

УДК 001.92

ББК 72

ОБЗОР НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ ДЛЯ ОЦЕНКИ ПУБЛИКАЦИОННОЙ ДЕЯТЕЛЬНОСТИ УЧЕНОГО

Штовба С.Д.¹, Штовба Е. В.²

*(Винницкий национальный технический университет,
Винница)*

Сегодня для оценивания результативности научной деятельности совместно с экспертными заключениями все чаще используют и наукометрические показатели. Проведен обзор основных наукометрических показателей, которые учитывают количество публикаций и количество цитирований как отдельно, так и совместно. Показаны способы учета дополнительной информации по количеству соавторов, по статусу журнала, продолжительности научной карьеры, договорным цитированиям и т.п. Рассмотрены “подводные камни” наукометрических показателей, связанных со скрытыми и неформальными ссылками, а также с ошибками в списке литературы.

Ключевые слова: наукометрия, индекс цитирования, индекс Хирша, скрытое цитирование, неформальное цитирование.

1. Введение

Сегодня для оценивания результативности научной деятельности совместно с экспертными заключениями все чаще используют и наукометрические показатели. Эти показатели основаны на количестве публикаций автора и на количестве

¹ Сергей Дмитриевич Штовба, доктор технических наук, профессор (shtovba@gmail.com, www.shtovba.vinnitsa.com).

² Елена Валериевна Штовба, кандидат экономических наук (olena.shtovba@yahoo.com).

ссылок на его работы. Возросший интерес к наукометрическим показателям вызван в первую очередь возможностью автоматизации процесса оценивания с использованием программных средств баз данных Web of Science, Scopus, Российской научной библиотеки (elibrary.ru). Кроме того, можно использовать бесплатные программы, например, Publish or Perish, работающих на данных поисковой системы научных публикаций Google Scholar. Дешевизна и быстрота проверки, а также отсутствие человеческого фактора, обуславливают популярность наукометрических показателей в экспресс-оценивании публикационной деятельности ученых. Пороговые ограничения по наукометрическим показателям представляют собой некий фильтр, который отсеивает слабых кандидатов и тем самым сокращает затраты на проведение дорогостоящего и трудоемкого экспертного оценивания качества научных результатов.

Кроме того, наукометрические показатели удобны для оценки фундаментальных исследований, результаты которых непосредственно не связаны с экономическим эффектом. Фундаментальные разработки направлены на развитие науки, поэтому их востребованность оценивают через отзыв научного сообщества на публикации с результатами исследований. Формально, этот отзыв выражают индексом цитирования – суммарным количеством публикаций, ссылающихся на рассматриваемые работы.

Известно, что как только какой-то показателей становится критерием принятия решений, придумываются способы его “накрутки”. Не исключение и классические наукометрические показатели – количество публикаций и индекс цитирования. Для их искусственного увеличения применяют дробления результатов для опубликования в нескольких статьях, опубликование одних и тех же результатов под разными названиями, публикации в нерейтинговых журналах, включение в число соавторов посторонних, самоцитирование и цитирование друзьями и т.п.

Целью статьи является обзор основных наукометрических показателей оценки публикационной активности ученого, которые фильтруют различные способы накруток количества публи-

каций и индекса цитирования. Кроме того, обращается внимание на недостатки наукометрических показателей, связанных: 1) с ошибками в списке литературы; 2) с забыванием имен классиков, когда авторы считают, что вклад предшественников настолько хорошо известен любому из этой области науки, что нет смысла об этом упоминать; 3) сокрытием первоисточников, т. е. включением в перечень цитируемой литературы не концептуальных работ, а их модификаций.

2. Показатели на основе количества публикаций

Сначала основным наукометрическим показателем было количество печатных работ ученого – суммарное или по отдельным типам: монографии, статьи, тезисы, публикации в изданиях, входящих в список ВАК, внесенных в электронные базы Web of Science, Scopus или eLibrary.ru, проиндексированных Google Scholar и т.п. Иногда учитывают объем публикаций, т.к. журнальная статья может занимать и 3 страницы, и 150. Часто по количеству публикаций устанавливается порог, превышение которого позволяет автору участвовать в некотором конкурсе или экспертизе. Например, с 1 сентября 2013 г. для защиты кандидатской диссертации в Украине необходимо опубликовать 4 статьи в национальных изданиях из специального списка и 1 статью в зарубежном журнале, а для докторской – 16 плюс 4. К конкурсу на гранты часто не допускаются проекты, авторы которых не опубликовали несколько статей в журналах из международных наукометрических баз.

Существуют и интегральные критерии, чаще всего в виде взвешенной суммы, когда баллы за публикацию определяются ее типом, например, за монографию начисляется 20 баллов, за статью в Scopus – 10 баллов, за тезисы – 1 балл. Другой вариант – учитывать статус издания. Для учета популярности издания баллы за публикацию взвешивают импакт-фактором журнала – коэффициентом, показывающей среднее число цитирования одной статьи журнала за определенный временной интервал, обычно за 2 года. Для учета престижности издания

баллы за публикацию умножают на взвешенный импакт-фактор журнала, который рассчитывается на основе алгоритма ранжирования веб-страниц – Google PageRank Algorithm. В взвешенном импакт-факторе учитывается репутация изданий, в которых цитируется рассматриваемый журнал [8].

Если считать только количество публикаций, то молодые ученые всегда будут проигрывать своим старшим коллегам. Поэтому, существуют относительные показатели, когда учитывают публикаций за определенный интервал времени, например, за последние 3 года. Другой вариант – разделить суммарное количество публикаций на научный стаж автора. По этому показателю первое место занимает советский химик Ю.Т. Стручков, который с 1981 по 1990 гг. опубликовал 948 научных работ, т.е каждые 4 дня писал 1 статью. За это достижение в 1992 г. ему присуждена Шнобелевская премия по литературе с намеком на то, что его просто вписывали в соавторы статей. На самом деле, Ю.Т. Стручков как раз свою часть этих статей и писал. Он “сидел на приборе” – создал в Институте органических соединений Академии наук лабораторию рентгеноструктурных исследований, поставив на поток свой метод определения кристаллических структур.

Среди современников выделим американского профессора Э. Тополя, который с 1980 г. опубликовал 1702 статьи¹, что соответствует средней производительности одна статья в неделю. Он тоже отмечен Шнобелевской премией по литературе в 1993 г., но не за количество публикаций, а за огромное число соавторов. В 1993 г. в журнале “New England Journal of Medicine” он и еще 976 соавторов опубликовали десятистраничную статью. Таким образом, на 1 страницу текста приходится около 100 соавторов. Сегодня, статьи с таким количеством соавторов не редкость, например, сотрудники Института физики высоких энергий из Протвино опубликовали статью, в которой 3185 соавторов, а кроме этого еще 80 работ, в каждой из которых

¹ По данным его профиля на Google Scholar.

более 3000 соавторов. Для выделения личного вклада, баллы за публикацию распределяют на всех соавторов. Обычно считается, что вклад всех соавторов равновеликий, поэтому баллы за публикацию делят на число соавторов.

Таким образом, в наукометрических показателях на основе количества публикаций может учитываться тип публикации, статус издания, объем работы и количество соавторов. Для искусственного увеличения количества публикаций используют такие типовые приемы, как дробление результатов для опубликования в разных изданиях, а также публикация почти идентичных статей под разными названиями. Поэтому, погоня за количеством публикаций часто снижает качество научных работ.

3. Показатели на основе количества цитирований

Индекс цитирования – это суммарное количество ссылок в научных публикациях на работы автора. Он отражает реакцию научного сообщества на публикации с результатами исследований. Как правило, плохие работы не цитируют, за исключением особых отношений между авторами.

Основные варианты обычного индекса цитирования имеют такие особенности:

- игнорируют самоцитирование или цитирование соавторами [3, 17], что существенно снижает рейтинг "ученого-затворника", публикации которого интересуют только его самого;
- игнорируют повторное цитирование одной работы одним и тем же ученым [16], что уменьшает влияние комплементарного, договорного цитирования по принципу "я – тебя, ты – меня";
- учитывают личный вклад ученого [12, 25], разделяя количество цитирований между соавторами;
- учитывают репутацию цитирующего издания, взвешивая количество ссылок в журнале на его импакт-фактор или другой аналогичный коэффициент [10, 26];

- учитывают интенсивность цитирований (среднюю длину списка литературы) в разных науках, которая, например, в биологии раз в 5 выше, чем в математике.

Кроме явных ссылок, указанных в списке литературы, существуют неформальное цитирование и скрытое цитирование. Неформальное цитирование состоит в указании источника информации в тексте работы без включения его в список литературы. Например, указание в тексте только фамилий и инициалов авторов предшествующих исследований или использование эпонимов, например, геометрия Лобачевского, распределение Вейбула – Гнеденко, принцип Беллмана – Заде и т.п. Часто используются термины без лингвистической связи с фамилией автора, например, метод наименьших квадратов, теория относительности, задача о Кёнигсберских мостах. При этом не упоминаются ни фамилии авторов – К. Гаусса, А. Эйнштейна, Л. Эйлера, ни названия соответствующих работ. Через 10–30 лет после публикаций статей-шедевров [2] на них все чаще начинают ссылаться неформально. Очевидно, что этот процесс ускоряет включение научных результатов в учебники и учебные пособия. Наукометрические исследования [22] показали, что в статьях, опубликованных в 2005 г., соотношение между неформальными и обычными ссылками составило:

- для эффекта Доплера – 1575;
- для функции Грина – 397;
- для уравнения Фоккера – Планка – 100;
- для Ч.В. Рамана – 16.5;
- для метода Вентцеля – Крамерса – Бриллюэна – 14.2;
- для М. Планка – 7.4;
- для теории функционала плотности – 6.4;
- для А. Эйнштейна – 5.8;
- для Э. Шрёдингера – 5.5.

Скрытое цитирование [4] состоит в использовании идей без прямой ссылки на ее автора, но с возможностью идентификации первоисточника через цепочку цитирований. В истории науки есть много примеров, когда концептуальные статьи цитируют

значительно реже, чем работы по их модификации. Впечатляющим примером, из приведенных в [1], являются статьи О. Фолина [15] и О. Лоури [21]. О. Лоури модифицировал реактив О. Фолина для колориметрического определения белка, добавив еще один компонент. Сегодня, по данным Google Scholar, статью О. Лоури процитировали 248 093 раз, тогда как работу О. Фолина – 1971. Следовательно, среди 248 093 работ, цитирующих статью [21] и, соответственно, использующих идеи О. Фолина, на которых она основана, почти нет таких, которые явно ссылаются на его концептуальную статью [15].

В [5] предложен наукометрический показатель оценки ученого с учетом скрытого цитирования. Он имеет две составляющие. Первая соответствует обычному индексу цитирования. Вторая рассчитывается по количеству скрытых цитирований. Добавив к обычному индексу цитирования с некоторым весом число скрытых ссылок, получим наукометрическую оценку ученого с учетом скрытой диффузии знаний. Аналогичные принципы оценивания персонала действуют в сетевом маркетинге, когда сотрудник получает баллы как за свои прямые продажи, так и за продажи, которые осуществила завербованная им команда. Здесь персональные продажи сотрудника соответствуют прямому цитированию, а продажи его команды – скрытому цитированию. Эти же принципы применяются в спорте для оценивания по системе "гол плюс пас" атакующих игроков команды. Практическая ценность предложенного в [5] показателя заключается в том, что он позволяет идентифицировать креативных ученых, на основе идей которых научное сообщество создало значительного количества востребованных (высокоцитированных) работ. При этом, сами генераторы идей остались в тени. За такими учеными часто "охотятся" лидерские исследовательские компании, но автоматически их обнаружить было достаточно сложно из-за низких наукометрических показателей. Теперь идентификацию скрытых зачинщиков научного мейнстрима можно формализовать, выявив ученых, которые одновременно имеют большое значение предложенного в [5] индекса

цитирования и малое значением традиционного индекса цитирования.

При принятии решений на основе индекса цитирований и его модификаций необходимо помнить о невозможности точно установить все источники информации, которые использованы при подготовке работы. Во-первых, автор включает в перечень литературы только наиболее релевантные источники, уровень использования которых превышает некоторый порог [1]. Во-вторых, в списке литературы часто встречаются ошибки, например, в медицинских журналах их частота составляет около 30% [7, 14].

4. Показатели на основе количества цитирований и количества публикаций

Чтобы выявить ученых, пишущих много и качественно, в 2005 г. физик Х. Хирш предложил новый показатель – индекс Хирша [18]. Индекс Хирша или h -индекс – это максимальное целое число h , указывающее, что автор опубликовал h статей, каждая из которых процитирована хотя бы h раз. Эти h статей составляют ядро Хирша или h -ядро. Чтобы попасть в ядро Хирша, статью должны процитировать хотя бы h раз. Чтобы получить высокий индекс Хирша надо писать много, при этом не дробя результаты по нескольким публикациям. Простота расчетов и нечувствительность к типовым приемам искусственного улучшения вышепересмотренных показателей мгновенно сделали индекс Хирша популярным наукометрическим индикатором.

Недостатки индекса Хирша связаны с тем, что в нем не учитываются: 1) насколько превышен порог цитирований в ядре Хирша; 2) длина “хвоста”, т.е. количество публикаций, не вошедших в ядро и уровень их цитирования. Для компенсации этих недостатков предложены более 30 модификаций индекса Хирша, основные из которых сведены в табл. 1. Формулы этой таблицы написаны в предположении, что количество цитирова-

ний публикаций отсортировано по убыванию и задано таким вектором:

$$C = (c_1, c_2, \dots, c_N),$$

где c_i – количество цитирований i -ой публикации, причем

$$c_1 \geq c_2 \geq \dots \geq c_N, \quad i = \overline{1, N};$$

N – общее количество публикаций.

Таблица 1. Хиршподобные наукометрические показатели

Показатель	Описание	Отношение	Источник
g-индекс	Максимальное число g самых популярных статей, получивших совместно не меньше g^2 ссылок. Учитывает превышение суммарного цитирования ядра Хирша по сравнению с минимумами требованиями.	$h \leq g$	[13]
hg-индекс	Среднее геометрическое h -индекса и g -индекса: $hg = \sqrt{h \cdot g}$.	$h \leq hg \leq g$	[6]
h_α -индекс	индекс равен h_α , если на каждую из h_α публикаций приходится не менее $\alpha \cdot h_\alpha$ ссылок, а на каждую из оставшихся публикации меньше $\alpha \cdot h_\alpha$ цитирований, $\alpha \in \{1, 2, 3, \dots\}$.	$h_1 = h$	[11]

h_I -индекс	Индивидуальный индекс Хирша: $h_I = \frac{h}{N_A},$ где N_A – среднее количество соавторов статей из ядра Хирша.	$h_I \leq h$	[25]
е-индекс	Квадратный корень избыточного цитирования ядра Хирша: $e = \sqrt{\sum_{j=1, h} c_j} - h^2.$	–	[27]
А-индекс	Среднее число цитирований ядра Хирша: $A = \frac{1}{h} \sum_{j=1, h} c_j.$	$R = \sqrt{h \cdot A}$	[23]
R-индекс	Квадратный корень из суммарного цитирования ядра Хирша: $R = \sqrt{\sum_{j=1, h} c_j}.$	$R = \sqrt{h \cdot A}$	[20]
m-индекс	Медиана количества цитирований ядра Хирша.	$h \leq m$	[9]
m-quotient	Относительный индекс Хирша: $m_q = \frac{h}{y}$, где y – возраст первой статьи автора.	$h = m_q \cdot y$	[18]
i10	Количество статей, каждая из которых получила не менее 10 цитирований.		Google Scholar

MaxProd	$MaxProd = \max_{i=1, h} (j \cdot c_j)$	$MaxProd \geq h^2$	[19]
---------	---	--------------------	------

Индекс Хирша является целочисленным индикатором. При достижении автором больших значений индекса Хирша сильно проявляется его инерционность, вязкость – он может годами оставаться постоянным. В этом случае для формализованного отслеживания деятельности ученого и прогнозирования результативности исследований применяют рациональные модификации индекса Хирша: Sh-индекс [4] и h_{rat} -индекс [24]. Целая часть этих показателей эквивалентна обычному индексу Хирша. Дробная часть показывает насколько автор приблизился к следующему значению индекса Хирша.

В Sh-индексе дробная часть рассчитывается как доля наполнения статьями следующего ядра Хирша (рис. 1). Другими словами, дробную часть Sh-индекса можно интерпретировать как долю выполнения плана по статьям для получения следующего значения индекса Хирша. Математически дробная часть рассчитывается так:

$$\Delta = \frac{\sum_{j=1, h} r_j}{h+1},$$

где $r_j = \begin{cases} 1, & \text{если } c_j > h \\ 0, & \text{иначе} \end{cases}$ – индикатор принадлежности статьи следующему ядру Хирша.

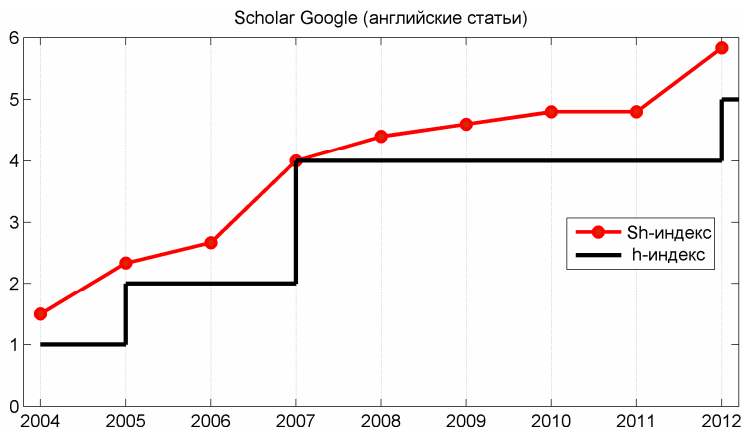


Рис. 1. Динамика h -индекса и Sh -индекса Сергея Штовбы

В h_{rat} -индексе единичный отрезок $[h, h+1]$ разбивается на $2h(C)+1$ равные части в соответствии с минимально необходимым количеством цитирований для получения следующего значения индекса Хирша. Затем к h – количеству публикаций из текущего ядра Хирша – прибавляется доля выполнения плана цитирования по формированию следующего ядра Хирша. Математически дробная часть рассчитывается так:

$$\Delta = 1 - \frac{(h+1 - c_{h+1}) + \sum_{j=1, h} (1 - r_j)}{2h+1}.$$

5. Выводы

Проведен обзор основных наукометрических показателей, которые учитывают количество публикаций и количество цитирований как отдельно, так и совместно. Показаны способы учета дополнительной информации по количеству соавторов, по статусу издания, продолжительности научной карьеры, договорным цитированиям и т.п. Рассмотрены “подводные камни” индексов цитирования, связанных со скрытыми и неформальными ссылками, а также с ошибками в списке литературы.

Литература

1. КАРА-МУРЗА С.Г. Цитирование в науке и подходы к оценке научного вклада // Вестник АН СССР. – 1981. – №5. – С.68-75.
2. ПИСЛЯКОВ В.В. *Наука через призму статей* // Публичные лекции "Полит.ру". – 2011. – URL: http://polit.ru/article/2011/12/21/pislyakov_2011/
3. СОЙФЕР В.Н. *Международная Соросовская программа образования. Часть 2. Результаты именных конкурсов* // Соросовский образовательный журнал. – 1996. – №1. – С. 4–16.
4. ШТОВБА С.Д., ШТОВБА Е.В. *Sh-индекс – новая дробная модификация индекса Хирша* // Научные труды Винницкого национального технического университета. – 2011. – №3. URL: http://www.nbuu.gov.ua/e-journals/vntu/2011_3/2011-3_ru.files/ru/11sds moh_ru.pdf
5. ШТОВБА С.Д., ШТОВБА О.В. *Застосування принципів сітьового маркетингу для оцінювання дифузії наукових знань* // Вісник Вінницького політехнічного інституту. – 2012. – №2. – С. 39–44.
6. ALONSO S., CABRERIZO F., HERRERA-VIDE MA E., HERRERA F. *hg-index: a new index to characterize the scientific output of researchers based on the hand g- indices* // Scientometrics. – 2010. – Vol. 82, №2. – P. 391–400.
7. AWREY J., INABA K., BARM PARAS G., RECINOS G., TEIXEIRA P., CHAN L., TALVING P., DEMETRIADES D. *Reference accuracy in the general surgery literature* // World journal of surgery. – 2011. – Vol. 35., №3. – P. 475–479.
8. BOLLEN J., RODRIQUEZ M.A., VAN DE SOMPEL H. *Journal status* // Scientometrics. – 2006. – Vol. 69, №3. – P. 669–687.
9. BORN MANN L., MUTZ R., DANIEL H. Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from

- biomedicine // Journal of the American Society for Information Science and Technology. – 2008. – Vol. 59, №5. – P. 830–837.
10. BUELA-CASAL G. *Assessing the quality of articles and scientific journals: Proposal for weighted impact factor* // Psychology in Spain. – 2004. – Vol. 8, №1. – P. 60–76.
 11. ECK N.V., WALTMAN L. *Generalizing the h- and g-indices* // Journal of Informetrics. – 2008. – Vol. 2, №4. – P. 263–271.
 12. EGGHE L. *Mathematical theory of the h- and g-index in case of fractional counting of authorship* // Journal of the American Society for Information Science and Technology. – 2008. – Vol. 59, №10. – P. 1608–1616.
 13. EGGHE L. *Theory and practice of the g-index* // Scientometrics. – 2006. – Vol. 69, №1. – P. 131–152.
 14. EICHORN P., YANKAUER A. *Do authors check their references? A survey of accuracy of references in three public health journals* // American Journal of Public Health. – 1987. – Vol. 77. – №8. – P. 1011–1012.
 15. FOLIN O., CIOCALTEU V. *On tyrosine and tryptophane determinations in proteins* // The Journal of Biological Chemistry. – 1927. – Vol. 73, №2. – P. 627–650.
 16. FRANCESCHINI F., MAISANO D., PEROTTI A., PROTO A. *Analysis of the ch-index: an indicator to evaluate the diffusion of scientific research output by citers* // Scientometrics. – 2010. – Vol. 85. – P. 203–217.
 17. GARCIA-PEREZ M.A. *The Hirsch h index in a non-mainstream area: methodology of the behavioral sciences in Spain* // The Spanish Journal of Psychology. – 2009. – Vol. 12, №2. – P. 833–849.
 18. HIRSCH J.E. *An index to quantify an individual's scientific research output* // Proceedings of the National Academy of Sciences of the USA. – 2005. – Vol. 102, №46. – P. 16569–16572.
 19. KOSMULSKI M. *MAXPROD – a new index for assessment of the scientific output of an individual, and a comparison* // Cybermetrics. – 2007. – Vol. 11. – №. 1.

20. LIANG B.J.L., ROUSSEAU R., EGGHE L. *The R- and AR-indices: complementing the h-index* // Chinese Science Bulletin. – 2007. – Vol. 52, №6. – P. 855–863.
21. LOWRY O.H., ROSBROUGH N.J., FARR A.L., RANDALL R.J. *Protein measurement with the Folin phenol reagent* // The Journal of Biological Chemistry. – 1951. – Vol. 193, №1. – P. 265–275.
22. MARX W., CARDONA M. *The citation impact outside references –formal versus informal citations* // Scientometrics. – 2009. – Vol. 80., №1. – P. 1–21.
23. ROUSSEAU R. *New developments related to the Hirsch index* // Science Focus. – 2006. – Vol. 1, №4. – P. 23–25 (in Chinese). English version at <http://eprints.rclis.org/6376/>
24. RUANE F., TOL R. *Rational (successive) h-indices: An application to economics in the Republic of Ireland* // Scientometrics. – 2008. – Vol. 75, №2. – P. 395–405.
25. SCHREIBER M. *A modification of the h-index: The h(m)-index accounts for multi-authored manuscripts* // Journal of Informetrics. – 2008. – Vol. 2, №3. – P. 211–216.
26. VAN NOORDEN R. *A profusion of measures* // Nature. – 2010. – Vol. 465. – P. 864–866.
27. ZHANG C.-T. *The e-index, complementing the h-index for excess citations* // PLoS ONE. – 2009. – Vol. 4, №5. URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005429>

A SURVEY OF SCIENTOMETRIC INDICATORS FOR ASSESSMENT OF PUBLICATION ACTIVITY OF A RESEARCHER

Serhiy Shtovba, Vinnytsia National Technical University, Vinnytsia, Ukraine, Doctor of Science, professor (*Khmelnitske Shose, 95, Vinnytsia, Ukraine, www.shtovba.vinnitsa.com, shtovba@gmail.com*).

Olena Shtovba, Vinnytsia National Technical University, Vinnytsia, Ukraine, Candidate of Science, associate professor (*Khmelnitske Shose, 95, Vinnytsia, Ukraine, olena.shtovba@yahoo.com*).

Abstract: Today, in order to evaluate the productivity of research activities, along with expert opinions scientometric indices are being increasingly used. This article provides an overview of the main scientometric indices that are taken into account the number of publications and the number of citations both separately and together. Ways to incorporate additional information such as the number of co-authors, journal status, duration of a scientific career, bargain citations, etc. are shown. Pitfalls of scientometric indices associated with the hidden and informal citations, as well as errors in the reference list are shown.

Keywords: scientometrics, citation index, h-index, hidden citation, informal citation.

Статья представлена к публикации
членом редакционной коллегии ...заполняется редактором...

Поступила в редакцию ...заполняется редактором...
Опубликована ...заполняется редактором...