

НАУКОМЕТРИЯ. БИБЛИОМЕТРИЯ

УДК 001.811

А. И. Земсков

ГПНТБ России

О некоторых библиометрических индексах

Производительность научного труда может измеряться как в абсолютных, так и в относительных показателях. Автор излагает суть индекса Хирша и показывает процесс его вычисления. В качестве иллюстрации возможностей системы Российского индекса научного цитирования приведены персональные библиометрические показатели ряда авторов. Сделан вывод о возможности библиометрии оценить крупномасштабные тренды в науке.

Эта статья – журнальный вариант выступления автора под названием «Чемпионы Хирша – краткие комментарии; Столяров Ю. Н., Шрайберг Я. Л., Гиляревский Р. С.» на конференции «ЛИБКОН–2015».

Ключевые слова: цитирование, библиометрия, абсолютные показатели, относительные показатели, индекс Хирша, Декларация Сан-Франциско по оценке научных исследований, Российский индекс научного цитирования.

UDC 001.811

Andrey Zemskov

Russian National Public Library for Science and Technology, Moscow, Russia

On several indices of bibliometry

The research productivity could be assessed both by absolute and relative metrics. The author explains the essence of the Hirsh index and demonstrates how it is calculated. The author refers to the Russian Citation Index database to demonstrate the personal bibliometric indices of the number of the well known in the library community authors, so one could compare his/her own evaluations with the bibliometric data. In conclusion, the bibliometry potential to evaluate large scale scientific trends is revealed.

Keywords: citation, bibliometry, absolute indices, relative indices, Hirsh index, San Francisco Declaration on Research Assessment (DORA), Russian Citation Index.

Citation data technique is based on the assumption that for sufficiently large amount of citations of the article the real motivations, would they be positive or negative, are mutually neutralized, and finally their number reflects an impact of the given article. The absolute indicators include: the total number of publications of the author, or employees of the organization for all time, for the year or for any given period of time (publication activity); number of articles in periodicals for the year or for a certain period of time; number of citations for all time or for any period of time (this period is often referred to as "citation window"); amount can refer either to any particular journal or some selected group. As for the relative indicators, there are: the impact factor of the journal and citation indexes. The impact factor of the journal of this year is the ratio of the number of citations (in a predetermined collection of journals) in a given year to the number of publications in this journal for a certain period. The evaluation of the quality of the article, based on the impact factor of the journal is quite risky. Careless use of the journal impact factor has caused some friction in the scientific community. At 2012 the American Society of cell biology has adopted the San Francisco Declaration. The impact factor should not serve as a single evaluation method for employment, promotion or making decisions about grants. The Hirsch index, proposed in 2005 is an alternative to the classical citation index. The criterion is based on the account number of researchers publications and number of citations of these publications. There are presented bibliometric data for some librarians and authors of scientific papers. The unique contribution of bibliometry in the process of scientific communication and its great importance is the fact that it gives you the big picture, which is very important for the administrators and the general public. The use of bibliometry can positively affect the referencing in several aspects: its analysis revealed the laws, which are invisible for a single scientist. Combining methods of expert evaluation with bibliometric methods is better protected from the influence of nepotism.

Вопрос о том, в каком контексте применяются методы цитирования в ходе процесса научных коммуникаций – в положительном (чтобы похвалить) или в негативном (чтобы обругать), – не изучен.

В настоящее время использование данных цитирования базируется на том предположении, что при достаточно большом количестве найденных ссылок различные мотивации, служащие побудительным стимулом для цитирования той или иной статьи, взаимно нейтрализуются, и то, что остаётся, – это сам факт внимания к статье.

Мы также знаем, что в различных отраслях науки сложились очень разные обычаи цитирования. Статьи о результатах основных биомедицин-

ских исследований цитируются в шесть раз чаще, чем статьи по математике. Это нужно обязательно учитывать при сравнении учреждений, работающих в различных областях науки. Пока же не существует общепринятой теории причин, вызывающих цитирование той или иной работы.

В библиометрии используются как абсолютные, так и относительные показатели (индексы).

К основным абсолютным показателям относятся:

полное количество публикаций данного автора или сотрудников данной организации за всё время, за год или за какой-либо иной период времени (публикационная активность);

количество статей в периодическом издании за год или за иной период времени;

количество цитирований работ данного автора или сотрудников данной организации за всё время или за какой-либо период времени (этот период часто называют «окно цитирования»); количество может либо относиться к конкретному журналу, либо к отобранной группе журналов (например, наиболее авторитетных), либо учитываться в целом.

Эффективность научных исследований измеряется, в частности, числом публикаций. Публикации при оценке науки считают, как правило, по базам данных *Web of Science* и/или *Scopus*. Число публикаций прямо пропорционально зависит от вложенных в науку средств, а также зависит от объёма потребляемой научной информации. Основная форма научной информации – электронная.

Из многих относительных показателей назовём лишь импакт-фактор журнала и индексы цитирования.

Импакт-фактор журнала по состоянию на текущий год – это отношение числа цитирований (в заранее установленном круге журналов) за данный год к числу публикаций в этом журнале за определённый период. Например, в системе РИНЦ импакт-фактор рассчитывается на основе данных по цитированию журнала в массиве публикаций, входящих в РИНЦ, за предыдущие два года (или пять лет). При этом данные берутся из публикаций того года, для которого рассчитывается импакт-фактор.

При расчёте импакт-фактора число ссылок, сделанных в расчётном году из всех обрабатываемых в РИНЦ журналов, на те статьи, которые опубликованы в данном журнале за предыдущие два года (или пять лет), делится на общее число этих статей. Полученный показатель отражает среднее число цитирований одной статьи в журнале. Например, при расчёте пятилетнего импакт-фактора за 2013 г. суммарное число ссылок, сделанных в 2013 г. на статьи, опубликованные в журнале в период с 2008 по 2012 г. включительно, делится на общее число статей, опубликованных в выпусках журнала за эти годы.

Рост публикаций в мире по одному из продуктов библиометрии – журнальному импакт-фактору – опережает темпы увеличения числа публикаций в целом. Подходы к оценке качества научной работы на основе журнального импакт-фактора (*JIF* – подсчёт среднего количества цитирований статьи) сейчас широко распространены, однако зачастую оказываются в тисках неправильного использования, в особенности при оценках отдельной статьи. Вызывает беспокойство тот факт, что технология *JIF* является коммерческим секретом, и то, что вполне возможна значительная доля мошенничества.

Конечно, оценивать качество статьи на основе импакт-фактора журнала, в котором статья опубликована, довольно рискованно. (Это всё равно, что считать человека хорошим только потому, что он живёт в хорошей квартире.)

Следует отметить, что слишком необдуманное использование импакт-фактора журнала вызвало определённые трения в научном сообществе. На ежегодной встрече Американского общества клеточной биологии (*American Society for Cell Biology*) в декабре 2012 г. была принята Декларация Сан-Франциско по оценке научных исследований (*San Francisco Declaration on Research Assessment – DORA*), которую к настоящему времени подписали более 4 тыс. учёных и 160 научных организаций. Цель этой Декларации – прекратить практику корреляции величины импакт-фактора журнала с заслугами конкретного учёного, поскольку это ведёт к неточностям и искажениям научных исследований. Утверждается, что импакт-фактор не должен служить заменой «оценок качества отдельной научной статьи или учитываться при приёме на работу, продвижении по службе или принятии решений о грантах».

Индекс цитируемости – это отношение числа публикаций данного автора к суммарному количеству цитирований его работ.

Индекс Хирша (*h-индекс*) – это наукометрический показатель, предложенный в 2005 г. американским физиком Хорхе Хиршем (*Jorge Hirsch*) в качестве альтернативы классическому индексу цитируемости. Критерий основан на учёте числа публикаций исследователя и числа их цитирований. Индекс вычисляется исходя из распределения цитирований публикаций данного исследователя. Обычно распределение количества публикации $N(q)$ в зависимости от числа их цитирований q в очень грубом приближении соответствует гиперболе: $N(q) \approx \text{const} \times q^{-1}$. Координата точки пересечения этой кривой с прямой $N(q) = q$ и будет равна *h-индексу*.

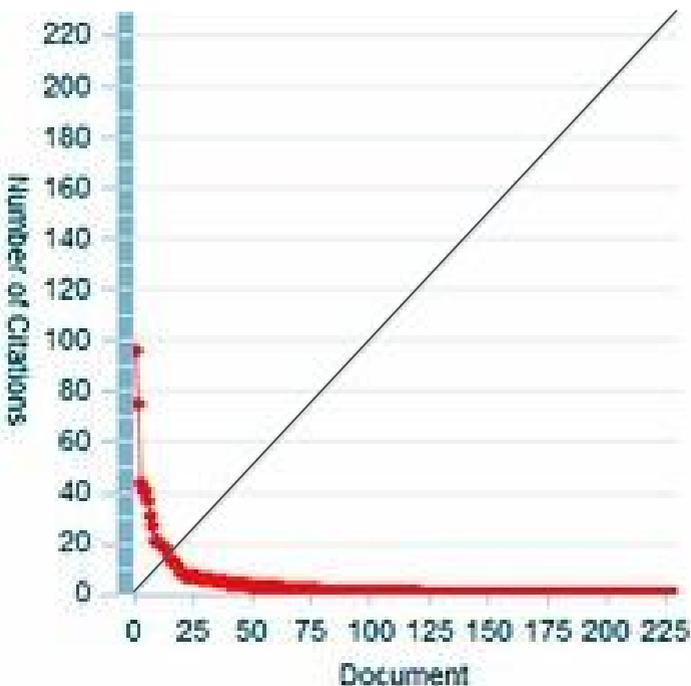


Рис. 1. Распределение цитируемости работ одного исследователя, опубликовавшего 230 статей и имеющего *h-индекс* 15 (данные системы *Scopus*); горизонтальная шкала – количество статей, вертикальная – количество её цитирований

Согласно Хиршу, учёный имеет индекс h , если h из его N_p статей цитируются как минимум h раз каждая, в то время как оставшиеся $(N_p - h)$ статьи цитируются не более чем h раз каждая.

Например, *h-индекс*, равный 10, означает, что учёным было опубликовано не менее 10 работ, каждая из которых была процитирована 10 и более раз. При этом количество работ, процитированных меньшее число раз, может быть любым. Аналогичным образом может быть рассчитан *h-индекс* для конкретного научного журнала, организации, страны.

Этот индекс хорошо работает лишь при сравнении учёных, занимающихся одной областью исследований, поскольку в разных отраслях науки связанные с цитированием традиции различаются: например, в биологии и медицине *h-индекс* намного выше, чем в физике.

Индекс Хирша в определённой степени показывает итоги творческой деятельности учёного (а не его активность в данное время), поскольку *h-индекс* является кумулятивным показателем и может только увеличиваться

от текущего значения. Это – полезный, но не абсолютный показатель, зависящий от научной дисциплины, стажа научной деятельности учёного и целевой аудитории.

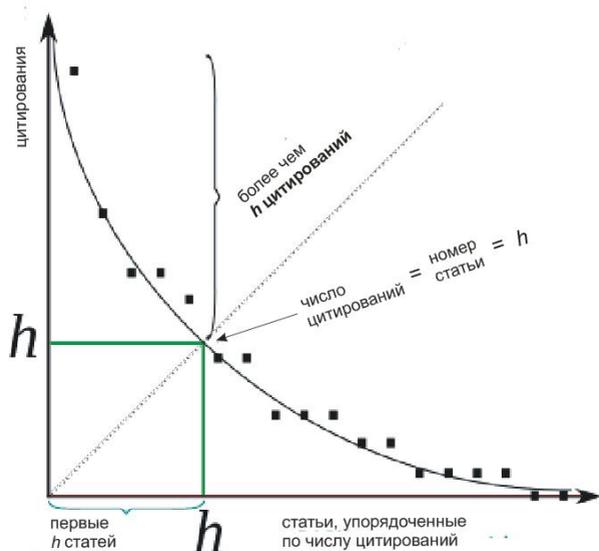


Рис 2. Схема расчёта индекса Хирша

Более детально процесс вычисления h -индекса представлен на рис. 2. Если отложить на горизонтальной шкале статьи в порядке убывания их цитирований, а на вертикальной – число цитирований, то h -индекс – первое целое число снизу от биссектрисы. В норме h -индекс автора, например учёного-физика, примерно равен продолжительности его научной карьеры.

Хорхе Хирш считает, что в физике (в реалиях США) h -индекс, равный 10–12, может служить одним из определяющих факторов для решения о предоставлении исследователю постоянного места работы в крупном исследовательском университете; уровень исследователя с h -индексом 15–20 соответствует членству в Американском физическом обществе; индекс 45 и выше может означать членство в Национальной академии наук США.

Индекс Хирша может вычисляться с использованием как бесплатных баз данных в интернете (*Google Scholar*, *eLibrary.ru*, *ADS NASA*), так и баз с платной подпиской (*Scopus* или *ISI Web of Science*). Заметим, что h -индекс, подсчитанный для одного и того же человека с использованием различных БД, будет различен – как и другие наукометрические характеристики, он зависит от области охвата выбранной БД.

Кроме того, *h-индекс* может подсчитываться с учётом и без учёта самоцитирования; предполагается, что отбрасывание ссылок авторов на собственные статьи даёт более объективные результаты.

Индекс Хирша, разумеется, не идеален. Нетрудно придумать ситуацию, когда он даёт совершенно неверную оценку значимости исследователя. В частности, короткая карьера учёного приводит к недооценке значимости его работ. Так, *h-индекс* французского математика Эвариста Галуа (*Evariste Galois*; 1811–1832), погибшего в юности и успевшего опубликовать лишь четыре статьи, равен 4 и останется таким навсегда. Если бы Альберт Эйнштейн прекратил деятельность в начале 1906 г., его *h-индекс* остановился бы на 4 или 5, несмотря на чрезвычайно высокую значимость статей, опубликованных им в 1905 г.

Существует неустранимый временной лаг (запаздывание) между датой написания статьи и датой её опубликования (от нескольких месяцев до года), а также запаздывание между датой публикации и началом цитирования. Разница между динамикой цитирования (плавный рост за 2–3 года до максимальных значений и довольно медленный спад) и темпом выгрузки (немедленное начало и быстрый спад интереса) очень характерна и говорит в пользу формирования сетевых архивов открытого доступа.

В таблице приведены библиометрические данные нескольких заметных в библиотечном сообществе авторов научных работ. Богатые возможности системы РИНЦ демонстрируют как сильные стороны, так и типичные недостатки методики библиометрического анализа, в том числе и индекса Хирша. Ясно, например, что авторитет и позитивное значение деятельности Е. Ю. Гениевой никоим образом не отражаются в формальных показателях. В то же время изучение подробной информации о собственной научной производительности и цитируемости, быть может, окажется полезным для некоторых читателей журнала.

РИНЦ. Сравнительные данные индексирования авторов
(на 10 ноября 2015 г.)

Автор	Число публикаций	Число цитирований	Среднее цитирование одной публикации	Индекс Хирша	Количество и процент самоцитирований	Год публикации наиболее цитируемой работы и число ссылок на нее
Соколов А. В.	206	1 703	8,26	15	100 (5,9)	2002 – 554
Литаревский Р. С.	90	1 045	11,61	13	28 (2,7)	1968 – 206; 1976 – 142; 1996 – 164
Писляков В. В.	38	610	16,0	10	42 (6,9)	2009 – 296
Шрайберг Я. Л.	126	696	5,52	9	43 (6,2)	2001 – 88; 2003 – 89; 2007 – 88
Егепов Б. С.	121	572	4,7	8	7 (1,2)	1976 – 93; 1980 – 98
Глухов В. А.	33	231	7,0	8	20 (8,7)	2000 – 60
Столяров Ю. Н.	154	723	4,7	8	98 (13,6)	2005 – 66
Катёнов Н. Е.	125	560	4,48	7	205 (36,6)	2007 – 57
Леонов В. П.	80	181	2,26	5	27 (14,9)	1986 – 20; 2000 – 20; 2003 – 20; 2005 – 20

Продолжение таблицы

Автор	Число публикаций	Число цитирований	Среднее цитирование одной публикации	Индекс Хирша	Количество и процент самоцитирований	Год публикации наиболее цитируемой работы и число ссылок на нее
Вислый А. И.	32	162	5,4	3	2 (1,2)	1995 – 95
Маркусова В. А.	105	640	6,1	10	99 (15,5)	2009 – 86
Земсков А. И.	100	375	3,75	8	13 (3,5)	2003 – 64, 2007 – 62
Цветкова В. А.	54	222	4,1	6	33 (19,5)	2006 – 58
Клюев В. В.	95	140	2,1	4	7 (5)	1999 – 27
Сухнасян Э. Р.	263	399	1,5	5	172 (43,1)	2004 – 60
Гениева Е. Ю.	12	28	2,33	2	0	2006 – 11
Гриханов Ю. А.	9	23	2,55	1	2 (8,7)	2001 – 5
Игуменова Н. П.	38	37	0,97	2	13 (35,1)	2006 – 10
Никонорова Е. В.	39	117	3,0	4	3 (2,6)	2005 – 18

Несмотря на определённые ограничения в использовании библиометрических индикаторов (часть из которых уже упоминалась), наличие проверенных источников исходной информации, доступность и простота использования обеспечивают широкое применение таких индикаторов, как импакт-фактор и связанный с ним Отчёт о журнальных цитированиях (*Journal Citation Reports, JCR*) – это БД, содержащая материалы по импакт-факторам и ранжировкам многих журналов по общественным наукам и наукам о жизни.

Родственный импакт-фактору – показатель *Eigenfactor*, в котором учитываются публикации с большим коэффициентом, цитируемые в высокоавторитетных журналах, и вычитается самоцитирование.

Ещё один индикатор – *SCImago Journal Rank (SJR indicator)* призван отражать научную значимость журнала; он учитывает как полное количество цитирований, так и авторитетность (престиж) источников цитирования. Алгоритм селекции разработан для использования в очень больших и разнородных массивах цитирования. Результатом является ранжировка по «средней значимости статьи».

Другие индексы, используемые в библиометрии

g-индекс также рассчитывается на основе распределения цитирований публикаций учёного, но по иному алгоритму: для данного множества статей, отсортированных в порядке убывания количества их цитирований.

При равном *h-индексе* величина *g-индекса* будет больше у того исследователя, у которого на самые цитируемые статьи приходится больше цитирований.

i-индекс используется для оценки публикационной активности научной организации наряду с *h-индексом*: научная организация имеет индекс *i*, если не менее *i* учёных имеют *h-индекс* не менее *i*. Таким образом, величина *i-индекса* зависит от количества наиболее высокоцитируемых исследователей, работающих в данной организации; *i-индекс* будет низким там, где высокий *h-индекс* обеспечивают публикации одного-двух соотрудников.

Выводы

Уникальный вклад библиометрии в процесс научного общения и её огромное значение состоит в том, что она создаёт картину в целом, а это весьма важно и для администраторов и для широкой публики. На многообразном полотне библиометрии отчётливо видны контуры зарождающихся явлений, например взаимное пересечение каких-либо научных дисциплин, пока ещё не оформленное институционально. Интерпретация этих связей, какие-либо неожиданные противоречия с устоявшимися представлениями – всё это должно изучаться экспертами или по крайней мере с участием специалистов.

Научное рецензирование (*peer review*) призвано вносить коррективы в общую картину, созданную библиометрией. Использование библиометрии может положительно сказаться на рецензировании в нескольких аспектах. Поскольку библиометрия опирается на обработку больших массивов данных, при их анализе выявляются такие закономерности, которые не видны одному учёному.

Ещё один очень важный аспект взаимодействия библиометрии и рецензирования состоит в том, что библиометрический анализ в какой-то мере может контролировать, проверять или, точнее, поверять, научное рецензирование. Быстрый спад интереса к какой-либо научной дисциплине, недавно весьма популярной и имеющей ранее созданную инфраструктуру, а значит, по инерции стремящейся поддержать былое могущество, не может надёжно фиксироваться системой научного рецензирования ввиду врождённой избирательности и уже израсходованного на неё ресурса. С этой точки зрения совместное аналитическое исследование методами и рецензирования, и библиометрии лучше защищено от влияния кумовства и студенческих привязанностей.

Библиометрические индикаторы поддерживают результаты рецензирования и могут скорректировать их в том случае, если индивидуальные оценки вступают в противоречие с агрегированными данными. С другой стороны, и библиометрические данные также могут быть подправлены суждениями рецензентов в том случае, если формальные алгоритмы почему-то не работают. Это объединение библиометрических методов с рецензированием, так называемое информированное рецензирование, может служить целям обеспечения прозрачности для широкой публики и в то же время сохранять все признаки научного подхода.

Общий вывод: измерения на базе цитирования оказываются действенными только при работе с большими массивами.

Индекс Хирша – полезный, но не абсолютный показатель, поскольку зависит от научной дисциплины, стажа научной деятельности и целевой аудитории.

Andrey Zemskov, Cand. Sc. (Physics and Mathematics), Associate Professor, leading researcher, Russian National Public Library for Science and Technology;

andzem@gpntb.ru

17, 3-rd Khoroshevskaya st., 123436 Moscow, Russia