

## 1. Введение

В настоящем пособии рассматривается ряд библиометрических индикаторов, связанных с цитируемостью научных публикаций. Дается определение и демонстрируется расчет таких показателей, как импакт-фактор (двухлетний, пятилетний, совокупный, средневзвешенный, относительный), средняя цитируемость статей организации/автора, средняя нормализованная цитируемость (с нормализацией по областям науки и по журналам), SNIP, индекс Хирша и др.

Мы ставили себе целью в доступной форме представить алгоритм расчета индикаторов и, что еще более важно, донести до читателя их «физический смысл». Продемонстрировать особенности различных показателей, типичные ошибки при их расчете, дать практические рекомендации к их применению и корректной интерпретации. Значительное место занимают и подробные «кнопочные» указания — как в той или иной базе данных найти или самостоятельно рассчитать соответствующий индикатор.

### Примечание

Мы отдаем себе отчет в том, что интерфейс любой современной информационной системы претерпевает косметические или даже более серьезные изменения 1–2 раза в год, из-за чего вскоре данная составляющая нашего пособия неизбежно окажется устаревшей. Более того, базы данных пополняются новой информацией и, иногда, новыми индексируемыми источниками, что приводит к изменению чисел, фигурирующих в заданиях. Тем не менее ряд примеров по работе с ресурсами может преодолеть будущие изменения в онлайн-интерфейсах, а кроме того мы надеемся, что аккуратный читатель сумеет, имея под рукой подробные инструкции, сообразить, как надо выполнить то или иное действие в новом, изменившемся контексте.

Упомянем о некоторых соглашениях, которых мы собираемся придерживаться на протяжении данного текста. Во-первых, мы оставляем за рамками пособия вопрос о степени отражения научной значимости той или иной работы полученными ею библиографическими ссылками.

Эти споры ведутся не одно десятилетие, у них свое дискуссионное пространство, на которое не хотелось бы вторгаться в рамках данного, ставящего более технические вопросы, пособия. Если говорить несколько заостряя, в настоящем издании понятия «цитируемость» и «научный уровень» будут считаться синонимами.

Во-вторых, английское «citation», как справедливо указывалось, например, А. В. Полетаевым, означает не «цитирование» или тем более «цитату», а лишь библиографическую ссылку. Строго говоря, следует говорить не о цитируемости, а о числе полученных ссылок, «citation index» переводить как «указатель ссылок» и т. д. Тем не менее в силу ограниченности языка, необходимости использовать синонимы и трудноискренимой практики, уже сложившейся в русскоязычной литературе, в рамках настоящего издания «цитирование» будет эквивалентом «ссылке». При этом, конечно, не подразумевается, что если одна статья «цитирует» другую, то в первой содержится *выдержка, фрагмент* из второй (то, что обозначается английским словом «quotation»). Речь всего лишь о том, что вторая статья фигурирует в списке использованной литературы первой статьи.

## 2. Библиометрические инструменты. Базы данных Web of Science и Scopus

---

В настоящем пособии будут использоваться две международные библиометрические базы данных — Web of Science (WoS) и Scopus. Обе они *подписные* — они не размещены в открытом доступе, и для доступа к ним необходима платная подписка той организации, к которой вы относитесь (работаете, учитесь).

### **2.1. Принципы организации библиометрических баз данных; методологические замечания**

Что включает в себя «библиометрическая база данных» (или, как еще говорят, «база данных (научного) цитирования»), из чего она состоит и что именно в ней можно найти? Основной контент библиометрических баз

данных (как WoS, так и Scopus) — это научные журналы. При этом в базе данных *не содержится полного текста статей* этих журналов. О каждой статье хранится (и может быть выдана пользователю) следующая информация:

— библиографические сведения о статье («выходные данные»: автор(ы), название статьи, название журнала, год выхода, том, номер, страницы);

— аннотация статьи (реферат) — в том случае если она имелась в исходном тексте публикации; базы данных не составляют самостоятельно аннотации тех произведений, в которых рефераты отсутствуют;

— ключевые слова; иногда это несколько наборов ключевых слов — например, слова, приписанные статье в оригинальном тексте публикации («авторские ключевые слова»), и ключевые слова, «назначенные» базой данных исходя из внутреннего тезауруса ключевых слов и автоматических алгоритмов;

— тематика (рубрика), приписанная статье, и тип публикации (см. далее);

— организации, в которых работают авторы, т. н. «аффилиации», с почтовыми адресами мест работы и, иногда, электронными адресами авторов;

— *список цитируемой в статье литературы* — именно это поле является ключевым для базы данных и делает ее «библиометрической» базой данных;

— различные второстепенные поля (ISSN, язык оригинального документа, название и адрес издательства и др.), которые не важны для круга рассматриваемых нами вопросов.

Итак, если мы говорим, что «журнал расписывается (индексируется) библиометрической базой данных», это означает, что о каждой публикации данного журнала мы сможем узнать из базы данных сведения, перечисленные выше. Полного текста статей при этом, повторим, в базе не будет.

Помимо журнального контента, обе ведущие международные библиометрические базы включают также некоторое количество трудов конференций и, с недавнего времени, книг. При этом содержание полей, попадающих в базу данных, аналогично тому, что мы указали для журналов. Только в случае трудов конференций структурной единицей будет не статья, а доклад на конференции; в случае книг — глава из книги.

### Примечание

В случае сборников трудов конференций речь идет именно о «текстах докладов», а не о тезисах. Обычно текст такого доклада занимает от 4 страниц и более и в обязательном порядке имеет список использованной литературы. За исключением своего предназначения, он ничем не отличается от статьи в журнале (и впоследствии часто становится статьей). В серьезных научных монографиях также имеется список литературы по главам. Это дает возможность базам данных трактовать доклады и главы в полном подобии статьям и индексировать эти документы на тех же принципах.

В базах научного цитирования каждый журнал имеет «тематическую привязку», он отнесен к той или иной научной дисциплине (может быть отнесен более чем к одной). При этом при расчете практически всех индикаторов, о которых пойдет речь в настоящем пособии, тематическая рубрика *статьи* определяется на основании тематической рубрики *журнала*, в котором она опубликована. Все статьи одного журнала имеют одну и ту же рубрику (рубрики). Мы будем говорить, что статья/журнал относятся к той или иной дисциплине или научной области. Два последних понятия используются нами как синонимы.

Кроме дисциплины каждой публикации в библиометрических базах данных присваивается *тип документа*. Это может быть научная статья (Article), научный обзор (Review), письмо (Letter), книжная рецензия (Book Review) и т. д. При расчете ряда библиометрических индикаторов могут учитываться лишь некоторые типы публикаций — чаще всего это Article и Review. Более того, замечено, что разные типы публикаций обычно получают разное число ссылок. Так, документы типа Review, «научный обзор», в среднем цитируются более активно, чем типа Article, «научная статья». Тем не менее в данном пособии словом «статья» мы будем обозначать, как часто это делается в обыденной практике, любую публикацию в журнале. В тех разделах, где будет важно отличать «публикацию» от «статьи» (т. е. от публикации типа «научная статья»), читатель с легкостью поймет это различие из контекста. Кроме того, для обозначения типа документа «научная статья» мы часто будем прибегать к английскому «Article», это дополнительно избавит от двусмысленностей.

В библиометрии статью считают публикацией той или иной лаборатории/организации/страны на основании институциональных адресов, которые указаны авторами в статье и, соответственно, перене-

сены в поле «affiliation» (или «address») базы данных. В данном случае при анализе не интересуются деталями авторской биографии. Аргументы «на самом деле этот ученый приехал из другой страны», «данный автор работал во время написания статьи еще в одной организации», «рукопись долго рассматривали и на момент публикации сотрудник уже давно уволился» не принимаются во внимание в ходе стандартного библиометрического исследования. «То, что указано в статье» — это окончательный ответ на вопрос, в какой организации работал автор, ее написавший.

### **Примечание**

Строго говоря, то же самое верно и для авторства. На самом деле мы не можем знать, действительно автор участвовал в написании статьи — или был добавлен в список авторов из почтения, в результате административного давления, по взаимному соглашению и т. д. Для библиометрии он — автор статьи. Добавим также, что автор, конечно, может указывать в статьях сразу несколько мест работы, и именно в таком виде информация будет перенесена в базы цитирования.

Про отнесение публикаций к авторам, организациям и странам следует пояснить еще один момент, связанный с соавторством. Существует несколько способов счета статей, написанных несколькими авторами. Самый простой из них — «полный счет», whole counting, который предполагает, что каждому из соавторов засчитывается по одной статье. Написал автор статью в одиночку или в сотрудничестве с 20 коллегами — это все равно плюс одна публикация в его статистике. То же самое для организаций и стран: если организация (страна) хотя бы один раз указана в списке мест работы авторов статьи, эта публикация считается как целиком принадлежащая ей (и всем остальным организациям/странам соавторов наравне).

Помимо этого есть несколько способов «дробного счета», fractional counting. Если авторов в статье  $n$ , то считается, что каждый написал  $1/n$  статьи, и сложением именно таких долей находится суммарная публикационная активность автора по всем его работам. То же самое для организаций — каждая статья распределяется по всем организациям в равных долях или в долях, пропорциональных числу авторов из соответствующей организации.

### Внимание!

В рамках настоящего пособия будет использоваться только метод whole counting. Его отличает простота, однако следует помнить, что он не обеспечивает аддитивность большинства показателей. Например, если мы используем этот способ счета, то, отвечая на вопрос «сколько всего статей написали авторы организации», нельзя приводить арифметическую сумму числа статей, написанных каждым автором, — ввиду возможного (и вероятного) соавторства ученых организации при создании некоторых публикаций.

## 2.2. База данных Web of Science и аналитические надстройки

Web of Science (производитель: компания Thomson Reuters) охватывает более 12000 научных журналов, также содержит труды конференций и незначительное на данный момент количество книг. Вход в WoS:

<http://isiknowledge.com/wos>

WoS является ядром библиометрических продуктов Thomson Reuters, их главной частью. Помимо WoS также доступен (по подписке) ряд аналитических инструментов, в частности базы данных Journal Citation Reports (JCR) и Essential Science Indicators (ESI). Все эти продукты используют WoS как свою основу и лишь собирают, агрегируют из WoS данные по различным «информационным единицам» (журналам, ученым и т. д.).

JCR — база данных по библиометрическим показателям *журналов как целого*. Публикуются данные о количестве выходящих в журнале статей, числе полученных журналом ссылок, хронологическом распределении сделанных/полученных ссылок, импакт-факторе журнала (см. далее) и др. Вход в JCR:

<http://isiknowledge.com/jcr>

ESI — база данных по библиометрическим индикаторам авторов, организаций, стран, журналов. Публикуются данные о количестве вышедших у автора/организации/страны/журнала статей и их цитируемости (по журналам информация менее подробная, чем в JCR). Ограничения: в ESI включены только те организации и авторы, которые попали в 1% наиболее цитируемых хотя бы в одной научной дисциплине; и только те журналы и страны, которые попали в 50% наиболее цитируемых хотя бы в одной научной дисциплине. Помимо этого в ESI имеется специальный раздел по мировым высокоцитируемым статьям — попавшим в 1% самых цити-