



DOI: 10.19181/smtp.2022.4.3.11

EDN: MRQTVX

КЛАССИФИКАЦИЯ ПУБЛИКАЦИЙ ПО ТИПАМ ДОКУМЕНТОВ В WEB OF SCIENCE И SCOPUS: СХОДСТВА, РАЗЛИЧИЯ И ИХ ЗНАЧЕНИЕ ПРИ АНАЛИЗЕ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ

Мохначева Юлия Валерьевна¹

¹Библиотека по естественным наукам РАН, Москва,
Россия

АННОТАЦИЯ

В статье представлены сравнительные данные, касающиеся типов одних и тех же публикаций в двух ведущих мировых научно-информационных системах – Web of Science и Scopus – и у поставщиков первичной информации. В работе показаны существенные различия в типизации документов на выборке из 4338 наиболее активно цитируемых работ с российским участием различных типов за период 2010–2020 гг. В статье показано, что тип публикации – критерий далеко не очевидный: одной и той же работе может быть присвоен разный тип в зависимости от источника информации. Выявленные несоответствия информации из различных источников дают основания для определённого скепсиса в корректности постановки задач при анализе публикационной активности, касающегося учёта публикаций только определённых типов. Результаты проведённого исследования позволяют утверждать, что в случае необходимости учёта публикаций только определённых типов, сведения о типах документов необходимо сверять с исходными (первичными) сведениями у поставщиков первичной информации.

КЛЮЧЕВЫЕ СЛОВА:

публикационная активность, типы документов, Web of Science, Scopus, ошибки баз данных, учёт публикаций различных типов

ДЛЯ ЦИТИРОВАНИЯ:

Мохначева Ю. В. Классификация публикаций по типам документов в Web of Science и Scopus: сходства, различия и их значение при анализе публикационной активности // Управление наукой: теория и практика. 2022. Т. 4, № 3. С. 154–170. DOI: 10.19181/sntp.2022.4.3.11. EDN: MRQTVX

ВВЕДЕНИЕ

Ключевую роль в оценке научных исследований во всём мире играют наукометрические показатели: количество публикаций, цитируемость, импакт-факторы журналов и пр. В России ситуация с анализом публикационной активности заметно усложнилась в марте 2022 г., когда страна оказалась под колоссальным санкционным давлением: основной ресурс, являвшийся «золотым стандартом» наукометрических данных, закрыл доступ наших пользователей к своей информации^{1,2}. Кроме того, известны случаи массовых отказов зарубежных издательств в публикации работ, авторами (соавторами) которых являются российские исследователи.

К сожалению, для адекватной оценки научной продуктивности, несмотря на многочисленные недостатки, обойтись без наукометрических данных невозможно, так как другие подходы (экспертные оценки) обладают рядом существенных недостатков [1–4]. В публикации [2] отмечается, что не существует единого идеального метода оценки научных исследований, а для естественных и точных наук библиометрическая методология намного предпочтительнее рецензирования. Кроме того, экспертная оценка подразумевает очень высокие затраты и длительное время выполнения, а субъективное мнение рецензентов может зависеть от положительного или отрицательного отношения к тому или иному учёному. На субъективные оценки могут влиять реальные или потенциальные конфликты интересов; склонность давать более положительные оценки результатам известных исследователей, чем более молодым и менее авторитетным специалистам в этой области; недостаточная компетентность и кругозор, что вызывает неспособность распознать качественные аспекты исследований [1–4]. Как отмечается в публикациях [2–3], экспертная методология не универсальна, поскольку устанавливается автономно различными оценочными группами и/или отдельными рецензентами, что подвергает сравнения, связанные с этой методологией, потенциальным искажениям. К основным недостаткам рецензирования можно отнести: предвзятость рецензентов, чьи мнения часто не основаны на научных достоинствах, но зависят от личных качеств авторов, заявителей или самих рецензентов; рецензированию не хватает прогностической обоснованности, потому что существует незначительная, а часто просто отсутствует связь между мнениями рецензентов и последующей полезностью работы для научного сообщества, что впоследствии подтверждается частотой цитирования в более поздних научных работах; рецензирование задерживает публикации и препятствует публикации новых, инновационных и нетрадиционных идей; экспертиза является трудоёмким и дорогостоящим процессом и т. д. [1–4].

¹ Clarivate to Cease all Commercial Activity in Russia // Clarivate.com: [сайт]. URL: <https://clarivate.com/news/> (дата обращения: 15.08.2022).

² Доступ к Scopus на момент подготовки статьи сохранялся (июль 2022 г.)

Основная проблема при использовании наукометрических показателей – некорректная постановка задачи. Примером может служить требование учёта публикаций только определённых типов – исследовательских статей и обзоров. При этом все остальные типы публикаций – труды конференций, письма, заметки, книги, главы из книг и т. д. – оказываются вне поля зрения. Возможно, с точки зрения администраторов такой подход вызван рядом справедливых аргументов, однако на практике не всё так просто и однозначно. Проблема заключается в том, что тип публикации – критерий далеко не очевидный: одной и той же работе может быть присвоен разный тип в зависимости от источника информации. Единый стандарт присвоения типа документу отсутствует [5]. Издательства, журналы и базы данных используют свои собственные номенклатуры и определения: Web of Science (WoS) и Scopus поддерживают собственные системы присвоения типов документов, при этом корректность и точность отнесения документов к тому или иному типу вызывает вопросы, так как алгоритм этого действия не просматривается из соответствующей документации [5].

Большое количество ошибок в типизации документов влечёт за собой искажённую картину структуры публикационной активности и вызывает затруднения в процессе нормализации цитируемости [6–8]. Проблемам, связанным с особенностями и ошибками отнесения документов к тем или иным типам системами WoS и Scopus, посвящён ряд публикаций [5–13]. В статьях [6, 8] отмечается, что в Web of Science тип публикаций *Review Article* (обзорная статья) присваивается документу автоматически, если в рукописи содержится более 100 библиографических ссылок. Согласно исследованию [6] точность идентификации обзорных статей в WoS составляет 87%, и то же самое может относиться к Scopus. Кроме того, если работы опубликованы в журналах или в разделах изданий, в названии которых имеется слово *Review* (обзор), то они относятся к *Review Article* (обзорным статьям). Аналогичная ситуация с отдельными публикациями – если в названии документа присутствует слово *Review* (обзор), то они также относятся к *Review Article* (обзорным статьям) [8; 11; 14]. В руководстве Scopus объясняется методология отнесения публикаций к тому или иному типу [15]. Вот, например, как характеризуются обзоры: «Обзоры, как правило, имеют обширную библиографию. Учебные материалы, в которых рассматриваются конкретные вопросы в литературе, также считаются обзорами... В обзорах отсутствуют наиболее типичные разделы оригинальных статей, такие как материалы, методы и результаты исследований» [15]. Или, например, исследовательские статьи: «Статьи в рецензируемых журналах обычно занимают несколько страниц, чаще всего разделены на разделы: аннотация, введение, материалы и методы, результаты исследований, выводы, обсуждение и список литературы... Доклады, технические и исследовательские заметки, краткие сообщения также относятся к статьям и могут быть не более одной страницы. Статьи в отраслевых журналах обычно короче, чем в рецензируемых журналах и могут быть до одной страницы» [15]. К слову сказать, в базе данных MEDLINE/PubMed типы документов подчиняются стандартизированному и контролируемому словарю MeSH, используя в основном предложенные типы публикаций поставщиками данных – научными журналами [6].

Несоответствия при отнесении одних и тех же публикаций системами WoS и Scopus к совершенно разным типам документов приводят к негативным последствиям. Как уже упоминалось выше, из-за перекоса распределения цитирования различных типов публикаций искажаются нормализованные оценки цитируемости документов по научным тематикам [5; 7]. В статье [5] подчёркивается, что при оценке научной продуктивности и значимости публикации должны верифицироваться на соответствие типам документов. Автор отмечает, что при интерпретации результатов библиометрических исследований следует иметь в виду, что исключение публикаций некоторых типов может привести к непреднамеренной утере документов, имеющих значительную научную ценность. Так, в определённых дисциплинах и предметных областях значительная часть результатов исследований публикуется не в журналах, а в других источниках. Например, хорошо известно, что в компьютерных науках результаты исследований, как правило, публикуются в материалах конференций, в то время как книги и учебники являются более важными источниками в социальных и тем более в гуманитарных науках [10]. В работе [10] отмечается, что в некоторых дисциплинах часто отдаётся предпочтение менее популярным типам публикаций, которые могут не индексироваться в определённых базах данных. Например, в социальных и гуманитарных науках значительная часть научных результатов предназначена для широкой публики и, следовательно, часто публикуется в виде писем, отчётов, обзоров книг и пр., которые часто не включаются в основные базы данных.

Поскольку WoS часто относит материалы конференций, опубликованные в журналах, одновременно к статьям и к материалам конференций [10; 16–17], то количество публикаций определённого типа в WoS может быть искажено: при анализе по типам документов эти работы будут учитываться дважды, в то время как количество цитирований на публикацию может уменьшиться [10]. Кроме того, в статье [10] отмечены факты того, что в WoS главы авторских книг часто ошибочно индексируются как отдельные публикации.

ОСНОВНЫЕ ЗАДАЧИ, ЦЕЛИ И ОБЪЕКТ ИССЛЕДОВАНИЯ

Основная задача исследования – показать доли несоответствий в типах одних и тех же документов от различных поставщиков информации. Объект исследования – массивы наиболее активно цитируемых российских публикаций различных типов за период 2010–2020 гг. в базах данных Web of Science Core Collection (WoS CC) и Scopus по состоянию данных на март–апрель 2022 г. Цель исследования – обозначение проблемы, связанной с влиянием специфических особенностей различных ресурсов на результаты проводимого анализа публикационной активности.

МЕТОДОЛОГИЯ И МЕТОДЫ ИССЛЕДОВАНИЯ

Контрольными источниками данных³ служили ресурсы WoS CC (Clarivate Analytics) и Scopus (Elsevier). Временной интервал исследования: 2010–2020 гг.⁴

По каждому году исследуемого периода в WoS CC (все базы данных⁵) производился поиск публикаций с российским участием. Далее данные отфильтровывались по основным типам публикаций: *Articles* (статьи); *Proceedings Papers* (труды конференций); *Review Articles* (обзоры); *Book Chapters* (главы из книг); *Letters* (письма); *Data Papers* (статьи с массивами данных); *Books* (книги). Как уже говорилось во Введении, особенностью WoS CC является то, что ряд публикаций относится одновременно к двум типам публикаций. В нашем случае были выявлены следующие наиболее значимые комбинации⁶: *Article/Book Chapter* (статья/глава из книги); *Review/Book Chapter* (обзор/глава из книги); *Article/Proceedings Paper* (статья/труды конференций) и *Article/Data Paper* (статья/статья с массивами данных). Такие публикации рассматривались отдельными массивами: т. е. если публикация в базе данных относилась одновременно и к статье, и к трудам конференций, то такие документы рассматривались как статья/материалы конференций, не пересекаясь с массивом публикаций, которые были отнесены, например, только к статьям. Процесс сбора данных производился методом исключения записей, отнесённых одновременно к двум типам, не допуская при этом дублирования одной и той же публикации, благодаря чему каждый документ был учтён только один раз.

Массивы записей по каждому из указанных выше типов документов были проранжированы по нисходящей цитируемости. Такой подход позволил достичь значительной доли пересечения записей в WoS CC и Scopus⁷. Таким образом, для исследования отбирались публикации с наибольшей совокупной цитируемостью каждого из указанных выше типов за каждый год периода 2010–2020 гг.: в большинстве случаев – примерно по 50 документов. В некоторых случаях при отсутствии необходимого числа процитированных документов – в зависимости от числа этих публикаций. Посредством WoS

³ Контрольный источник данных – ресурс, данные из которого берутся за основу, а на их основе проводится дальнейшее сравнение.

⁴ На момент сбора данных (март–апрель 2022 г.) доступ к WoS и Scopus сохранялся. Тем не менее из-за санкционного давления со стороны зарубежных поставщиков научной информации с 1 мая 2022 г. был прекращён доступ к WoS. Доступ к Scopus на момент подготовки данной статьи (июль 2022 г.) сохранялся.

⁵ Science Citation Index Expanded – 1975 – 01.05.2022; Social Sciences Citation Index – 1975 – 01.05.2022; Arts & Humanities Citation Index – 1975 – 01.05.2022; Conference Proceedings Citation Index- Science – 1990 – 01.05.2022; Conference Proceedings Citation Index - Social Science & Humanities – 1990 – 01.05.2022; Book Citation Index – Science – 2005 – 01.05.2022; Book Citation Index – Social Sciences & Humanities – 2005 – 01.05.2022; Emerging Sources Citation Index (ESCI) – 2017 – 01.05.2022.

⁶ К значимым комбинациям относились только те массивы, в которых было не менее 300 документов и как минимум 100 из них были процитированы хотя бы 1 раз.

⁷ Доля анализируемых документов, которая не пересекалась в WoS и Scopus, составила 6,9% (300 из 4338 документов).

СС были собраны в отдельные массивы: статьи – 654 документа; обзоры – 551; письма – 550; статьи/главы из книг – 550; обзоры/главы из книг – 242; статьи/труды конференций – 547; статьи/статьи с массивами данных – 200; труды конференций – 549 документов.

В WoS СС отсутствует такой тип публикаций как *Notes*⁸ (заметки) [11], к которому относится достаточно большой сегмент работ в Scopus с российским участием. Для этого типа документов в качестве контрольного ресурса служила база данных Scopus: было отобрано 495 документов данного типа.

Таким образом, в исследовании использовались данные о 4338 публикациях наиболее значимых типов за 2010–2020 гг. с российским участием.

Все публикации, отнесённые к тому или иному типу документов, сравнивались с данными других систем: если контрольный массив был сформирован посредством WoS СС, то данные об этих публикациях сверялись с базой данных Scopus и с сайтами изданий (поставщиками первичной информации). И наоборот: если контрольный массив публикаций формировался с помощью Scopus, то данные сравнивались с WoS СС и с сайтами изданий.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Как уже говорилось во Введении, структура типов документов в WoS СС и Scopus сильно отличается. В WoS СС часты случаи, когда одному документу может быть приписано два типа, в то время как в Scopus документ всегда соответствует только одному (типу).

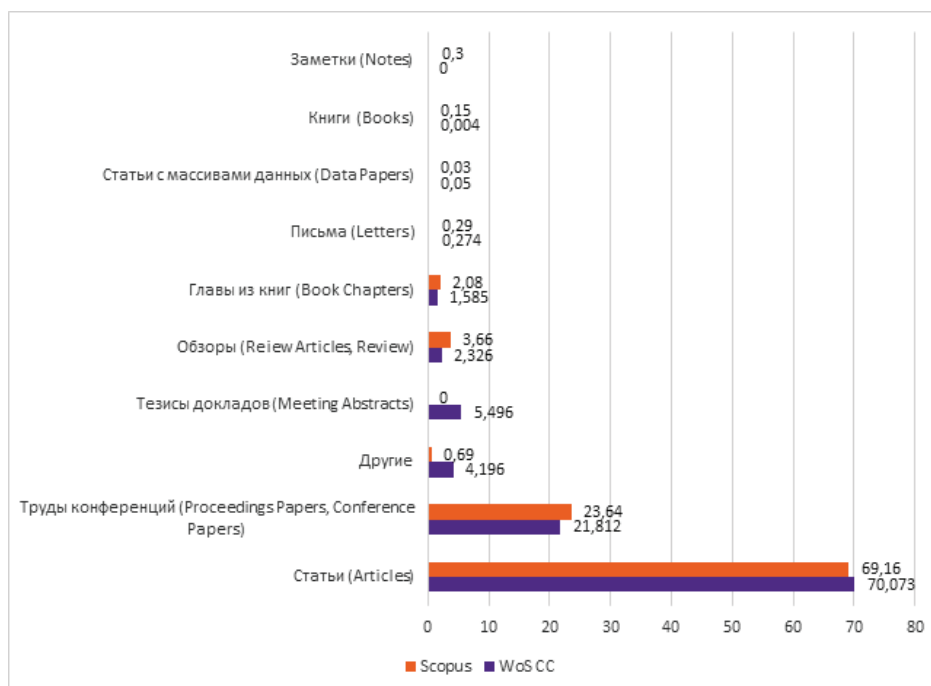


Рис. 1. Долевое распределение публикаций с российским участием по основным типам в WoS СС и Scopus. 2010–2020 гг., %

⁸ В WoS СС такая категория была постепенно упразднена [8].

На рис. 1 представлено долевое распределение всех публикаций с российским участием по различным типам документов в WoS CC и Scopus за 2010–2020 гг.

По данным рис. 1 видно, что величины долей статей, трудов конференций, писем, статей с массивами данных в WoS CC и Scopus в общем массиве с российским участием близки друг к другу. Однако имеются и значительные различия. Прежде всего – это наличие специфических типов документов, на которые приходится ощутимая доля публикаций с российским участием: тезисы докладов (WoS CC) и заметки (Scopus). Кроме того, в WoS CC заметно больше, чем в Scopus, публикаций «других» типов, что объясняется как разными подходами систем к типизации документов, так и содержательными особенностями баз данных. В данном случае мы видим долевое распределение публикаций на уровне документальных потоков в WoS CC и Scopus. Если сравнивать типы документов на уровне публикаций, то обнаружится, что часто типы одних и тех же документов отличаются в зависимости от поставщиков информации – баз данных или издательств.

В результате сравнительного анализа одних и тех же публикаций по типам в WoS CC (контрольный массив, 100%) с данными Scopus и сайтов изданий было получено долевое распределение, представленное на рис. 2.

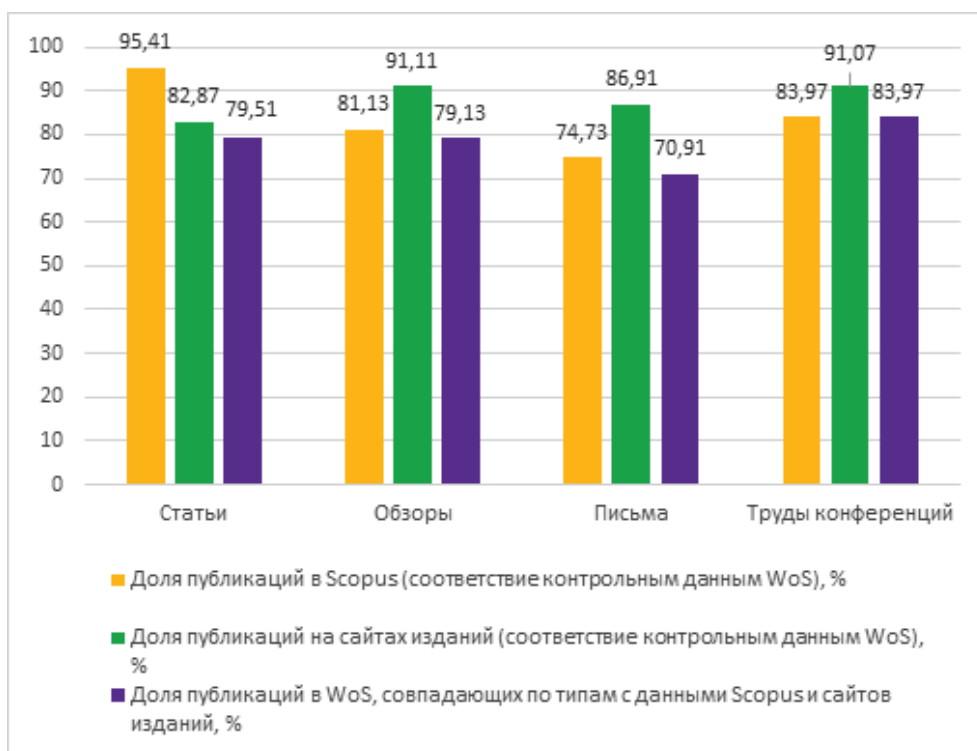


Рис. 2. Доли соответствия одних и тех же документов в Scopus и на сайтах изданий с контрольными данными WoS CC⁹, %

По данным рис. 2 видно, что наибольшее совпадение в типизации документов у контрольных массивов (WoS CC) наблюдалось с поставщиками первичной информации – сайтами изданий (издательства). Ощутимые различия

⁹ Контрольные массивы публикаций в WoS CC соответствуют 100% по каждому типу документов.

у контрольных массивов (WoS CC) наблюдались со Scopus. Если одни и те же документы с типом «статья» совпадают в WoS CC и Scopus на 95,4%, то разница по обзорам, письмам и трудам конференций весьма существенная – от 19% до 25%. Кроме того, как видно из рис. 2, ещё больше несовпадений типов одних и тех же публикаций наблюдается одновременно в WoS CC, Scopus и на сайтах изданий: от 16% до 29%.

Остановимся подробнее на выявленных соответствиях и несоответствиях в типах документов согласно информации, представленной в различных источниках информации (табл. 1, 2). Данные в табл. 1 показаны в следующей логической последовательности: столбец контрольных данных – данные из ресурса, по которому производился основной поиск документов для последующего сравнения – WoS CC. Второй столбец – типы уже выявленных контрольных публикаций в другой базе данных – Scopus и доли соответствий с контрольными данными, т. е. доли пересечений по типам документов с контрольными данными. Третий столбец – данные по типам документов из первичных источников информации (сайты изданий): здесь показаны пересечения не только с контрольной информацией в WoS CC, но и с другой (не контрольной) – Scopus. Данные табл. 2 показывают аналогичное сравнение, но в качестве контрольной базы данных использовался Scopus.

Таблица 1

Доли соответствий и несоответствий типов одних и тех же публикаций с российским участием в 2010–2020 гг. согласно данным WoS CC (контрольные данные), Scopus и поставщиков первичной информации (официальные сайты изданий)

Типы публикаций в WoS CC и количество документов ¹⁰	Типы публикаций в Scopus: соответствие контрольным данным в WoS CC		Типы публикаций на сайтах изданий: соответствие контрольным данным в WoS CC и Scopus	
	Тип публикации	Доля публикаций, %	Тип публикации	Доля публикаций, %
Статьи (654 документа, 100%)	Статьи	95,41	Статьи	79,511
			Письма	5,963
			Доклады	5,199
			Короткие сообщения	2,141
			Методические рекомендации	1,376
			Предложение редакторов	0,612
			Программное обеспечение	0,306
			Заметки	0,153
			Главы из книг	0,153
	Обзоры	3,823	Статьи	2,599
			Обзоры	1,07
			Специальные сообщения	0,153
	Заметки	0,153	Статьи	0,153
Труды конференций	0,612	Статьи	0,612	

¹⁰ Методология формирования массива документов определённых типов рассмотрена в соответствующем разделе данной статьи.

Продолжение табл. 1				
Обзоры (551 документ, 100%)	Обзоры	81,13	Обзоры	79,129
			Статьи	1,633
			Рекомендации	0,181
			Доклады	0,181
	Статьи	17,24	Обзоры	10,526
			Статьи	5,082
			Рекомендации	0,907
			Комментарии	0,181
			Письма	0,181
			Методические рекомендации	0,181
			Специальные сообщения	0,181
	Труды конференций	0,726	Обзоры	0,544
			Труды конференций	0,181
Краткие обзоры	0,544	Обзоры	0,544	
Данные о публикациях отсутствуют	0,363	Обзоры	0,363	
Письма (550 документов, 100%)	Письма	74,73	Письма	70,909
			Статьи	3,455
			Краткие сообщения	0,182
			Реплики	0,182
	Статьи	21,64	Письма	13,273
			Статьи	8,182
			Комментарии	0,182
	Обзоры	0,909	Письма	0,545
			Комментарии	0,182
			Статьи	0,182
	Заметки	1,818	Письма	1,455
			Комментарии	0,364
	Краткие обзоры	0,545	Письма	0,364
Статьи			0,182	
Труды конференций	0,364	Письма	0,364	
Статьи-Главы из книг (550 документов, 100%)	Главы из книг	51,636	Главы из книг	51,636
	Данные о публикациях отсутствуют	28	Главы из книг	26,909
			Статьи	0,727
			Книги	0,182
			Труды конференций	0,182
	Статьи	16,364	Главы из книг	13,636
			Статьи	2,364
			Обзоры	0,182
			Труды конференций	0,182
	Труды конференций	0,909	Главы из книг	0,909
	Обзоры	2	Главы из книг	1,091
			Статьи	0,727
			Обзоры	0,182
Книги	1,091	Главы из книг	0,909	
		Книги	0,182	

Продолжение табл. 1				
Обзоры-Главы из книг (242 документа, 100%)	Обзоры	4,959	Обзоры	4,132
			Главы из книг	0,826
	Главы из книг	62,4	Главы из книг	61,983
			Статьи	0,413
	Статьи	21,07	Обзоры	9,504
			Статьи	9,091
			Главы из книг	1,24
	Книги	2,066	Доклады	1,24
			Главы из книг	1,653
	Данные о публикациях отсутствуют	9,504	Обзоры	0,413
			Главы из книг	8,264
			Доклады	0,826
Статьи-Труды конференций (547 документов, 100%)	Статьи	73,68	Статьи	70,018
			Труды конференций	1,28
			Письма	0,183
			Доклады	0,914
			Обзоры	0,731
			Краткие сообщения	0,548
	Труды конференций	19,38	Статьи	17,367
			Труды конференций	1,097
			Обзоры	0,366
			Краткие сообщения	0,366
	Обзоры	6,764	Технические записки	0,183
			Обзоры	5,667
Статьи			0,914	
Краткие обзоры	0,183	0,183	Статьи	0,183
Статьи-Статьи с массивами данных (200 документов, 100%)	Статьи	28,5	Статьи с массивами данных	25,5
			Статьи	2
			Обзоры	0,5
			Примечание к данным	0,5
	Статьи с массивами данных	71	Статьи с массивами данных	64,5
			Дескриптор данных	6,5
Заметки	0,5	0,5	Статьи с массивами данных	0,5
Труды конференций (549 документов, 100%)	Труды конференций	83,97	Труды конференций	83,971
	Статьи	6,011	Статьи	3,643
			Главы из книг	1,639
			Труды конференций	0,729
	Главы из книг	1,275	Главы из книг	0,911
			Труды конференций	0,364
	Данные о публикациях отсутствуют	8,743	Труды конференций	6,011
			Статьи	1,821
Главы из книг			0,911	

Как видно из табл. 1, типы одних и тех же документов могут сильно различаться в зависимости от поставщиков информации. Как уже упоминалось выше, в WoS и Scopus существуют свои собственные методологии классификации документов по типам. Однако здесь мы сталкиваемся с очень непростой проблемой: как быть, если при учёте публикаций существуют требования по отсечению публикаций определённых типов. Мы видим (табл. 1), что, например, к трудам конференций (версия WoS CC) относятся статьи, главы из книг, обзоры (версии Scopus) и, согласно данным поставщиков первичной информации, – сайтов изданий.

Что же касается типов публикаций, которые часто не рекомендуют учитывать – «письма», «труды конференций», «заметки» и т. д., то, например, из 550 «писем» контрольного массива (WoS CC) 21,64% значились в Scopus как «статьи», а, согласно информации у первичных поставщиков информации, к типу «статья» было отнесено 12% «писем» контрольного массива.

Не совсем понятна в WoS CC логика и смысл одновременного отнесения к «статьям» и «главам из книг» ряда публикаций. Мы видим (табл. 1), что большая часть этих документов в Scopus и у поставщиков первичной информации относилась к «главам из книг». На долю «статей» приходилось: 16,4% – в Scopus и лишь 3,8% – по версии первоисточников. Такая типизация документов не выглядит оправданной ещё и потому, что другие аналогичные публикации (из тех же самых сборников) в других случаях относятся только к одному типу по версии WoS CC: «статьям», «материалам конференций», «обзорам» и т. д.

Ещё более странная ситуация сложилась с одновременным отнесением в WoS CC публикаций к «статьям» и «трудам конференций». Так, по версии Scopus лишь 19,38% этих работ отнесены к «трудам конференций». Остальные – к «статьям» и «обзорам». По версиям поставщиков первичной информации, на долю «трудов конференций» пришлось всего лишь около 3,47%. Остальные работы обозначены как: «статьи» (88,48%); «обзоры» (6,76%) и другие типы публикаций («письма», «краткие сообщения», «технические записки») – около 5%.

Если допустить, что при анализе публикационной активности будут отсечены «труды конференций» и (или) другие «малозначимые» типы публикаций, то велики риски утраты сведений о значимых работах, что особенно негативно может отразиться при анализе публикационной активности на мезо- и микроуровнях (отдельные научные организации и учёные).

Аналогичная ситуация наблюдалась и в случае с «заметками» в Scopus (табл. 2).

Таблица 2

Доли соответствий и несоответствий типов одних и тех же публикаций с российским участием в 2010–2020 гг., представленных в Scopus (контрольные данные) как «Заметки», с типами документов в WoS CC и у поставщиков первичной информации (официальные сайты изданий)

Тип публикаций в Scopus (контрольные данные)	Публикации в WoS CC		Публикаций на сайтах изданий	
	Тип публикации	Доля публикаций, %	Тип публикации	Доля публикаций, %
Заметки (495 документов)	Редакционные материалы	49,091	Комментарии	26,67
			Статьи	15,96
			Новости	2,22
			Ответы на комментарии	1,62
			Заметки	1,01
			Письма	0,61
			Предложения	0,61
			Эссе	0,2
			Рисунки	0,2
	Статьи	31,515	Статьи	17,98
			Заметки	5,66
			Комментарии	3,23
			Объявления	2,02
			Короткие сообщения	1,41
			Письма	0,81
			Приложения	0,2
			Лекции	0,2
			Обзоры	0,2
			Дополнения	0,2
	Данные о публикациях отсутствуют	14,747	Статьи	8,28
			Заметки	2,42
			Комментарии	1,01
			Труды конференций	1,01
			Письма	0,4
			Редакционные материалы	0,4
			Методические рекомендации	0,2
			Предложения	0,2
			Доклады	0,2
			Короткие сообщения	0,2
	Письма	2,02	Письма	1,41
			Комментарии	0,61
	Новости	1,212	Новости	0,61
			Заметки	0,61
Статьи-Труды конференций	0,404	Комментарии	0,2	
		Труды конференций	0,2	
Обзоры	0,404	Статьи	0,4	
Статьи-Статьи с массивами данных	0,202	Статьи с массивами данных	0,2	
Исправления	0,202	Исправления	0,2	
Заметки	0,202	Новости	0,2	

Как видно из табл. 2, наибольшие доли «заметок» выявлены в WoS CC под типами документов: «редакционные материалы» – 49% и «статьи» – чуть более 35%. У поставщиков первичной информации исследуемые «заметки» были отнесены к очень разнообразному спектру типов документов: «комментарии» и «ответы» на них; «новости»; «письма»; «предложения»; «эссе»; «рисунки»; «объявления»; «приложения»; «дополнения»; «исправления» и пр. Неудивительно, что около 15% «заметок» Scopus не были отражены в WoS CC, прежде всего по причине несоответствия требованиям индексируемого контента.

ВЫВОДЫ

На основании данных, представленных в этой статье, можно сделать несколько выводов:

- как WoS, так и Scopus не предоставляют в полной мере адекватных и корректных данных о типах документов, а методология отнесения публикаций к различным типам документов непрозрачна и вызывает множество вопросов;
- при анализе публикационной активности особенно на мезо- и микроуровнях необходимо сверять типы публикаций с первичной информацией о документах на сайтах поставщиков;
- используя метод «отсечения» определённых типов публикаций при анализе публикационной активности велики риски утраты учёта очень значимых публикаций.

Представленные в данной публикации сведения показывают как несовершенство наукометрических систем, так и связанные с этим риски в процессе формулирования задач для проведения анализа публикационной активности. Однако отдельно стоит оговориться, что несмотря на наличие множества недостатков, альтернативы наукометрическим данным из авторитетных информационных ресурсов при анализе публикационной активности не существует.

ЛИТЕРАТУРА

1. Лопатина Н. В. О новых подходах к оценке научной деятельности: рассуждения авторов-рецензентов / Н. В. Лопатина, В. А. Цветкова // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2021. № 4. С. 38–42. DOI: 10.36535/0548-0019-2021-04-5. EDN: SEWRHC.
2. Abramo G. Evaluating research: from informed peer review to bibliometrics / G. Abramo, C. A. D'Angelo // Scientometrics. 2011. Vol. 87, № 3. P. 499–514. DOI: 10.1007/s11192-011-0352-7.
3. Pendlebury D. A. The use and misuse of journal metrics and other citation indicators // Scientometrics. 2009. Vol. 57, № 1. P. 1–11. DOI: 10.1007/s00005-009-0008-y.
4. Bornmann L. Scientific Peer Review // Annual Review of Information Science and Technology. 2011. Vol. 45. P. 197–245. DOI: 10.1002/aris.2011.1440450112.

5. *Donner P.* Document type assignment accuracy in the journal citation index data of Web of Science // *Scientometrics*. 2017. Vol. 113. P. 219–236. DOI: 10.1007/s11192-017-2483-y.
6. *Blümel C.* Studying review articles in scientometrics and beyond: a research agenda / C. Blümel, A. Schniedermann // *Scientometrics*. 2020. Vol. 124. P. 711–728. DOI: 10.1007/s11192-020-03431-7.
7. *Ioannidis J. P. A.* Citation metrics: A primer on how (not) to normalize / J. P. A. Ioannidis, K. Boyack, P. F. Wouters // *PLoS Biology*. 2016. Vol. 14. e1002542. DOI: 10.1371/journal.pbio.100254.
8. *Harzing A.-W.* Document categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences? // *Scientometrics*. 2013. Vol. 94. P. 23–34. DOI: 10.1007/s11192-012-0738-1.
9. The role of editorial material in bibliometric research performance assessments / T. van Leeuwen, R. Costas, C. Calero-Medina, M. Visser // *Scientometrics*. 2013. Vol. 95. P. 817–828. DOI: 10.1007/s11192-012-0904-5.
10. *Pranckutė R.* Web of science (WoS) and Scopus: The titans of bibliographic information in today's academic world // *Publications*. 2021. Vol. 9, №1. Art. №12. DOI: 10.3390/publications9010012.
11. *Sigogneau A.* An analysis of document types published in journals related to physics: Proceeding papers recorded in the Science Citation Index database // *Scientometrics*. 2000. Vol. 47. № 3. P. 589–604.
12. *Yeung A. W. K.* Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers // *Current Science*. 2019. Vol. 116, № 11. P. 1909–1914. DOI: 10.18520/cs/v116/i11/1909-1914.
13. *Yeung A. W. K.* Document type assignment by Web of Science, Scopus, PubMed, and publishers to “Top 100” papers // *Malaysian Journal of Library & Information Science*. 2021. Vol. 26, № 3. P. 97–103. DOI: 10.22452/mjlis.vol26no3.5.
14. *Ketcham C. M.* The impact of review articles / C. M. Ketcham, J. M. Crawford // *Laboratory Investigation*. 2007. Vol. 87. P. 1174–1185. DOI: 10.1038/labinvest.3700688.
15. Scopus Content Coverage Guide // Elsevier.com : [сайт]. URL: https://www.elsevier.com/__data/assets/pdf_file/0007/69451. P. 11–12. (дата обращения: 16.08.2022).
16. *Visser M.* Large-Scale Comparison of Bibliographic Data Sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic / M. Visser, N. Jan Van Eck, L. Waltman // *Arxiv.org* : [сайт]. URL: <https://arxiv.org/ftp/arxiv/papers> (дата обращения: 16.08.2022).
17. *Moed H. F.* Trends in Russian Research Output Indexed in Scopus and Web of Science / H. F. Moed, V. Markusova, M. Akoev // *Scientometrics*. 2018. Vol. 116. № 2. P. 1153–1180. DOI: 10.1007/s11192-018-2769-8.

Статья поступила в редакцию 05.08.2022.

Одобрена после рецензирования 22.08.2022. Принята к публикации 29.08.2022.

СВЕДЕНИЯ ОБ АВТОРЕ

Мохначева Юлия Валерьевна *j-v-m@yandex.ru*

Кандидат педагогических наук, ведущий научный сотрудник, заведующая отделом, Библиотека по естественным наукам РАН, Москва, Россия

AuthorID РИНЦ: 68138

ORCID ID: 0000-0001-5780-485X

Scopus ID: 54880572900

Web of Science ResearcherID: AAI-7181-2020

DOI: 10.19181/smtp.2022.4.3.11

CLASSIFICATION OF PUBLICATIONS BY DOCUMENT TYPES IN WEB OF SCIENCE AND SCOPUS: SIMILARITIES, DIFFERENCES AND THEIR VALUE IN THE ANALYSIS OF PUBLICATION ACTIVITY

Yuliya V. Mokhnacheva¹

¹Library for Natural Sciences of the RAS, Moscow, Russia

Abstract. The article presents comparative data concerning the types of the same publications in two of the world's leading scientific information systems – Web of Science and Scopus – and publishers. The paper shows significant differences in the typification of documents in a sample of 4338 most actively cited works with Russian authorship of various types for the period 2010-2020. The article shows that the type of publication is not an obvious criterion: the same work can be assigned a different type depending on the source of information. The revealed discrepancies of information from various sources give grounds for a certain skepticism in the correctness of the formulation of tasks in the analysis of publication activity concerning the accounting of publications of only certain types. The results of the study suggest that it is necessary to take into account only certain types of publications, information about the types of documents should be verified with the original information from publishers.

Keywords: publication activity, document types, Web of Science, Scopus, database errors, accounting of publications of various types

For citation: Mokhnacheva, Y. V. (2022). Classification of Publications by Document Types in Web of Science and Scopus: Similarities, Differences and their Value in the Analysis of Publication Activity. *Science Management: Theory and Practice*. Vol. 4, no. 3. P. 154–170. DOI: 10.19181/smtp.2022.4.3.11

REFERENCES

1. Lopatina, N. V. and Tsvetkova, V. A. (2021). On new approaches to the evaluation of scientific activity: arguments of reviewers. *Nauchno-tekhnicheskaya informatsiya. Seriya 1: Organizatsiya i metodika informatsionnoi raboty*. No. 4. P. 38–42. DOI: 10.36535/0548-0019-2021-04-5 (In Russ.).
2. Abramo, G. and D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*. Vol. 87, no. 3. P. 499–514. DOI: 10.1007/s11192-011-0352-7.
3. Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Scientometrics*. Vol. 57, no. 1. P. 1–11. DOI: 10.1007/s00005-009-0008-y.
4. Bornmann, L. (2011). Scientific Peer Review. *Annual Review of Information Science and Technology*. Vol. 45. P. 197–245. DOI: 10.1002/aris.2011.1440450112.

5. Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics*. Vol. 113. P. 219–236. DOI: 10.1007/s11192-017-2483-y.
6. Blümel, C. and Schniedermann, A. (2020). Studying review articles in scientometrics and beyond: a research agenda. *Scientometrics*. Vol. 124. P. 711–728. DOI: 10.1007/s11192-020-03431-7.
7. Ioannidis, J. P. A., Boyack, K. and Wouters, P. F. (2016). Citation metrics: A primer on how (not) to normalize. *PLoS Biology*. Vol.14. e1002542. DOI: 10.1371/journal.pbio.1002542.
8. Harzing, A.-W. (2013). Document categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences? *Scientometrics*. Vol. 94. P. 23–34. DOI: 10.1007/s11192-012-0738-1.
9. van Leeuwen, T., Costas, R., Calero-Medina, C. and Visser, M. (2013). The role of editorial material in bibliometric research performance assessments. *Scientometrics*. Vol. 95. P. 817–828. DOI: 10.1007/s11192-012-0904-5.
10. Prancutė, R. (2021). Web of science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*. Vol. 9, no. 1. Art. 12. DOI: 10.3390/publications9010012.
11. Sigogneau, A. (2000). An analysis of document types published in journals related to physics: Proceeding papers recorded in the Science Citation Index database. *Scientometrics*. Vol. 47, no. 3. P. 589–604.
12. Yeung, A. W. K. (2019). Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers. *Current Science*. Vol. 116, no. 11. P. 1909–1914. DOI: 10.18520/cs/v116/i11/1909-1914.
13. Yeung, A. W. K. (2021). Document type assignment by Web of Science, Scopus, PubMed, and publishers to “Top 100” papers. *Malaysian Journal of Library & Information Science*. Vol. 26, no. 3. P. 97–103. DOI: 10.22452/mjlis.vol26no3.5.
14. Ketcham, C. M. and Crawford, J. M. (2007). The impact of review articles. *Laboratory Investigation*. Vol. 87. P. 1174–1185. DOI: 10.1038/labinvest.3700688.
15. Scopus Content Coverage Guide. *Elsevier.com*. URL: https://www.elsevier.com/___data/assets/pdf_file/0000/0000/0000.pdf. P. 11–12 (accessed 16.08.2022).
16. Visser, M., Jan Van Eck, N. and Waltman, L. Large-Scale Comparison of Bibliographic Data Sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Arxiv.org*. URL: <https://arxiv.org/ftp/arxiv/papers/2005/2005.10732.pdf> (accessed 16.08.2022).
17. Moed, H. F., Markusova, V. and Akoev, M. (2018). Trends in Russian Research Output Indexed in Scopus and Web of Science. *Scientometrics*. Vol. 116, no. 2. P. 1153–1180. DOI: 10.1007/s11192-018-2769-8.

The article was submitted on 05.08.2022.

Approved after reviewing 22.08.2022. Accepted for publication 29.08.2022.

INFORMATION ABOUT THE AUTHOR

Mokhnacheva Yuliya *j-v-m@yandex.ru*

Candidate of Pedagogics, Leading Researcher, Head of Department Library for Natural Sciences of the RAS, Moscow, Russia

AuthorID ПИИЦ: 68138

ORCID ID: 0000-0001-5780-485X

Scopus ID: 54880572900

Web of Science ResearcherID: AAI-7181-2020