

Типы документов, индексируемых в базах данных WoS и Scopus: сходства, различия и их значение при анализе публикационной активности

Представлены сравнительные характеристики типов одних и тех же публикаций, проиндексированных в двух ведущих мировых научно-информационных базах данных – Web of Science и Scopus и поставщиками первичной информации. На выборке из 3843 наиболее активно цитируемых различных типов статей с российским участием показаны существенные различия в типизации документов за период 2010–2020 гг. Выявлено, что тип публикации – критерий далеко не очевидный: одной и той же работе может быть присвоен разный тип в зависимости от источника – поставщика информации. Установленные несоответствия дают основания для определённой неуверенности в корректности постановки задач при анализе публикационной активности, касающемся учёта публикаций только определённых типов. Результаты настоящего исследования позволяют утверждать, что при учёте публикаций только определённых типов, данные 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

DOI: 10.36535/0548-0019-2023-01-4

Основная проблема при использовании наукометрических показателей в оценке публикационной активности – некорректная постановка задачи. Примером может служить требование учёта публикаций только определённых типов – исследовательских статей и обзоров. При этом все остальные типы – труды конференций, письма, заметки, книги, главы из книг и т. д. – оказываются вне поля зрения. Проблема заключается в том, что тип публикации – это критерий далеко не очевидный: одной и той же работе может быть присвоен разный тип в зависимости от источника информации. Единый стандарт отнесения документа к определённому типу отсутствует. Издательства, журналы и базы данных используют собственные номенклатуры и определения. Так, *Web of Science (WoS)* и *Scopus* поддерживают свои системы присвоения типов документам, при этом корректность и точность отнесения документов к тому или иному типу вызывает вопросы, так как алгоритм этого действия не просматривается в соответствующих методических установках [1].

Большое количество ошибок в типизации документов влечёт за собой искажённую картину структуры публикационной активности и вызывает затруднения для нормализации цитируемости. В статьях [2, 3] отмечается, что в *Web of Science* тип публикаций *Review Article* (обзорная статья) присваивается документу автоматически, если в рукописи содержится более 100 библиографических ссылок. Согласно ис-

следованию [2] точность идентификации обзорных статей в *WoS* составляет 87%, и то же может относиться к *Scopus*. Кроме того, если работы опубликованы в журналах или в разделах изданий, в названии которых имеется слово *Review* (обзор), то автоматически 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) [3–5].

В руководстве *Scopus* объясняется методика отнесения публикаций к тому или иному типу [6]. Например, обзоры характеризуются следующим образом: «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» [6], а исследовательские статьи так: «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» [6]. В базе данных

MEDLINE/PubMed типы документов подчиняются стандартизированному и контролируемому словарю MeSH, учитывая в основном типы, предложенные поставщиками информации – научными журналами [2].

Несоответствия при отнесении одних и тех же публикаций к совершенно разным типам документов системами WoS и Scopus приводят к негативным последствиям: из-за «перекося» в распределении цитирования различных типов публикаций искажаются нормализованные оценки цитируемости документов по научным тематикам [1, 7]. В статье [1] подчёркивается, что при оценке научной продуктивности и значимости публикации должны верифицироваться на соответствие типам документов и отмечается, что при интерпретации результатов библиометрических исследований следует иметь в виду, когда исключаются публикации некоторых типов – это 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: хорошо известно, что в компьютерных науках результаты исследований, как правило, публикуются в материалах конференций, в то время как книги и учебники – это более важные источники в социальных и тем более в гуманитарных науках. В работе [8] 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, публикуется в виде писем, отчётов, обзоров, которые, как правило, не включаются в основные базы данных.

Поскольку WoS часто относит материалы конференций, опубликованные в журналах, одновременно и к статьям, и к материалам конференций [8–10], 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. Кроме того, в статье [8] приведены факты того, что в WoS главы в авторских книгах часто ошибочно индексируются как отдельные публикации. Ошибкам и несоответствиям отнесения публикаций к различным типам в различных базах данных посвящены публикации [11–13].

В нашем исследовании объектом были 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.; целью – обозначение проблемы, связанной с влиянием специфических особенностей различных ресурсов на результаты анализа публикационной активности; контрольным источником данных¹ мы назначили ресурсы WoS CC (Clarivate Analytics); временной ин-

тервал – 2010–2020 гг.². Более подробно результаты проведённого исследования опубликованы в работе [14].

По каждому году изучаемого периода в WoS CC (все базы данных³) выполнялся поиск публикаций с российским участием. Далее данные отфильтровывались по основным типам: *Articles* (статьи); *Proceedings Papers* (труды конференций); *Review Articles* (обзоры); *Book Chapters* (главы из книг); *Letters* (письма); *Data Papers* (статьи с массивами данных); *Books* (книги). Особенностью WoS CC является то, что ряд публикаций относится одновременно к двум типам. В нашем случае были выявлены следующие наиболее значимые комбинации⁴: *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 как отдельные массивы, т. е. если публикация в базе данных относилась одновременно и к статье, и к трудам конференций, то она рассматривалась как статья/материалы конференций, не пересекаясь с массивом публикаций, которые были отнесены, например, только к статьям. Сбор 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 нами типов документов были проранжированы по нисходящей цитируемости. Такой подход позволил установить значительную долю пересечения записей в WoS CC и Scopus⁵. Для нашего исследования отбирались публикации с наибольшей совокупной цитируемостью каждого из указанных типов за каждый год периода 2010–2020 гг.: в большинстве случаев – примерно по 50 документов, при отсутствии необходимого числа процитированных документов – в зависимости от числа этих публикаций. С помощью WoS CC 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 –

² 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.

³ *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.

⁴ 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.

⁵ Доля анализируемых документов, которая не пересекалась в WoS и Scopus, составила 6%.

¹ Контрольный источник данных – ресурс, из которого берутся данные и на их основе проводится дальнейшее сравнение.

200; труды конференций – 549 документов. Таким образом, в нашем исследовании были сведения о 3843 публикациях с российским участием наиболее значимых типов за 2010–2020 гг.

Все публикации, отнесённые к тому или иному типу, сравнивались с данными других систем: контрольный массив (БД *WoS CC*) сверялся с БД *Scopus* и с сайтами изданий – поставщиков первичной информации. Таким образом было получено долевое распределение публикаций. Наибольшее совпадение в типизации документов у контрольных массивов наблюдалось с поставщиками первичной информации – сайтами изданий

(издательства). Ощутимые различия у контрольных массивов наблюдались с БД *Scopus*. Если одни и те же документы типа «статья» совпадали в *WoS CC* и *Scopus* на 95,4%, то разница по обзорам, письмам и трудам конференций весьма существенная – от 19% до 25%. Кроме того, ещё больше несовпадений одних и тех же публикаций по типам 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 из различных источников, представленной в таблице.

Соответствия и несоответствия (%) 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 <i>WoS CC</i> / кол-во документов ¹	в <i>Scopus</i> : соответствие контрольным данным в <i>WoS CC</i> ²		на сайтах изданий: соответствие контрольным данным в <i>WoS CC</i> и <i>Scopus</i> ³	
	тип	доля, %	тип	доля, %
Статьи / 654	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	
Труды конференций	0,612	Статьи	0,612	
Обзоры / 551	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
	Краткие обзоры	0,544	Труды конференций	0,181
данные о публикациях отсутствуют	0,363	Обзоры	0,544	
Письма / 550	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	

Типы публикаций				
в WoS CC/ кол-во документов ¹	в Scopus: соответствие контрольным данным в WoS CC ²		на сайтах изданий: соответствие контрольным данным в WoS CC и Scopus ³	
	тип	доля, %	тип	доля, %
Статьи-Главы из книг /550	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	0,909	Труды конференций	0,182
	Обзоры	2	Главы из книг	0,909
Главы из книг			1,091	
Статьи			0,727	
Книги	1,091	Обзоры	0,182	
		Обзоры	0,182	
		Обзоры	0,182	
Обзоры-Главы из книг / 242	Обзоры	4,959	Главы из книг	4,132
			Главы из книг	0,826
	Главы из книг	62,4	Главы из книг	61,983
			Статьи	0,413
			Обзоры	9,504
	Статьи	21,07	Статьи	9,091
			Главы из книг	1,24
			Доклады	1,24
	Книги	2,066	Главы из книг	1,653
			Обзоры	0,413
данные о публикациях отсутствуют	9,504	Главы из книг	8,264	
		Доклады	0,826	
		Статьи	0,413	
Статьи-Труды кон- ференций / 547	Статьи	73,68	Статьи	70,018
			Труды конференций	1,28
			Письма	0,183
			Доклады	0,914
			Обзоры	0,731
	Труды конференций	19,38	Краткие сообщения	0,548
			Статьи	17,367
			Труды конференций	1,097
			Обзоры	0,366
	Обзоры	6,764	Краткие сообщения	0,366
Технические записки			0,183	
Обзоры			5,667	
Краткие обзоры	0,183	Статьи	0,914	
		Труды конференций	0,183	
Статьи-Статьи с массивами данных / 200	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
Заметки	0,5	Дескриптор данных	6,5	
		Статьи с массивами данных	0,5	
Труды конференций / 549	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	

ПРИМЕЧАНИЯ:

¹ контрольные данные из ресурса, по которому производился основной поиск документов для последующего сравнения – WoS CC;

² 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;

³ 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*.

Как видно из таблицы, одни и те же документы могут сильно различаться по типам в зависимости от поставщиков информации. Как уже упоминалось, в базах данных *WoS* и *Scopus* существуют свои собственные методики 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, например, что по версии *WoS CC* к трудам конференций относятся статьи, главы из книг, а по версии *Scopus* – обзоры и, согласно данным поставщиков первичной информации, сайты 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 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. Например, (см. таблицу), большая часть этих документов в БД *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 аналогичные публикации (из тех же самых сборников) в других случаях относятся только к одному типу по версии *WoS CC* – к «статьям», «материалам конференций», «обзорам» и т. д.

Ещё более странная ситуация сложилась с отнесением в *WoS CC* публикаций одновременно к «статьям» и «трудом конференций». Так, по версии *Scopus* лишь 19,38% этих работ отнесены к «трудом конференций», остальные – к «статьям» и «обзорам». По версиям поставщиков первичной информации, на долю «трудом конференций» пришлось всего лишь около 2,56%, остальные публикации обозначены как «статьи» (88,48%), «обзоры» (6,76%) и другие типы («письма», «краткие сообщения», «технические записки») – около 2,5%.

Если допустить, что при анализе публикационной активности отдельных учёных и научных организаций будут отсечены «трудом конференций» и (или) другие «малозначимые» типы публикаций, то велики риски утраты сведений о значимых работах, что особенно негативно может отразиться при исследованиях публикационной активности на мезо- и микроуровнях (отдельных научных организаций и учёных).

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

- как *WoS*, так и *Scopus* не предоставляют в полной мере адекватных и корректных данных о типах документов, а их методика отнесения публика-

ций к различным типам документов непрозрачна и вызывает много вопросов;

- при анализе публикационной активности особенно на мезо- и микро- уровнях 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 очень значимых публикаций.

Представленные в настоящей статье сведения показывают как несовершенство наукометрических систем, так и связанные с этим риски в процессе формулирования задач для изучения анализа публикационной активности.

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. *Scopus Content Coverage Guide* // Elsevier.com: – URL: https://www.elsevier.com/_data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf. P. 11–12. (дата обращения: 16.08.2022).
7. Ioannidis J.P.A., Boyack K., Wouters P.F. Citation metrics: A primer on how (not) to normalize // *PLoS Biology*. – 2016. – Vol. 14. – e1002542. DOI: 10.1371/journal.pbio.100254.
8. Prancutè 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.
9. Visser M., M. Visser, Van Eck N. Jan, 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> (дата обращения: 16.08.2022).

10. Moed H.F., Markusova V., Akoev M. Trends in Russian Research Output Indexed in Scopus and Web of Science // *Scientometrics*. – 2018. – Vol. 116, № 2. – P. 1153–1180. DOI: 10.1007/s11192-018-2769-8.
11. 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.
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. 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: <https://doi.org/10.19181/smtp.2022.4.3.11>.

Материал поступил в редакцию 27.10.22.

Сведения об авторе

МОХНАЧЕВА Юлия Валерьевна – кандидат педагогических наук, ведущий научный сотрудник, заведующий отделом Библиотеки по естественным наукам РАН, Москва
e-mail: j-v-m@yandex.ru