

## 50 миллионов: примерное количество существующих научных статей

Перевод статьи, опубликованной в журнале *Learned Publishing* в июле 2010 года: Arif E. JINHA  
Article 50 million: an estimate of the number of scholarly articles in existence.

Автор статьи – Arif E. JINHA канадский ученый из Оттавского университета.

### Article 50 million: an estimate of the number of scholarly articles in existence

URL: <https://onlinelibrary.wiley.com/doi/abs/10.1087/20100308>

**Аннотация.** Сколько всего научно-исследовательских статей существует? Статьи в журналах впервые появились в 1665 г., и общее их количество в 2009 году уже преодолело отметку в 50 миллионов. Данная сумма была получена на основе данных о глобальном годовом объеме публикаций, представленных в 2006 году, а также на базе анализа годового объема публикаций и темпов роста за последнее десятилетие.

#### Введение

Начиная с первой модели современного журнала *Le Journal des Sçavans*, опубликованного во Франции в 1665 г., а затем и *Philosophical Transactions*, опубликованными королевским обществом (*Royal Society*) в Лондоне годом позже (1), число действующих научных журналов стабильно увеличивается. К 2006 году по грубым подсчетам насчитывалось 23 750 журналов (2). Существует прямая взаимосвязь между исследователями, журналами и статьями (3). Бьорк и др. (2) (Björk et. al) утверждали, что изменения в динамике исследований (или анализа) научных публикаций, спровоцированные революцией в области коммуникаций, сделали саму статью актуальной сегодня, как базовую молекулярную единицу распространения исследования.

Разница была выявлена благодаря изучению глобального выпуска исследований, проведенных в последнее десятилетие, в которых предоставляются данные о темпах роста и годовые показатели о статьях и исследователях (3-6). В то время как исследователи завершают свою карьеру, появляется ещё больше новых исследователей, журналы закрываются, однако еще больше новых начинают свою работу. Изменения в числе действующих исследователей и журналов отражает динамику публикации и исследования, а также увеличение полного объема текущего производства (5). Однако, статья носит статичный характер, что делает ее идеальным критерием. Созданные и опубликованные статьи почти никогда не уничтожаются. Они всегда могут быть восстановлены, а благодаря цитированию, каждая статья занимает определенную позицию в архитектуре, которую исследователи могут продолжать строить. По сути, статьи зародились стараниями журналов и их издателей, но они переживали и смерть журналов. Несмотря на то, что направления науки развивают различные области исследования, в конечном итоге, в самой науке нет установленных рамок — это единая система документирования письменных знаний.

Соответственно метрика, которая описывает количественное целое данной системы — общее количество всех современных статей в научных журналах, существующих в настоящий момент, или существовавших в любое другое время, может быть полезной как отправная точка для исследования структуры самой системы. Также, получение более точных расчетов всего объема исследований поможет ученым в достижении значительных результатов. Такого рода расчеты позволят ученым обозначить географию производства знаний, определять пути поиска статей и извлекать контент, при этом обеспечивая его сохранность и доступность. В данной работе представлены расчеты общего количества всех рецензируемых статей, опубликованных во всем мире с 1665 года. В разделе приложения находится ссылка на воспроизведенные ранее исследования, отражающие текущее число действующих журналов. Результаты за 2009 год.

#### Литературный обзор

Исследование объема публикаций научных статей в рецензируемых журналах, а также статей из обычных журналов никогда не было точным. Однако, существует несколько работ, в которых была предпринята

попытка определить объем выпускаемых научных статей, относящихся к послевоенному периоду «большой науки» (Big Science), также как и более современные работы начиная с 90х годов и до нашего времени.

В 1963 году Дерек де Солла Прайс (Derek de Solla Price) составил график роста числа журналов с 1665 г. по 2000 г. и предсказал, что к 2000 году будет существовать астрономическое количество журналов, 1 миллион. Прайс также определил ключевую взаимосвязь между инвестициями, количеством исследователей, журналов, аннотаций и статей. Данная взаимосвязь была перенесена и в более поздние исследования. Расчеты числа журналов по всему миру были произведены Кингом и соавторами (King et al.) в 1977 году: 57 4000, и Meadows and Singleton в 1995 году: 70 000 – 80 000 (7). Более поздние исследования сообщают, что цифры намного скромнее, чем в представленных ранее расчетах и прогнозах. Мейб и Амин (Mabe, Amin) во введении своей работы 2001 года описывали, что усовершенствования системы классификации Ulrich's (8) позволяют получить более реалистичные расчеты (5). Основываясь на этом, Мейб представил аргументы в пользу нового подхода в своей статье 2003 года (3). Результаты расчетов, произведенных ранее были завышены, по той причине, что исследователи не имели возможность отличить рецензируемые журналы от других периодических изданий, а также действующие журналы от уже закрытых. Важно, что мы не можем воспринимать темпы роста как экспоненциальные или кумулятивные, как это предполагал Прайс, что и объясняет, почему сегодня мы не имеем ничего похожего с 1 миллионом журналов (3, 7).

Мейб (3) использовал поисковые термины в системе классификаций Ulrich's для фильтрации научных, рецензируемых и активных журналов, а также и функцию «AND NOT» для нескольких терминов, которая не позволяет ресурсам базы данных быть включенным в описание научного журнала. В 2001 году Мейб рассчитал общее количество журналов: 14 694. Он также проверил выявленные Прайсом взаимосвязи между количеством исследователей, журналов и статьями, чтобы уточнить, насколько обоснованы оказались расчеты. Это было сделано путем определения количества журналов, проиндексированных в базе данных ISI, а для расчета количества не проиндексированных журналов в базе применялся закон Брэдфорда. Второй метод дал результат в 16 000 журналов (3). Следовательно, первые расчеты, которые более точно определяли параметры, а также включали в себя только действующие, рецензируемые и научные журналы, могут быть использованы в данном исследовании.

Тенопир и Кинг (6) (Tenopir and King) оценили глобальный годовой объем публикаций в 1 миллион статей на рубеже двадцатого века. Расчеты основаны на эмпирических данных о количестве активных исследователей и среднем объеме исследований на автора исследования. Бьорк и др. получили результат в 23 750 журналов за 2006 год с использованием метода, представленного Мейбом (3, 5), хотя Бьорк и др. не использовали функцию «AND NOT» в своем исследовании (в любом случае это не являлось необходимым). Далее авторы смогли произвести первый расчет глобального годового объема выпуска статей. Для этого они выделяли журналы, индексированные по ISI, которые, как правило, публикуют больше статей, чем журналы, не относящиеся к ISI, а затем определяли среднее количество статей на журнал для каждой категории (косвенным методом для ISI, описанным в исследовании, и статистической выборкой для не проиндексированных в ISI журналов). Расчет суммы журналов, умноженной на среднее количество статей на журнал для каждой категории, дает результат в 1 346 000 (округленный) статей за 2006 год. Интересно, что хотя журналы ISI и представляют 36% от общего числа активных журналов, статьи ISI, в свою очередь, представляют 70% от общего количества статей (2).

Во время обзора данной статьи на форуме American Scientist Open Access возникла оживленная дискуссия о большом разбросе в расчетах, что демонстрирует сложность в получении точных цифр (9). Моррис (10) обсуждал ограничения использования базы данных Ulrich's, заявляя:

«Издатели каталога полностью зависят от информации, поставляемой журналами, которые в него входят. Новые журналы часто не вносятся в базу сразу. Соответственно, не может быть никаких четких гарантий относительно полноты, актуальности и точности той информации» (стр. 299).

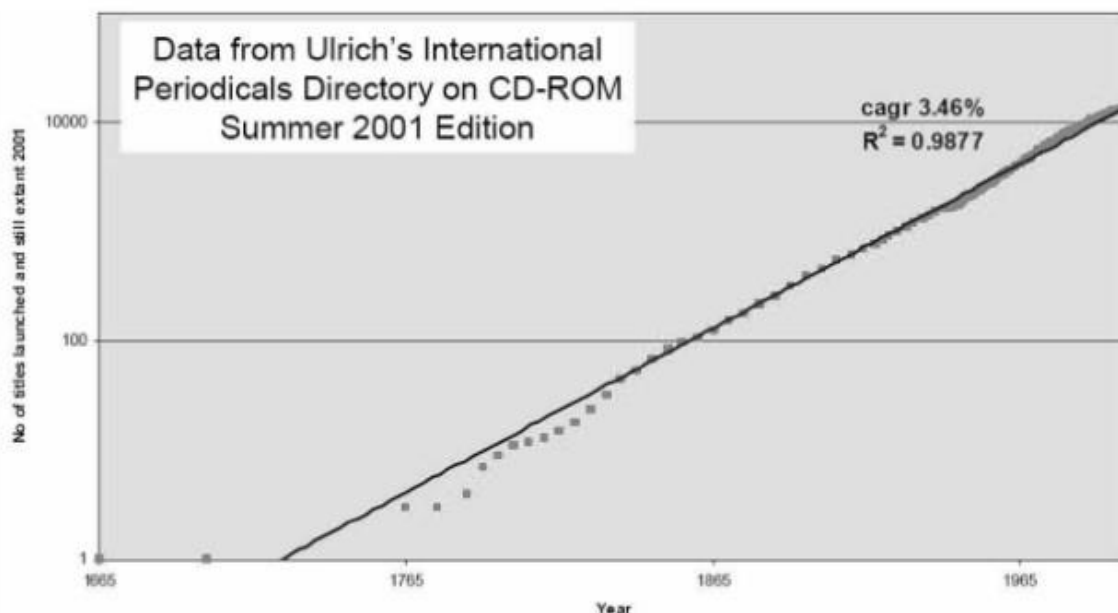
На форуме AMSCI была выражена обеспокоенность тем, что огромное количество небольших журналов, особенно издающихся на других языках, и публикующихся в экономически развитых странах, с большей вероятностью будут исключены, что приведет к искаженному взгляду на мир научной публикации, и к недооцененности его размеров (9). Эта проблема обсуждалась Тенопиром и Кингом (7) в их недавней книге «Будущее научных изданий» (The Future of Academic Publishing). Однако, за исключением ручного метода исчисления журналов, Ulrich's остается самой обширной базой данных для определения общих

показателей во всем мире и наиболее чувствительной к фильтрации по ключевым различиям, таким как активные, рецензируемые и научные журналы (7, 10). Более того, результаты, полученные из Ulrich's были последовательны в том, что мы понимаем о взаимоотношениях между количеством исследований, журналов и статей, а также темпами роста (2, 3). В данном исследовании был воспроизведен базовый метод поиска по Ulrich's для определения числа журналов в 2009 году. Из этого рисунка мы можем получить оценку глобального годового объема выпуска статей, если не допустить значительных изменений среднего числа статей в журналах проиндексированных в ISI, а также не проиндексированных в ISI журналах с 2006 года.

### Методы

Расчеты основаны на измерении объема научной продукции в 2006 году, которые представили Бьорк и др. (2), опираясь на предположение, что Мейб (3), Уэйр (4) (Ware), Мейб и Амин (5), а также Тенопир и Кинг (6) правы в том, что сообщают о стабильном увеличении числа исследователей, журналов и статей за три столетия. В то время как по данным, представленным авторами, темпы роста числа журналов составляют в среднем 3,26%, Уэр сообщает о росте выпуска статей примерно 3% в год (4), что дает удвоение чуть менее, чем за 24 года.

Мы выбрали 1726 год в качестве начального года для наших расчетов, поскольку он соответствует началу линии устойчивого роста журналов, показанной на рисунке 1 (воспроизведено из каталога периодических изданий Ulrich's 2001 года по Мейбу (3)).



**Рисунок 1.** Количество журналов, выпущенных за год

*Источник: Международный каталог периодических изданий Ульриха, воспроизведенный с разрешения Mabe (3).*

В литературе сообщается о стабильном росте на протяжении более двух-трех веков (3, 4). Самый ранний период публикации, с 1665 года до середины XVIII-го века, демонстрирует менее предсказуемый рост. Ulrich's не дает результатов по количеству журналов в 1726 году, поэтому этот метод не может быть применен для определения начальной цифры. Соответственно будет более разумным начать 1726 год с числа, большего нуля, но незначительно малого в отношении глобального количества на сегодня. Кроме того, количество статей может быть установлено как значение, которое представляет более точные результаты расчетов общего объема публикаций за последнее десятилетие, при применении кривой роста в 3%, как форма обратного отображения.

Когда количество статей для 1726 года установлено равным 344, кривая соответствует расчетам общего объема публикации Тенопира и Кинга (6) на рубеже тысячелетий, а также расчетам Бьорк и др. (2) для количества статей в 2006 году. Это было сделано с помощью электронной таблицы Excel. Предполагаемый годовой объем публикаций за каждый год с 1726 года, а также совокупный итог приведены по ссылке в приложении. Мейб (3) сообщает, что в период «большой науки» с 1946 по 1976 год, темпы роста журналов были самыми большими из года в год, при этом показатели были ниже средних до Второй мировой войны и после 1976 года. Автор использовал те же множители для изменений скорости роста количества журналов, чтобы скорректировать скорость роста количества статей и рассчитать сумму с учетом данной изменчивости. Поскольку результаты обоих расчетов были практически идентичны, здесь представлены более простые расчеты, основанные на среднем темпе роста.

### Результаты

Данный метод позволил получить результаты в 50 миллионов статей к концу 2008 года, причем ожидается, что в 2009 году эта цифра превысит 50 миллионов. Фактически, год, в котором сумма всех научных статей превысила 50 миллионов, не может быть точно определен, но мы можем сообщить, что этот результат является первой оценкой, которая следует из текущих данных об общем количестве всех научных статей, которые существовали на момент написания в 2009 году и публикации в 2010 году. Автор с юмором отмечает, что эта статья могла стать 50-миллионной!

Предполагаемая общая сумма на 31 декабря 2008 года = 49 234 626.

Предполагаемая общая сумма на 31 декабря 2009 года = 50 712 009.

Воспроизводя расчеты Björk et al. (2) и используя поисковые термины «академический/научный» (Academic / Scholarly), «рецензируемый» (Refereed) и «действующий» (Active) в базе Ulrich's, общее количество действующих журналов за 2009 год составляет 26 406.

Предполагая небольшое изменение в пропорции журналов проиндексированных в ISI (36%), и используя средние значения, представленные Бьорк и др. (2) для статей на журнал проиндексированный в ISI, и не проиндексированный в ISI, общее число опубликованных статей, оцениваемое в 2009 году по соотношению журналов к статьям, рассчитывается путем оценки количества журналов ISI и вычитания этого количества из общего числа журналов в 2009 году, затем умножая каждую категорию на соответствующее среднее количество статей. Среднее количество статей на журнал, представленное Бьорк и др. для журналов ISI — 111,7; для не проиндексированных в ISI, среднее количество статей составляет 26,2.

Чтобы определить количество журналов ISI, мы умножаем общее количество журналов на 35%.

- $26\,406 \times 0,36 = 9\,506$  (т.е. предполагаемое количество журналов ISI).

Чтобы определить количество статей ISI, мы умножаем количество журналов на количество статей на журнал, для журналов ISI.

- $9\,506 \times 111,7 = 1\,061\,820$  (т.е. предполагаемое количество журналов ISI).

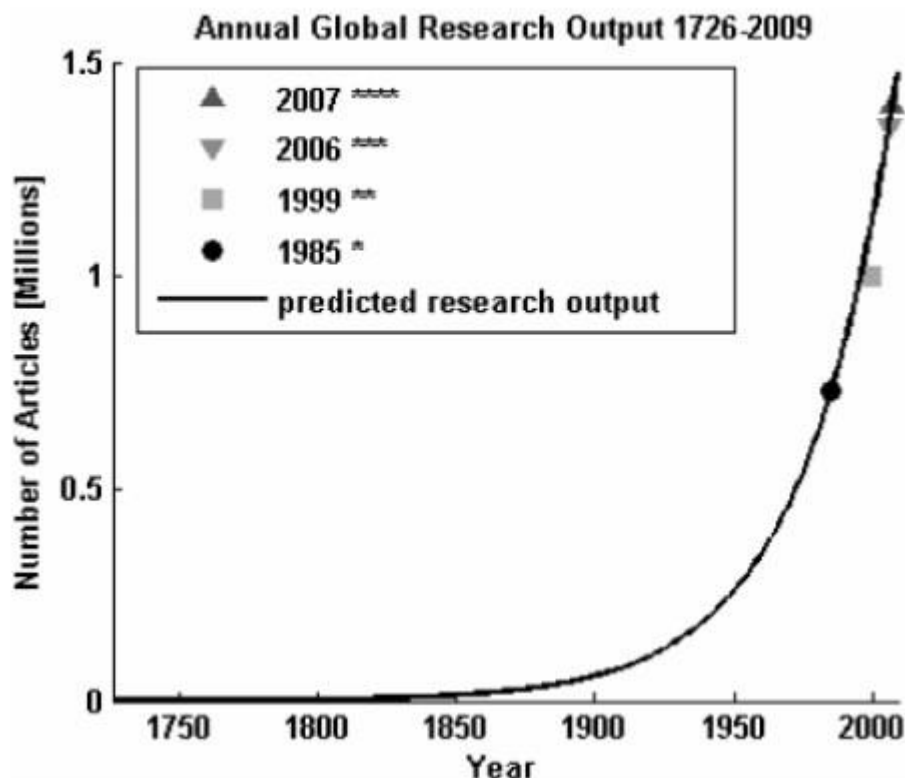
Чтобы определить количество журналов не проиндексированных в ISI, мы вычитаем количество журналов ISI из общего количества журналов.

- $26\,406 - 9\,506 = 16\,900$  (т. е. Примерное количество журналов, не относящихся к ISI).

Чтобы определить количество статей, не относящихся к ISI, мы умножаем количество журналов, не относящихся к ISI, на количество статей на журнал для журналов, не относящихся к ISI.

- $16\,900 \times 26,2 = 442\,780$ .

Чтобы определить годовой объем выпуска статей за 2009 год, мы суммируем количество статей ISI и количество не проиндексированных статей в ISI.



**Рисунок 2.** Расчетный годовой объем глобального исследования в годовой рост на 3%

\* 1985 год (2009 год — 25 лет): удвоение годового объема производства для изделий в возрасте до 24 лет; \*\* 1999 год: соответствует расчетам Тенопира и Кинга (6) для результатов исследований в конце 1990-х годов — 1 миллион статей в год; \*\*\* 2006 год: очень близко соответствует расчетам Björk et al. (2), за 2006 год — 1,35 миллиона статей; \*\*\*\* 2007 год: близко соответствует оценке Уэра (4) за тот же период — 1,4 миллиона изделий в год.

$$1\,061\,820 + 442\,780 = 1\,504\,600.$$

Суммируя их, общий выпуск статей в 2009 году составил 1 504 600. Эти данные показаны на рисунке 2.

Общее количество статей, рассчитанных на 2009 год с использованием кривой роста 3% и составленной в таблице Excel, составляет 1 477 382. Разница между расчетами по методу, использованному Бьорк и др., и методом с использованием кривой роста составляет менее 2%.

### Обсуждение и вывод

Оценка общего количества научных статей, которые существуют в мире, является приблизительной, а не точной цифрой. Тем не менее, исследование размера, роста и состава глобального научного ресурса продвинулось вперед в этом десятилетии. Мы можем более точно определить глобальный годовой объем научного ресурса благодаря нашему пониманию (i) взаимосвязи между количеством исследователей, журналов и статей; (ii) ежегодными темпами роста числа действующих изданий и количества опубликованных статей; (iii) отношений между журналами, проиндексированными в ISI, и не проиндексированными в ISI; и (iv) усовершенствованием системы классификации Ulrich's. Тем не менее, необходимы дальнейшие исследования, чтобы проверить надежность каждого из отношений и, полноту базы данных Ulrich's.

50 миллионов статей в рецензируемых журналах- это впечатляющее наследие и мощный ресурс для человечества. Для управления данным ресурсом таким образом, чтобы он был справедливым, полезным и устойчивым, нам было бы полезно постоянно интересоваться тем, где мы находимся с точки зрения доступа, оцифровки, поиска и индексации, а также сохранения этой глобальной библиотеки знаний.



### Благодарности

Автор хотел бы выразить благодарность Азиму Джинхе (Azim Jinha) за помощь в расчетах, а также за советы и редакционную помощь Робина Бикрофта (Robin Beecroft, Searchlighter) и Мустафу Диака (Moustapha Diack).

### Список литературы

1. Brown, H. 1972. History and the learned journal. *Journal of the History of Ideas*, 33: 365–378. <http://www.jstor.org/stable/27090412>.
2. Björk, B., Roos, A. and Lauri, M. 2008. Global annual volume of peer reviewed scholarly articles and the share available via different open access options. *Proceedings of the ELPUB 2008 Conference on Electronic Publishing*, Toronto, Canada, June 2008. <http://oacs.shh.fi/publications/elpub-2008.pdf>.
3. Mabe, M. 2003. The growth and number of journals. *Serials*, 16: 191–197. Ware, M. *Scientific Publishing in Transition: An Over-view of Current Developments*. Bristol, Mark Ware Consulting, 2006.
4. Mabe, M. and Amin, M. 2001. Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51: 147–162. <http://dx.doi.org/10.1023/A:1010520913124>.
5. Tenopir, C.W. and King, D.W. *Towards Electronic Journals*. Washington DC, Special Libraries Association, 2000.
6. Tenopir, C.W. and King, D.W. 2009. The growth of journals publishing. In Cope, B., and Phillips A. (eds), *The Future of the Academic Journal*. Chandos Publishing/Woodhead Publishing Ltd. ISBN 1 84334 416 5.
7. Ulrich's Periodicals Directory (Ulrichsweb.com) Ulrich's has been a global source of periodicals information since 1932.
8. American Scientist Open Access Forum. 2009 Archives. See discussions with subject line 'Number of Scholarly Journals in the World.' <http://listserver.sigmaxi.org/sc/wa.exe?A1=ind09&L=american-scientist-open-access-forum&F=10>.
9. Morris, S. 2007. Mapping the journal publishing landscape, how much do we know? *Learned Publishing*, 20(4): 299–310. <http://dx.doi.org/10.1087/095315107X239654>.