

Пристайные списки литературы: статистический анализ

Кочетков Д. М. (УрФУ, Екатеринбург)

На вопрос, сколько должно быть источников в списке литературы, не так просто ответить (хотя отечественные журналы часто вводят количественные ограничения для списков литературы – например, не менее 20 источников). Однако разные научные области имеют различные традиции цитирования и, соответственно, различную длину пристайных списков литературы. Мы проанализировали документы в 9 предметных областях базы данных Scopus. Мы также сравнили 2000 самых цитируемых документов по предметным категориям за все время и за последние несколько лет. Оказалось, что, несмотря на многократный рост количества публикаций, количество цитирований осталось практически неизменным. Рост наблюдается только в социальных и гуманитарных науках. В итоге мы определили среднюю длину пристайных списков в различных предметных категориях в разрезе по трем типам документов (статья, материалы конференции, обзорная статья). Наконец, мы решили взять 50 самых цитируемых статей в одной из категорий (для примера мы выбрали «Economics, Econometrics and Finance») и с помощью корреляционного анализа выяснить, есть ли зависимость между цитированием, количеством ссылок, возрастом статьи, количеством документов (опыт) и индексом Хирша первого автора в БД Scopus. Анализ показал единственную зависимость между индексом Хирша и количеством работ автора, что логически объяснимо. В остальных случаях связей не выявлено.

Ключевые слова: пристайные списки литературы, потенциал цитирования, индекс цитирования.

It is not easy to answer the question how many sources should be included into the reference list (although Russian journals often introduce quantitative restrictions, e.g., no less than 20 sources). However, different research areas have different citation traditions and, accordingly, different length of reference lists. We analyzed the documents in nine subject areas retrieved

from Scopus. We also compared the 2000 most cited papers in subject categories of all time and over the recent few years. Despite the huge increase in the number of publications, it turned out that the number of citations remained practically unchanged. Growth was observed only in the social science and humanities. As a result, we have determined the average length of reference lists in various subject categories in terms of three types of documents (article, conference proceedings, review). Finally, we decided to take the 50 most cited papers in one of the categories (we chose «Economics, Econometrics and Finance») and find out using correlation analysis whether there is a relationship between citation, the number of references, the “age” of the article, the number of documents in Scopus (experience) and Hirsch index of the first author. The analysis has shown a certain relationship between the h-index and the number of papers by the author, which is logically explainable. In other cases, relationships have not been identified.

Key words: reference list, citation potential, citation index

Будучи библиографом, я довольно часто сталкивался с вопросом «Сколько должно быть источников в библиографическом списке научной статьи?» На этот вопрос не так просто ответить (хотя отечественные журналы часто вводят количественные ограничения для списков литературы – например, не менее 20 источников). Однако разные научные области имеют различные традиции цитирования [5] и, соответственно, различную длину пристатейных списков литературы. Юджин Гарфилд называл данный феномен «потенциалом цитирования» научной области и определял его как среднее количество процитированных источников на одну статью [3]. Поэтому начиная с 1980-х гг. прошлого столетия предпринимались попытки статистических исследований и создания наукометрических индексов, которые учитывали бы «привычки» цитирования в конкретной научной области [2, 6–9].

В настоящее время существует ряд индексов цитирования, которые учитывают предметную область. В частности, компания Elsevier использует метрику Source Normalized Impact per Paper, которая отображается в большинстве ее продуктов (например в базе данных Scopus). Source Normalized Impact per Paper (SNIP) был создан профессором Хэнком Мудом из CTWS, University of Leiden [1]. SNIP измеряет контекстуальное влияние публикации с помощью взвешивания ци-

тирований на основе общего количества цитирований в предметной категории. Цитирование имеет большую значимость в предметной категории с меньшим показателем цитирования и наоборот. Рассчитывается как отношение цитирований статьи к потенциальному количеству цитирований в предметной области. Это позволяет проводить прямое сравнение источников из разных предметных категорий. Оказывается, потенциал цитирования варьируется не только между предметными категориями – группами журналов в одной области исследования – или дисциплинами (например, журналы по математике, механике и социальным наукам имеют, как правило, меньше цитирований, чем журналы в категории “Науки о жизни”), но и между различными журналами в пределах одной предметной категории. Например, фундаментальные журналы имеют тенденцию иметь больше цитирований, чем прикладные или клинические журналы, а журналы, охватывающие развивающиеся отрасли, имеют больше цитирований, чем периодика по классическим предметам или журналы более общего профиля. SNIP нивелирует данные различия.

Со своей стороны, мы проанализировали документы в 9 предметных областях базы данных Scopus:

- Arts and Humanities
- Chemistry
- Computer Science
- Economics, Econometrics and Finance
- Engineering
- Materials Science
- Mathematics
- Physics and Astronomy
- Social Science

Анализ показывает, что наибольшее количество документов в категории «Engineering», наименьшее – в «Economics, Econometrics and Finance» и «Arts and Humanities» (рис. 1). Однако в этих областях достаточно высока доля документов 2012–2017 гг., таким образом, данные области можно назвать быстрорастущими. Лидером роста являются компьютерные науки, где доля документов последних лет 34%.

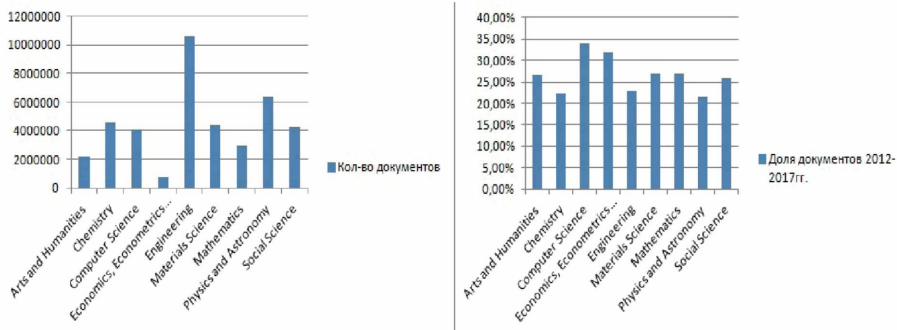


Рис. 1. Количество документов по предметным категориям и доля «свежих» документов

Мы также сравнили 2000 самых цитируемых документов по предметным категориям за все время и за последние несколько лет (рис. 2). Оказалось, что, несмотря на многократный рост количества публикаций, количество цитирований осталось практически неизменным. Рост наблюдается только в социальных и гуманитарных науках.

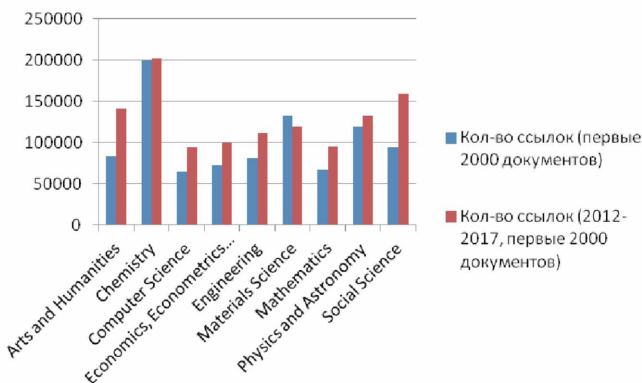


Рис. 2. Количество ссылок у 2000 самых цитируемых документов за все время и за 2012–2017 гг.

Среднее количество ссылок по предметным категориям и типам документов представлено в таблице. Для анализа мы выбрали 2000 самых цитируемых документов в каждой категории.

Среднее количество библиографических ссылок по предметным категориям и типам документов

Предметная категория	Среднее количество ссылок	Среднее количество ссылок по типу документа: article	Среднее количество ссылок по типу документа: conference paper	Среднее количество ссылок по типу документа: review
Arts and Humanities	41,31	27,32	27,25	62,36
Chemistry	99,68	67,85	37,84	153,24
Computer Science	32,20	28,89	17,13	64,55
Economics, Econometrics and Finance	36,06	30,64	30,24	47,23
Engineering	40,52	32,30	20,08	106,51
Materials Science	66,04	43,68	29,62	129,00
Mathematics	33,25	28,57	22,76	69,39
Physics and Astronomy	59,31	44,70	34,06	157,40
Social Science	46,97	36,82	36,34	63,19

Наконец, мы решили взять 50 самых цитируемых статей в одной из категорий (для примера мы выбрали «Economics, Econometrics and Finance») и с помощью корреляционного анализа выяснить, есть ли зависимость между цитированием, количеством ссылок, возрастом статьи, количеством документов у первого автора в БД Scopus (опыт) и индексом Хирша первого автора в БД Scopus. Иногда подобные исследования позволяют выделить интересные закономерности. Например, экономист Гарвардского университета Вэй Хуанг выявил корреляционную взаимосвязь между цитированием, положением автора по алфавиту в пристатейном списке литературы и длиной этого списка [4] (рис. 3).

	Возраст статьи	Количество цитирований	Количество ссылок	Количество работ у первого автора	Индекс Хирша первого автора	
Возраст статьи		1				
Количество цитирований	0,11772423		1			
Количество ссылок	-0,2201929	0,302463487		1		
Количество работ у первого автора	-0,068871966	-0,048937528	0,088154833		1	
Индекс Хирша первого автора	0,021917459	0,021741492	0,224144835	0,853484476		1
SJR, 2015	0,107023569	-0,179430011	-0,037860257	-0,079130105	-0,10207451	1

Рис. 3. Корреляционный анализ библиографических атрибутов статьи и автора

Анализ показал единственную зависимость между индексом Хирша и количеством работ автора, что логически объяснимо. В остальных случаях связей не выявлено. Таким образом, мы можем сделать главный вывод, что «священного Грааля» не существует: на цитирование статьи влияет исключительно ее содержание. Однако в последние годы в связи с массовым распространением движения открытого доступа (open access movement) мы можем предположить, что доступность статьи также влияет на ее цитирование. Данное предположение станет темой для дальнейшего исследования.

Литература

1. *About Source Normalized Impact per Paper (SNIP) // Journal Metrics [Electronic resource]. URL: <http://www.journalmetrics.com/snip.php> (дата обращения: 06.04.2016).*
2. *Braun T. World flash on basic research—The newest version of the facts and figures on publication output and relative citation impact of 100 countries 1981–1985 / T. Braun, W. Glänzel, A. Schubert // Scientometrics. 1988. V. 13. № 5. P. 181–188.*
3. *Garfield E. Citation indexing. Its theory and application in science, technology and humanities / E. Garfield. New York: Wiley, 1979. 296 p.*
4. *Huang W. Do ABCs get more citations than XYZs? / W. Huang // Economic Inquiry. 2015. V. 53. № 1. P. 773–789.*
5. *Lancho-Barrantes B. S. What lies behind the averages and significance of citation indicators in different disciplines? / B. S. Lancho-Barrantes, V. P. Guerrero-Bote, F. Moya-Aneón // Journal of Information Science. 2010. V. 36. № 3. P. 371–382.*

6. Leeuwen T. N. Van. *Development and application of journal impact measures in the Dutch science system* / T. N. Van Leeuwen, H. F. Moed // *Scientometrics*. 2002. V. 53. № 2. P. 249–266.

7. Marshakova-Shaikevich I. *The standard impact factor as an evaluation tool of science and scientific journals* / I. Marshakova-Shaikevich // *Scientometrics*. 1996. V. 35. № 2. P. 283–291.

8. Moed H. F. *Measuring contextual citation impact of scientific journals* / H. F. Moed // *Journal of Informetrics*. 2010. V. 4. № 3. P. 265–277.

9. Zitt M. *Modifying the journal impact factor by fractional citation weighting: The audience factor* / M. Zitt, H. Small // *Journal of The American Society for Information Science and Technology*. 2008. V. 59. № 11. P. 1856–1860.