

Наукометрическая интроспекция: что знают базы данных о самих себе

АВТОР

Владимир Владимирович ПИСЛЯКОВ,

заместитель директора Библиотеки Национального исследовательского университета «Высшая школа экономики»



В любой науке интересно рассмотреть ситуацию, когда инструмент исследования сам становится его объектом. Очень схематично говоря, когда линейкой измеряют линейку или весами взвешивают весы. Усложним: когда проверяют, подчиняется ли учебник физики законам, написанным в нём (почти всегда), или психолог — правилам преподаваемой психологии (редко).

В наукометрии такой подход особенно логичен, он ей внутренне присущ. Классик, один из основателей наукометрии Дерек де Солла Прайс начинал самую известную свою книгу «Little Science, Big Science» с вопроса: «Почему бы нам не применить научные инструменты к самой науке?» А ещё раньше один из отцов современной информатики Поль Отле опубликовал труд с названием «Livre sur le livre» («Книга о книге»). С регулярностью выходят работы, посвящённые наукометрической картографии наукометрии [1] или наукометрическому анализу ведущего наукометрического журнала (собственно, *Scientometrics*), да ещё и опубликованные в нём самом [2].

Поэтому, продолжая исследовать этот наукометрический «уроборос» (змею, кусающую свой хвост), интересно посмотреть: а что знают библиотечные базы данных о самих себе?

Но какие выбрать базы для исследования? Интересно, что будет в JSTOR о JSTOR или в EBSCO про EBSCO. Это одни из пионеров рынка электронных ресурсов (оба возникли ещё в 1990-х гг.). Но мы выберем наукометрическую базу Scopus (производит её компания Elsevier). В ней сейчас больше всего журналов с текущими выпусками, если брать проверенные временем инструменты. Это обеспечит большую статистическую достоверность нашим выводам.

ДИНАМИКА

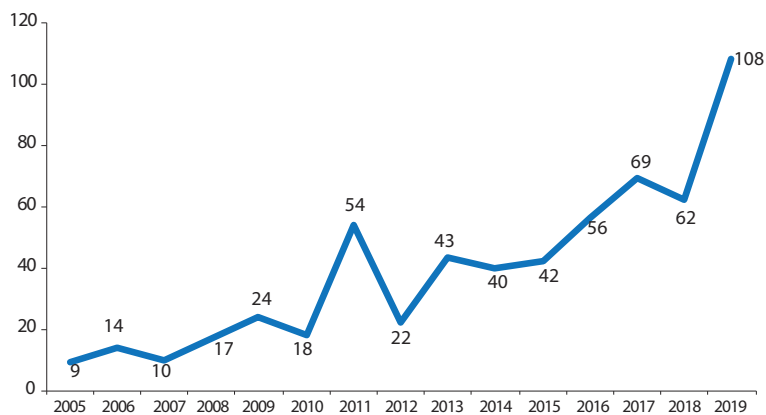
Прежде всего: как часто встречается Scopus в названиях научных статей, размещённых в самой базе данных, и какова динамика? Надо иметь в виду, что эльзевировская наукометрическая база вышла в рабочий режим в 2005 г., поэтому снизу мы ограничим наш поиск этим годом.

При всех наблюдаемых флуктуациях, статей, посвящённых Scopus, становится всё больше (рис. 1). Обратите внимание на те девять статей, которые появились в 2005 г., т.е. сразу после выхода продукта на рынок. Это интересно. Многие опубликованы в неизвестных нам журналах, например в хорватском химическом (!) издании. Там это рассказ о выходе нового наукометрического инструмента, что важно для учёных любой специализации. Известных библиотекарям журналов в этом наборе быстро среагировавших только два: *Serials Librarian* и *Library Journal*. Потому что во многих серьёзных академических изданиях процесс рецензирования длится дольше года.

Если взять другую экстремальную точку, то лидером среди источников 2019 г. оказывается не журнал. Это труды главной мировой наукометрической конференции: *Conference of International Society for Scientometrics and Informetrics — 2019 (ISSI 2019)*. Она проходила в Риме. На конференциях оперативнее всего сообщаются научные результаты, которые уже позже, как правило, публикуются в журналах, если пройдут там рецензирование.

Рис. 1

Число статей о Scopus в базе данных Scopus (всего 588)



Источник: БД Scopus

ОБЛАСТИ НАУКИ

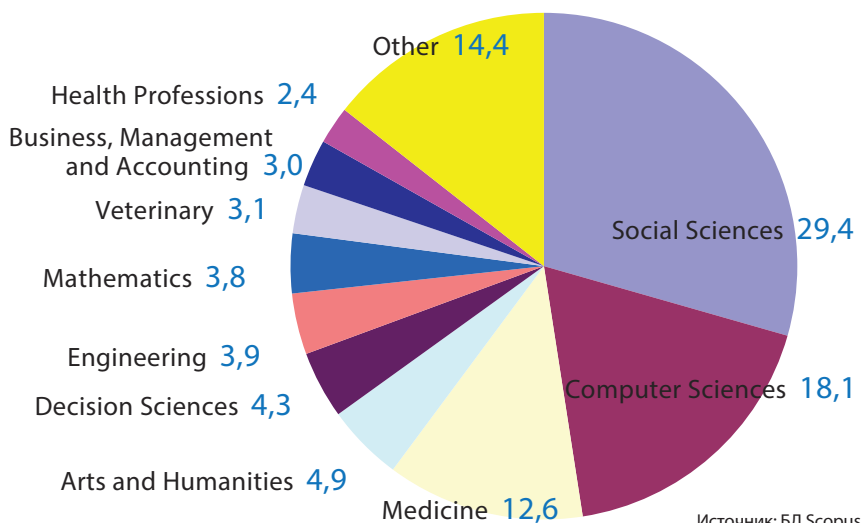
Дальше оценим раскладку по научным дисциплинам: в каких областях науки выходят статьи, посвящённые Scopus? (Напомним: именно посвящённые базе данных или использующие её как основной источник данных — её название встречается в их заголовках.)

Здесь картина, скажем так, не вызывает исследовательского удивления. Наукометрия — область знаний, находящаяся на пересечении общественных наук и естественных. Ведь наука как сфера человеческой деятельности — это социальный феномен. Но наукометрия измеряет его математическими методами. Это видно и в структуре рис. 2. На первых местах социальные и вычислительные науки (social sciences и computer sciences). Главные наукометрические журналы Journal of Informetrics и Scientometrics отнесены сразу к двум этим категориям, а статьи, опубликованные в них, считаются в каждую из дисциплин. Очевидно, что именно в них преобладают публикации про базу данных Scopus.

Менее тривиально третье место — за медицинскими науками. И хотя международный журнал собственно по медицинской информатике Journal of the Medical Library Association хорошо известен, не он здесь играет ведущую роль. Больше всего статей про Scopus неожиданно опубликовала Куба — с 20 статьями в Revista Cubana de Informacion en Ciencias de la Salud (с его прежним названием Acimed) и Revista Habanera de Ciencias Medicas. Дальнейшее распределение областей науки: гуманитарное знание, науки о принятии решений, технические науки — достаточно ожидаемо.

Рис. 2

Распределение статей о Scopus по областям наук



СРЕДНЯЯ ЦИТИРУЕМОСТЬ

Также интересно сравнить цитируемость статей, упоминающих базу данных Scopus, с цитируемостью статей без такого упоминания. Опять сделаем это по самой базе Scopus, но уже расширим поиск. Будем считать упоминаниями те статьи, где название базы данных встречается не только в заголовке, но и в аннотации (реферате) или в ключевых словах. Поскольку нет смысла искать такие публикации во всей мировой литературе, нужно сузить охват источников, на которых можно провести этот эксперимент. Давайте возьмём самый классический журнал Scientometrics (кстати, не все знают, что это перевод русского «наукометрия», а не наоборот: этот термин введён в мировую науку Василием Налимовым). Будем исследовать только статьи, опубликованные в этом журнале.

Далее, чтобы усилить нашу «интроспекцию», давайте для анализа воспользуемся новым вариантом методики CiteScore [3] — оценки средней цитируемости статьи в журнале. Индикатор CiteScore приводится для каждого журнала в самой базе Scopus. Он решает ту же задачу, что и хорошо известный нам импакт-фактор, но учитывает больший (четырёхлетний) временной промежуток. Кроме того, ссылки на журнал учитываются не только из отчётного года (года, для которого рассчитывается показатель), как в импакт-факторе, но и вообще из всех статей, вышедших за тот же четырёхлетний интервал. Как говорят в таком случае, здесь окно цитирования приравнено к публикационному окну. В этом есть несомненный приоритет, но имеются и недостатки. Чем больший мы исследуем интервал, тем шире охват. Но тем опаснее не отследить динамику: например, на журнал перестали ссылаться, а в метрике полученные им цитирования ещё три года будут учитываться. Сейчас вы всё поймёте на конкретных вычислениях.

Итак. Журнал Scientometrics, сравнение статей, где в главных библиографических полях есть упоминание базы данных Scopus, с остальными публикациями. Технические инструкции по сбору нужных данных мы оставим за рамками нашей статьи, однако все числа вы можете получить и сами, проведя поиск в Scopus и далее используя функцию Citation Overview. Итог таков: в 2016–2019 гг. в журнале Scientometrics вышли 137 статей с упоминанием базы данных (в основных полях) и 1339 публикаций без такого упоминания. То есть примерно каждая 11-я статья в той или иной мере касалась Scopus. При этом за весь тот же →

Источник: БД Scopus

промежуток времени (2016–2019) первый массив статей получил в целом 1493 ссылки (из всех источников в Scopus: журналов, книг, трудов конференций), второй был процитирован 6719 раз. Получается, что средняя цитируемость статей с упоминанием Scopus — $1493 : 137 = 10,5$; без упоминания — $6719 : 1339 = 5,0$.

Разница более чем в два раза! Но что именно мы узнали? Говорит ли это что-то о самой базе данных Scopus? Нет. Это говорит лишь об информационном поле вокруг неё, о том, что в наукометрическом сообществе статьи, которые ей посвящены, или её затрагивают, или просто упоминают этот инструмент, получают повышенное внимание и отклик, в среднем более активное обсуждение.

Также обратите внимание: если мы возьмём суммарное число статей, сумму знаменателей ($137 + 1339 = 1476$), а также суммарное число полученных ими ссылок ($1493 + 6719 = 8212$) и разделим второе на первое, то получим почти точный показатель CiteScore за 2019 год для всего журнала *Scientometrics*: $8212 : 1476 = 5,6$. Слово «почти» появилось здесь только по причине того, что методика CiteScore ещё накладывает некоторые ограничения на учитываемые в ней типы публикаций: эти детали в нашей процедуре мы опустили (а индикатор, кстати, получили при этом точный — можете посмотреть значение CiteScore 2019 для *Scientometrics*, эти данные открыты).

«БЛАГОДАРНОСТИ»

Наконец, есть новый пласт информации, индексируемый теперь наукометрическими базами данных, он появился лишь пять лет назад. Напомним: базы данных научного цитирования, «наукометрические базы», не содержат полных текстов статей. Это не полнотекстовые ресурсы, они индексируют (т.е. включают в себя) лишь некоторые библиографические поля: выходные данные, данные об авторах и, главное, списки используемых («цитируемых») источников. Полного текста в них нет.

Рис. 3

За что благодарят базу данных Scopus?



Источник: БД Scopus

Но не так давно в эти инструменты стали добавлять информацию о поддержке, о финансировании проведённого исследования. Смысл понятен: наукометрия граничит с управлением наукой, а в вопросах управления важны не только библиографическая информация и библиометрическая статистика, но и вопросы финансирования, затрат на соответствующие исследования. Поэтому в наукометрических базах появились поля «Финансирующая организация», «Данные о гранте» и даже «Текст, сообщающий о финансировании». Откуда берутся эти сведения? Конечно, только из самих статей. Что в статье написано, то и будет отражено в базе данных. Никаких внешних исследований и проверок не проводится.

Вот именно сопровождающий текст нас и интересует. Дело в том, что чаще всего в него включается весь раздел Acknowledgments («Благодарности»), а там... Там можно найти что угодно, от благодарности Богу до признания своему коту за вдохновение. И это попадает в наукометрические инструменты! Мы же остановимся посередине: посмотрим, как там фигурирует база данных Scopus (опять же, по данным самой Scopus).

В целом анализ полей, относящихся к финансированию, в Scopus уже проводился, даже именно на отечественной почве, обратите внимание на название поддержанного РФФИ гранта [4]. Поскольку автоматическое определение текстовой области Acknowledgments в статье — это сложный процесс, включающий в себя модные ныне алгоритмы искусственного интеллекта, для точности я просмотрел вручную 105 свежих статей 2020 г. выпуска, где в соответствующем поле был термин Scopus. Оказалось, что алгоритмы слишком много захватили из основного текста, не из Acknowledgments. Чаще всего — из раздела методики исследования, где Scopus активно используется (41 случай). Немного меньше — из биографии авторов (32 статьи), где учёные либо идентифицируют себя при помощи Scopus-ID, либо указывают свои Scopus-показатели (Scopus H-index, например), либо скромно хвастаются званиями обладателей премий Young Scopus Researcher Award, NASI Scopus Young Scientist Award, Scopus Eureka Prize, Scopus Researcher of the Year Award и т.д.

Такие данные любопытны, но в нашем контексте они лишние. На данный момент это явная недоработка Scopus: при поиске по Funding Text должен быть найден Funding Text, а не фрагменты основного «тела» статьи. Наша цель — выяснить, как упоминается база именно в разделе благодарностей, и сделать

сейчас это можно только вручную. После отбора таких статей за последний год остаётся 24. Посмотрите *рис. 3*: вот за что благодарят базу данных.

Чаще всего, конечно, это благодарность Scopus (реже — и владельцу, компании Elsevier) за источник данных для исследования. Далее это способ уточнить организацию, которой благодарны авторы за поддержку их работы, — добавляется Scopus ID. Достаточно часто упоминание Scopus фигурирует в названии грантов, по которым финансировались исследования, например University Grants for Scopus and Web of Science Publications, Scopus H-Index Based Research, Foundation Scopus и даже Research Grant of Publikasi Internasional Terindeks Scopus Untuk Tugas Akhir Mahasiswa B (это индонезийский язык). Также благодарят организации, предоставившие авторам доступ к Scopus, или даже называют имена конкретных ассистентов, выполнивших поисковые запросы в базе для статьи.

«Смешные истории во время съёмок» напоследок. В этом небольшом исследовании мне очень сильно помогало одно свойство самого термина Scopus — он, как я думал, уникален и обозначает только базу данных (например, нет ни одного автора с фамилией Scopus). Однако тщательное исследование опровергло и это. Есть гора Mount Scopus (внизу на фото, там располо-

жен Еврейский университет в Иерусалиме), и она фигурирует в нескольких адресах авторов. А ещё на *рис. 3* есть пункт «Идентификация проекта». На самом деле этот проект называется «Smart Converters for Optimized Power Usage and Storage». Прочитайте по первым буквам. Сплошное царство всепобеждающего электричества — и никаких баз данных цитирования! ■

ЛИТЕРАТУРА

1. Chen C. Mapping scientometrics (1981–2001) / C. Chen, K. McCain, H. White & X. Lin // Proceedings of the ASIST Annual Meeting. — 2002. — Vol. 39. — P. 25–34. — DOI:10.1002/meet.1450390103.
2. Hou H. The structure of scientific collaboration networks in scientometrics / H. Hou, H. Kretschmer & Z. Liu // Scientometrics. — 2008. — Vol. 75, no. 2. — P. 189–202. — DOI:10.1007/s11192-007-1771-3.
3. How are CiteScore metrics used in Scopus? [electronic resource] // Elsevier. — 2020. — Access mode: https://service.elsevier.com/app/answers/detail/a_id/14880/supporthub/scopus.
4. Стерлигов И.А. (руководитель). Использование данных о финансировании в Scopus для сравнительной оценки источников проектного финансирования российской науки: возможности и ограничения: грант РФФИ №18-311-00289.

Гора Скопус

