможность учитывать только первые k итерации или же вообще отбрасывать те работы, которые процитированы определенное число раз. Аналогично способ, связанный с выбором числа α , нацело делящимся на β , дает возможность отбрасывания малоцитируемых работ. Допустим, пусть $\beta = h$, а $\alpha = 2h$, тогда на каждой итерации будут выделяться прямоугольники размером $h \times 2h$ и в случае $h = 1$ нельзя будет выделить прямоугольники для цитируемых единожды работ.

На рисунке 7 приведен график изменения значений $w(\beta)$ -индексов с различными основаниями на каждом шаге итерации для примера из таблицы 1. Пред-

Рис. 7. Рост значений $w(\beta)$ -индексов на каждой итерации подсчета

ставленные закономерности в длинах цепей с каждым из оснований и в значениях $w(\beta)$ -индексов являются общими и присущи подавляющему большинству статистических рядов из цитирований работ реальных ученых.

На определенной итерации итоговое значение суммы изменяется не столь существенно, чем в самом начале. Например, в таблице 7 приведены все итерации подсчета $w(\beta)$ -индексов при $\beta=h, j$ и g . При $k=5$ $w(h)=33,066$ (82,529 % от итогового значения), $w(j)=43,85$ (63,689 % от итогового значения), $w(g)=53,333$ (52,892 % от итогового значения). Останов работы алгоритма подсчета можно инициировать при достижении какого-либо порогового значения доли учитываемых цитирований. Например, 90 % цитирований всех работ. Для рассматриваемого примера 90-процентная граница от общего числа цитирований преодолевается для $w(h)$ при $k=6$, для $w(j)$ при $k=10$, для $w(g)$ при $k=14$.

Введем в рассмотрение «настраиваемые» $w(\beta)$ -индексы.

Обозначим через $w'(\beta)$ -индекс такой индекс, который получается при $k=t$ итерации формирования итоговой суммы.

Обозначим через $w^{-r}(\beta)$ -индекс такой индекс, который не учитывает все работы, цитируемые не более r раз. Например, w^{-1} -индекс не учитывает все цитирования работ, которые процитированы только по одному разу.

Таблица 7. Итерации подсчета $w(\beta)$ -индексов для рассматриваемого примера

k	$w(h)$	$w(j)$	$w(g)$
1	13	14,6	15,333
2	20,333	24,35	28,333
3	25,733	32,017	37,833
4	30,066	38,35	45,833
5	33,066	43,85	53,333
6	36,066	48,85	59,833
7	38,066	53,35	65,833
8	39,066	56,85	70,833
9	40,066	59,85	75,833
10	—	62,85	79,833
11	—	64,85	82,833
12	—	66,85	85,833
13	—	67,85	88,833
14	—	68,85	91,833
15	—	—	94,833
16	—	—	96,833
17	—	—	98,833
18	—	—	99,833
19	—	—	100,833

В таблицах 8 и 9 приведены результаты подсчетов $w^l(\beta)$ и $w^{-r}(\beta)$ -индексов для рассматриваемой выборки при $\beta=h$. В столбцах N , N^l и N^{-r} — общее число цитирований, число цитирований, охваченных $w^l(\beta)$ и $w^{-r}(\beta)$ -индексами. В столбце η приведена доля учитываемых цитирований h -индексом, а в столбцах υ^l и υ^{-r} — доля учитываемых цитирований $w^l(\beta)$ и $w^{-r}(\beta)$ -индексами. Если в среднем h -индекс учитывает только 36,733 % цитирований всех работ ученого для рассматриваемой выборки, то $w^1(h)$ -индекс учитывает в среднем 67,323 % цитирований всех работ. Для $w^5(h)$ -индекса и $w^{10}(h)$ -индекса данный показатель еще выше: получены средние значения в 87,566 % и 92,533 % всех ссылок на работы ученых соответственно. Для $w^{-r}(h)$ -индексов получены такие результаты: w^{-1} -индекс учитывает 98,666 % цитирований всех работ, w^{-2} -индекс — 97,181 %, w^{-3} -индекс — 95,512 %. Это как раз и показывает, что малоцитируемые работы не столь значительно изменяют общую сумму и могут не рассматриваться в конкретных «настройках».

На рисунках 8 и 9 приведены в виде графика доли учитываемых цитирований от общего их числа для $w^l(\beta)$ и $w^{-r}(\beta)$ -индексов.

Таблица 8. $w'(h)$ -индексы со значениями $t = 1, 5, 10$

Позиция при ранжи- ровании по h -индексу	Уче- ный	N	h -ин- декс	η	$w(h)$ - индекс	N^1	N^5	N^{10}	v^1	v^5	v^{10}	w^1 -ин- декс	w^5 -ин- декс	w^{10} -ин- декс
1	A	15246	58	22,065	509,996	12534	14435	14796	82,212	94,681	97,048	216,103	307,032	350,13
2	B	6415	51	40,546	309,878	5337	6025	6181	83,196	93,92	96,352	104,647	160,045	188,712
3	C	4785	45	42,32	352,828	3776	4235	4384	78,913	88,506	91,62	83,911	130,161	158,661
4	D	8093	42	21,797	470,121	5377	7026	7445	66,44	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	36,06	302,965	1819	2502	2686	60,232	82,848	88,94	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	70,6	85,372	44,938	125,439	181,494
15	O	1527	32	67,06	88,414	1360	1500	1516	89,064	98,232	99,28	42,5	67,914	77,414
Средние значения				36,733	—	—	—	—	67,323	87,566	92,533	—	—	—

Таблица 9. $w^r(h)$ -индексы со значениями $r = 1, 2, 3$

Позиция при ранжи- ровании по h -индексу	N	h -ин- декс	η	$w(h)$ - индекс	N^{-1}	N^{-2}	N^{-3}	v^{-1}	v^{-2}	v^{-3}	w^{-1} -ин- декс	w^{-2} -ин- декс	w^{-3} -ин- декс
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	42,32	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	36,06	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	67,06	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	—	—	—

Рис. 8. Доли учитываемых цитирований при использовании $w^1(h)$ -, $w^5(h)$ -, $w^{10}(h)$ -индексов

Рис. 9. Доли учитываемых цитирований при использовании $w^1(h)$ -, $w^2(h)$ -, $w^3(h)$ -индексов

2.5. Модификация w -индекса с переменным значением основания

Расчеты показывают, что выбор числа β носит несколько условный характер и не обязательно осуществлять привязку именно к какому-либо известному показателю. Достаточно определить принцип выбора конкретного числа работ из полного множества ранжированных в невозрастающем порядке. Приведем один из таких способов.

Алгоритм 3. Правила подсчета $w(\beta)$ -индекса с переменным основанием:

1. Работы ученого q_i ранжируются в невозрастающем порядке по числу их цитирований N_i .

2. Определяется показатель N_q , характеризующий среднее число цитирований в пересчете на каждую работу:

$$N_q = \frac{N}{Q}, \quad (5)$$

где N — общее число цитирований всех Q работ ученого.

3. Устанавливается значение $i=1$.

4. Для первой в списке работы определяется показатель:

$$\beta_i = \frac{N_1}{N_q}, \quad (6)$$

где запись ... обозначает целое сверху от вычисляемого значения.

5. Определяется суммарное число цитирований β_i работ ученого:

$$s_i = \frac{\sum_{j=1}^{\beta_i} N_j}{\beta_i}. \quad (7)$$

6. Определяется число w^i :

$$w^i = \frac{s_i}{\beta_i}. \quad (8)$$

7. Первые β_i работ ученого удаляются из списка.

8. Проверяется, есть ли в списке хотя бы одна цитируемая работа. Если да, то осуществляется переход к п. 9, если нет — то к п. 10.

9. Вычисляется $i:=i+1$ и осуществляется переход к п. 2.

10. Подсчитывается сумма:

$$w = \sum_{i=1}^k w^i, \quad (9)$$

где k — число итераций формирования чисел β_i .

На этапе вычисления по формуле (5) среднего числа цитирований в пересчете на одну работу определяется число N_q , которое в дальнейшем определяет и значение основания β на соответствующем шаге подсчета. Если имеются работы с большим числом цитирований, резко отличающимся по величине от числа цитирований остальных работ, то по формуле (6) вычисляется большое по величине основание β (далее будет показан такой пример). Это приводит к выделению большого числа работ и усреднения между ними всех цитирований.

Получим число w для заданного в таблице 10 и на рисунке 9 примера. В данном случае нет всплесков числа цитирований (последовательно убывают с шагом в одно цитирование или остаются на том же уровне). Это, однако, не всегда так. Рассмотрим пример, приведенный на рисунке 11 и описанный в таблице 11. В данном случае основания изменяются в зависимости от итерации, что связано с округлением величины β_i по формуле (6). Как итог, на первом шаге итерации существенное число цитирований (107, что составляет примерно 77 % всех цитирований работ) распределяется между первыми девятью работами. Что определяется именно большим числом цитирований (можно сказать, всплеском) первой работы в списке.

Таблица 10. Подсчет $w(\beta)$ -индекса с переменным основанием (первый случай)

k	β_i	s_i	w^i
1	3	46	15,333
2	3	39	13
3	3	27	9
4	2	16	8
5	2	14	7
6	2	12	6
7	2	11	5,5
8	2	10	5
9	2	9	4,5
10	2	7	3,5
11	2	6	3
12	2	6	3
13	2	4	2
14	2	2	1
Сумма		209	85,833

Рис. 10. Геометрическая интерпретация w -индекса с переменным основанием (первый случай)

Таблица 11. Подсчет $w(\beta)$ -индекса с переменным основанием (второй случай)

k	β_i	s_i	w^i
1	9	107	11,889
2	2	6	3
3	3	9	3
4	2	2	1
5	2	5	2,5
6	3	2	0,667
7	2	2	1
8	2	2	1
9	2	2	1
10	3	2	0,667
Сумма		139	25,723

Представленный способ формирования наукометрического показателя более сложный и менее наглядный, чем иные. Тем не менее, можно формировать $w(\beta)$ -индекс с переменным основанием и оценивать эффективность работы ученых именно с помощью такого индекса.

Рис. 11. Геометрическая интерпретация w-индекса с переменным основанием при изменении их значений (второй случай)

Заключение

Широкое использование наукометрических показателей, основанных на цитируемости работ ученых, обусловлено простотой их подсчета и интерпретации. Однако большинство этих показателей учитывает только одну какую-нибудь особенность цитируемых работ и, как правило, выделяет наиболее значимые работы среди всего их многообразия. Существуют, конечно, и показатели, учитывающие полное множество цитирований всех работ ученого, но они тоже несовершенны и не принимают во внимание особенности распределения цитируемости.

Предложенные в статье способы формирования нового наукометрического показателя, основанного на цитируемости работ ученого, позволяют более точно оценивать его вклад в развитие соответствующей предметной области. Есть возможность настройки показателя с учетом специфики предметной области. В относительно узких областях науки, например, автоматики и вычислительной техники (где количество ученых существенно ниже, чем в таких областях, как биология и физика), существенным является число малоцитируемых работ. А в более широких областях малоцитируемые работы могут вообще не рассматриваться, поскольку весомого вклада не дают. Итерационный принцип подсчета индекса цитируемости позволяет учесть все эти особенности для различных областей науки.

Несомненное достоинство w -индекса и его модификаций в том, что он представляет собой единственное число и имеет вполне понятную и простую геометрическую интерпретацию. Недостатком является более сильное влияние малоцитируемых работ на общее значение показателя, чем высокоцитируемых. Но, как показывают исследования, w -индекс и его модификации более точно учитывают вклад ученого в свою предметную область, чем известные h -, j - и g -индексы, ориентированные на высокоцитируемые и наиболее популярные работы ученых. При этом частично решается проблема «накрутки» показателей, т. к. усреднение значений при итерационном подсчете позволяет учесть «аномальные» перепады в числе цитируемостей, прослеживающиеся при явных «накрутках», самоцитировании и цитировании соавторами. К сожалению, полностью проблему «накрутки» предложенные показатели не решают.

Принципы формирования w -индекса и его различных модификаций могут быть использованы и совместно с иными способами оценки деятельности ученых, такими как учет соавторов, уровень периодического издания и пр.

Новый принцип формирования наукометрического показателя, основанного на учете цитируемости работ ученого, позволяет дополнить известные показатели и в некотором смысле даже заменить их.

Библиографический список

1. Ball P. Index Aims for Fair Ranking of Scientists//Nature. – 2005. – Vol. 436. – P. 900. – doi.org/10.1038/436900a.
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. 16569–16572. – DOI: 10.1073/pnas.0507655102.
3. Jones T., Huggett S., Kamalski J. Finding a Way through the Scientific Literature: Indexes and Measures//World Neurosurgery. – 2011. – Vol. 76. – Issue 1–2. – Pp. 36–38. – doi.org/10.1016/j.wneu.2011.01.015.
4. Михайлов О. В. Цитирование и цитируемость в науке. – М.: ЛЕНАНД. – 2017. – 208 с.
5. Орлов А. И. Число цитирований – ключевой показатель результативности в фундаментальной науке//Россия: тенденции и перспективы развития. Ежегодник. Институт научной информации по общественным наукам Российской академии наук. 2018. С. 861–867.
6. Михайлов О. В. О возможной модификации индексов Хирша и Эгга с учетом соавторства//Социология науки и технологий. – 2014. – Т. 5. – № 3. – С. 48–56.
7. Марвин С. В. Альтернативная дробная модификация индекса Хирша, учитывающая количество авторов цитируемых статей//Управление большими системами. – 2015. – № 56. – С. 108–122.
8. Марвин С. В. Нормированная доленая цитируемость как универсальная характеристика научной публикации//Социология науки и технологий. – 2016. – № 1. – С. 95–108.
9. 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. – Issue 3. – Pp. 1107–1127. – DOI: 10.1007/s11192-017-2633-2.
10. Leydesdorff L., Bornmann L., Opthof T. h_x : the Scientist as Chimpanzee or Bonobo//Scientometrics. – 2019. – Vol. 118. – Pp. 1163–1166. – DOI: 10.1007/s11192-018-2994-1.

11. Hirsch J. E. An Index to Quantify an Individual's Scientific Research Output That Takes into Account the Effect of Multiple Coauthorship//*Scientometrics*. – 2010. – Vol. 85. – Issue 3. – Pp. 741–754. – DOI: 10.1007/s11192-010-0193-9.
12. Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости//*Управление большими системами*. – 2013. – № 44. – С. 248–261.
13. Штовба С. Д., Штовба Е. В. Обзор наукометрических показателей для оценки публикационной деятельности ученого//*Управление большими системами*. – 2013. – № 44. – С. 262–278.
14. Ramana K. V., Kandi S. A Newer Approach to Assessing a Researcher's Scientific Impact: Ramana's Researcher Index (r-Index)//*Biomedicine and Biotechnology*. – 2013. – Vol. 1. – Issue 2. – Pp. 9–10. – DOI: 10.12691/bb-1-2-2.
15. Полянин А. Д. Недостатки индексов цитируемости и Хирша и использование других наукометрических показателей//*Математическое моделирование и численные методы*. – 2014. – № 1. – С. 131–144.
16. Михайлов О. В. Парадоксы цитирования//*Вестник Российской академии наук*. – 2018. – Т. 88. – № 3. – С. 268–271.
17. Hirsch J. E. An Index to Quantify an Individual's Scientific Leadership//*Scientometrics*. – 2019. – Vol. 118. – Issue 1. – Pp. 673–686. – DOI: 10.1007/s11192-018-2994-1.
18. Hauer M. P., Hofmann X. C. R., Krafft T. D., Zweig K. A. Quantitative analysis of automatic performance evaluation systems based on the h-index//*Scientometrics* (2020) 123:735–751. DOI: 10.1007/s11192-020-03407-7.
19. Герасименко П. В. Моделирование показателей результатов творческой деятельности ученого по его публикациям и их цитированиям//*Автоматика на транспорте*. – 2019. – Т. 5. – № 4. – С. 505–514. – DOI: 10.20295/2412-9186-2019-5-4-493-504.
20. Герасименко П. В. Модификации индекса Хирша для дифференцированной оценки результатов творческой деятельности ученых//*Управление наукой и наукометрия*. – 2020. – Т. 15. – № 1. – С. 55–71.
21. Egghe L. Theory and Practice of the g-Index//*Scientometrics*. – 2006. – Vol. 69. – Pp. 131–152.
22. Михайлов О. В. Новая версия индекса Хирша – j-индекс//*Вестник Российской Академии наук*. – 2014. – Т. 84. – № 6. – С. 532–535. – DOI: 10.7868/S0869587314060085.
23. Сапожников В. В., Сапожников Вл. В., Ефанов Д. В. «Настраиваемый» индекс оценки деятельности ученого//*Социология науки и технологий*. – 2020. – Т. 11. – № 2. – С. 194–210. – DOI: 10.24411/2079-0910-2020-12009.
24. Сапожников В. В., Сапожников Вл. В., Ефанов Д. В. Новый подход к расчету показателей деятельности ученых//*Автоматика на транспорте*. – 2019. – Т. 5. – № 4. – С. 505–514. – DOI: 10.20295/2412-9186-2019-5-4-505-514.
25. Сапожников В. В., Сапожников Вл. В., Ефанов Д. В. Обобщенный индекс цитирования//*Автоматика, связь, информатика*. – 2020. – № 8. – С. 42–45.

Valery V. Sapozhnikov, D. Sc.

Vladimir V. Sapozhnikov, D. Sc.

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

Dmitry V. Efanov, D. Sc.

*LLC STC "Integrated Monitoring Systems", St. Petersburg
Graduate School of Transport,
Institute of Mechanical Engineering, Materials and Transport
Peter the Great St. Petersburg Polytechnic University, St. Petersburg
Department "Automation, Remote Control and Communication on Railway Transport",
Russian University of Transport, Moscow*

ITERATIVE METHODS OF SCIENTIFIC INDICATORS FORMATION WITH CONSIDERATION TO THE FULL SET OF CITED WORKS

The widely known scientometric indicators based on the number of citations of the scientist's works are analyzed. Some methods of forming a new scientometric indicator are proposed, taking into account the full set of citations of the scientist's works. These methods have a simple and understandable geometric interpretation and give the values of indicators in the form of a single number, which is convenient from the standpoint of comparing indicators of different scientists and forming ratings.

The efficiency of the use of the iterative principle of the citation index formation is demonstrated with a certain number of the first works in the list of the scientist's works ranked in the non-increasing order of the number of citations, the averaging of the number of citations, the exclusion of these works from consideration and the repetition of all procedures with a given method of highlighting the first works at each iteration step are distinguished. The method of selecting the first works from the list allows you to select the base of the chain of intermediate indicators.

In the work of the authors, it is proposed as a basis to take the Hirsch index at each iteration step (to select the Hirsch base) and build a chain of decreasing indices with averaging the number of citations at each iteration. This approach makes it possible to more fully take into account the scientist's contribution to the development of the corresponding subject area. Another number can be chosen as the basis, for example, the determination of the Egge or Mikhailov index can be carried out. The basis can also be variable – such cases are considered in the article.

The calculations presented in the article are supplemented by experimental results obtained using statistical information from the Russian Science Citation Index database. The proposed iterative method for the formation of the citation index allows the "adjustment" of the index and, if necessary, and lets not take into account low-cited works or take into account only a certain percentage of citations from their total number. This may be relevant for specific subject areas of science and scientific or educational organizations.

Assessment of the activities of scientists; scientometric indicators; accounting for the number of citations; Hirsch index; Egge index; Mikhailov index

DOI: 10.20295/2412-9186-2021-7-1-137-166

References

1. Ball P. (2005) Index Aims for Fair Ranking of Scientists. *Nature*, vol. 436, p. 900.
2. 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.

3. Jones T., Huggett S., Kamalski J. (2011) Finding a Way through the Scientific Literature: Indexes and Measures. *World Neurosurgery*, vol. 76, iss. 1–2, pp. 36–38.
4. Mikhaylov O. V. (2017) *Tsitirovaniye i tsitiruyemost' v nauke*. [Citation and citation rate in science]. Moscow, LENAND Publ., 208 p. (In Russian)
5. Orlov A. I. (2018) Chislo tsitirovaniy – klyuchevoy pokazatel' rezul'tativnosti v fundamental'noy nauke [The number of citations is a key performance indicator in fundamental science]. *Rossiia: tendentsii i perspektivy razvitiya. Yezhegodnik. Institut nauchnoy informatsii po obshchestvennym naukam Rossiyskoy akademii nauk*. [Russia: Trends and Development Prospects. Yearbook. Institute for Scientific Information on Social Sciences of the Russian Academy of Sciences], pp. 861–867. (In Russian)
6. Mikhaylov O. V. (2014) O vozmozhnoy modifikatsii indeksov Khirsha i Egga s uchetom soavtorstva [On a possible modification of the Hirsch and Egg indexes taking into account co-authorship]. *Sotsiologiya nauki i tekhnologii*. [Sociology of Science and Technology], vol. 5, no. 3, pp. 48–56. (In Russian)
7. Marvin S. V. (2015) Al'ternativnaya drobnaya modifikatsiya indeksa Khirsha, uchityvayushchaya kolichestvo avtorov tsitiruyemykh statey [Alternative fractional modification of the Hirsch index taking into account the number of authors of the cited articles]. *Upravleniye bol'shimi sistemami* [Management of large systems], no. 56, pp. 108–122. (In Russian)
8. Marvin S. V. (2016) Normirovannaya dolevaya tsitiruyemost' kak universal'naya kharakteristika nauchnoy publikatsii [Normalized fractional citation as a universal characteristic of a scientific publication]. *Sotsiologiya nauki i tekhnologii*. [Sociology of Science and Technology], no. 1, pp. 95–108. (In Russian)
9. 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.
10. Leydesdorff L., Bornmann L., Opthof T. (2019) ha: the Scientist as Chimpanzee or Bonobo. *Scientometrics*, vol. 118, pp. 1163–1166.
11. Hirsch J. E. (2010) An Index to Quantify an Individual's Scientific Research Output That Takes into Account the Effect of Multiple Coauthorship. *Scientometrics*, vol. 85, iss 3, pp. 741–754.
12. Tsyganov A. V. (2013) Kratkoye opisaniye naukometricheskikh pokazateley, osnovannykh na tsitiruyemosti [Brief description of scientometric indicators based on citation]. *Upravleniye bol'shimi sistemami* [Management of large systems], no. 44, pp. 248–261. (In Russian)
13. Shtovba S. D., Shtovba Ye. V. (2013) Obzor naukometricheskikh pokazateley dlya otsenki publikatsionnoy deyatel'nosti uchenogo [Review of scientometric indicators for assessing the publication activity of a scientist]. *Upravleniye bol'shimi sistemami* [Management of large systems], no. 44, pp. 262–278 (In Russian)
14. Ramana K. V., Kandi S. A. (2013) Newer Approach to Assessing a Researcher's Scientific Impact: Ramana's Researcher Index (r-Index). *Biomedicine and Biotechnology*, vol. 1, iss. 2, pp. 9–10.
15. Polyanin A. D. (2014) Nedostatki indeksov tsitiruyemosti i Khirsha i ispol'zovaniye drugikh naukometricheskikh pokazateley [Disadvantages of both citation indices and Hirsch and the use of other scientometric indicators]. *Matematicheskoye modelirovaniye i chislennyye metody* [Mathematical modeling and numerical methods], no. 1, pp. 131–144. (In Russian)
16. Mikhaylov O. V. (2018) Paradoksy tsitirovaniya [Paradoxes of citation]. *Vestnik Rossiyskoy akademii nauk* [Bulletin of the Russian Academy of Sciences], vol. 88, no. 3, pp. 268–271. (In Russian)
17. Hirsch J. E. (2019) An Index to Quantify an Individual's Scientific Leadership. *Scientometrics*, vol. 118, iss. 1, pp. 673–686.

18. Hauer M. P., Hofmann X. C. R., Krafft T. D., Zweig K. A. (2020) Quantitative analysis of automatic performance evaluation systems based on the h-index. *Scientometrics*, 123:735–751.
19. Gerasimenko P. V. (2019) Modelirovaniye pokazateley rezul'tatov tvorcheskoy deyatel'nosti uchenogo po yego publikatsiyam i ikh tsitirovaniyam [Modeling of indicators of the results of a scientist's creative activity based on his publications and their citations]. *Avtomatika na transporte* [Transport automation], vol. 5, no. 4, pp. 505–514. (In Russian)
20. Gerasimenko P. V. (2020) Modifikatsii indeksa Khirsha dlya differentsirovannoy otsenki rezul'tatov tvorcheskoy deyatel'nosti uchenykh [Modifications of the Hirsch index for a differentiated assessment of the results of the creative activity of scientists]. *Upravleniye naukoj i naukometriya* [Science management and scientometrics], vol. 15, no. 1, pp. 55–71. (In Russian)
21. Egghe L. (2006) Theory and Practice of the g-Index. *Scientometrics*, vol. 69, pp. 131–152.
22. Mikhaylov O. V. (2014) Novaya versiya indeksa Khirsha – j-indeks [New version of the Hirsch index – j-index]. *Vestnik Rossiyskoy akademii nauk* [Bulletin of the Russian Academy of Sciences], vol. 84, no. 6, pp. 532–535. (In Russian)
23. Sapozhnikov V. V., Sapozhnikov V. I., Yefanov D. V. (2020) «Nastraivayemyy» indeks otsenki deyatel'nosti uchenogo [“Customizable” index for assessing the activities of a scientist]. *Sotsiologiya nauki i tekhnologii*. [Sociology of Science and Technology], vol. 11, no. 2, pp. 194–210. (In Russian)
24. Sapozhnikov V. V., Sapozhnikov V. I., Yefanov D. V. (2019) Novyy podkhod k raschetu pokazateley deyatel'nosti uchenykh [A new approach to calculating the performance indicators of scientists]. *Avtomatika na transporte* [Transport automation], vol. 5, no. 4, pp. 505–514. (In Russian)
25. Sapozhnikov V. V., Sapozhnikov V. I., Yefanov D. V. (2020) Obobshchennyy indeks tsitirovaniya [Generalized citation index]. *Avtomatika, svyaz', informatika* [Automation, communication, informatics], no. 8, pp. 42–45. (In Russian)

*Статья представлена к публикации членом редколлегии С. А. Никищенковым
Поступила в редакцию 05.12.2020, принята к публикации 23.12.2020*

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, заместитель генерального директора по научно-исследовательской работе ООО НТЦ «Комплексные системы мониторинга», профессор Высшей школы транспорта Ин-

ститута машиностроения, материалов и транспорта Санкт-Петербургского политехнического университета Петра Великого (СПбПУ Петра Великого), профессор кафедры «Автоматика, телемеханика и связь на железнодорожном транспорте» Российского университета транспорта (МИИТ)
e-mail: TrES-4b@yandex.ru

© Сапожников Вал. В., Сапожников В. В., Ефанов Д. В., 2021